Causal Representation of the Simultaneous Localization And Mapping Problem

Joël Tari, Patrick Danès

Causal Representation of the Simultaneous Localization And Mapping Problem

Joel Tari[1 2] Patrick Danès[2]

[1] *joel.tari@akkodis.com*, Akkodis Research, 7 boulevard Henri Ziegler, Blagnac, France

[2] *patrick.danes@laas.fr*, LAAS-CNRS, 7 avenue du colonel Roche, Toulouse, France

Forewarning

This document is in a v0 stage and will be submitted as a monograph. Please be aware that its content will be subject to changes, additions, or corrections in subsequent versions. This work was primaraly conducted as part of a doctoral thesis (Tari, 2024) in a collaboration between Akkodis and CNRS[1]. The authors would like to particularly thank the members of the thesis comitee (Tari, 2024) for the stimulating discussions, as well as valuable suggestions and encouragements to publish this work.

[1]The authors acknowledge Akkodis, CNRS & ANRT institutions for supporting this work through the grant CIFRE 2018/0508. The first author would like to thank the Akkodis Research institution, Marie-Anne Bauda, Mathieu Jeudy and Benoit Baurens for according extra time on the project.

Abstract

We address the problem of spatial inference in robotics, commonly known as "Simultaneous Localization And Mapping" (SLAM), through the lens of causal inference. State-of-the-art approaches generally proceed by feeding the SLAM system with noisy sensor measurements and statistical assumptions, so as to get the joint posterior probability distribution of the decision variables conditioned on the measurements. Probabilistic inference techniques are used to produce the best posterior estimate. In contrast, in the proposed approach, the estimation of the decision variables is conducted on a scheme which makes use of causal assumptions. Since understanding the distinction between the probabilistic and the causal paradigms is paramount in this work, a doctrinal synthesis is provided, supported by an historical analysis. In particular, it is reminded that probability theory pioneers like Laplace handled uncertainty within a quasi-deterministic framework. Such a conception found an apropos application in the geodetic adjustment problem (triangulations for cartography), which shares many structural commonalities with SLAM. Using modern causal inference tools, in particular causal graphs and the structural causal model (SCM) championed by Pearl and co-authors, we develop a causal view of the SLAM problem in order to address some of the longstanding issues. We show how causal hypotheses can be articulated, before appropriately exploiting the conditional independences exhibited by the model. As a result, the adjustment problem boils down to a constrained least squares optimization problem. The overall framework is referred as CARLIT (Causal Approach to Represent Locomotion and Internalize Topography). The conceptual ramifications are analyzed, e.g., concerning the processing of loop-closure events or the incorporation of landmarks. While toy SLAM problems are used to showcase the approach, various evaluations are also provided on synthetic datasets as well as a real-scale experimentation.

Contents

Introduction

Simultaneous Localization and Mapping (SLAM) has long been an emblematic problem in Robotics. Its immediate goal is to estimate the hidden robot pose/trajectory and landmark positions from a sequence of local and relative measurements. However, informally, the techniques employed in pursuit of that objective are not just for mapping purposes. The solutions must often be integrated into a higher order perception system that aims to improve a robot's spatial understanding w.r.t. the achievement of a task. Indeed, the task is often not limited to mapping, but can be stated in the form of, e.g., a navigation objective, a collaborative objective, a search-for demand, or any combination of the mentioned forms. While mature techniques exist for the estimation of the pose/trajectory and the local map, it nonetheless remains that the term "spatial understanding" is still open to one's interpretation; concerns can be raised about whether the estimation aspects alone are sufficient to reach "spatial understanding".

In this work, a new formulation of SLAM will ultimately be proposed to towards this difficult question.

The SLAM problem in robotics

Before detailing the contours of our contribution, a general formulation of a typical SLAM problem is given as follows: a sensory agent (e.g., a robot) moves in an unknown, known, or partially known environment. In order to keep track of its position, it makes observations/measurements of its egomotion via some of its sensor suite. No matter how precise those observations are, the resulting estimate of pose/trajectory will, by accumulation of small measurement errors, drift over time from the true motion. As the characteristics of measurement errors are usually (imperfectly) known, the growing uncertainty in the trajectory can be tracked with probabilistic tools (such as covariance matrices). Thus, without the help of some correcting mechanism, the incremental mapping of the environment would suffer from errors accumulated in the trajectory and past observations. To alleviate this issue, a correction is made when previously mapped space is re-observed or re-visited. A geometric mismatch occurs if/when the tracked pose of the agent can not explain the measurement associated with the re-observation. Under the hypothesis of proper data correspondence and no outlier effect due to incorrect detection, this should prompt a SLAM algorithm to *simultaneously* (or jointly) improve the agent trajectory and map estimates to better explain the newly uncovered mismatch, reducing by ricochet the overall uncertainty. This phenomenon is commonly termed a *loop closure*. A loop closure is thus an opportunity to improve inference, but it can also be a computational challenge in dense and/or in large scale perception problems.

SLAM has been identified and well covered since the initial robotic perception papers in the 1980s (Brooks, 1985; Chatila and Laumond, 1985; Smith and Cheeseman, 1986; Smith et al., 1988). For a time, various filtering techniques were favored. Now, standard paradigm is as follows:

- the noisy local measurements (data) participate as potentials of a joint probability distribution function (pdf) of the hidden state (trajectory and elements of the map) given the observations; those potentials (a.k.a. factors) convey geometric model uniting the hidden states (e.g., pose to landmark geometry or pose to pose evolution), together with their noise characteristics;
- the joint pdf can be represented by probabilistic graphical model techniques (e.g., factor graphs, Bayesian networks) to convey the pattern of statistical independences between variables;
- as the joint pdf is typically difficult to exploit directly, the goal of the inference process is to summarize it into the *Maximum A Posteriori* (MAP) estimate of its hidden states given the observations. Optionally, if one is ready to accept computational trade-offs, or if the mission requires it, the recovery of marginal covariances can also be an achieved;
- the MAP estimate is obtained, by maximizing posterior pdf of the hidden states conditioned on the observations. Under the assumption of additive Gaussian dynamics and measurement noise, it naturally comes as the minimum of a sum of nonlinear least squares. Efficient (often sparse) matrix computations routines can be leveraged.

A more detailed analysis will be given in chapter 2. It should be made clear that SLAM, and largely spatial understanding, is still under active research due to: (a) new sensor devices to integrate and algorithmically exploit for mapping; (b) the evolving number of robot types; (c) the large number of queries and questions a SLAM system (or perception system) is expected to answer: where are we in the environment ? Where should we go to improve mapping ? What happens if we apply a given control input to the robot ? What would have happened had the robot moved to C instead of B ? etc. However, given that the probabilistic inference part of SLAM (a.k.a., the SLAM *back-end*) has matured along numerous contributions, one would think that these questions can be answered by reusing those existing techniques, or by adding new logic *on top of* them.

Initial research question

A case in point relates to our initial research question. How can we make several robots collaborate efficiently to build a map in an unknown or partially known environment ? In the journey to tackle this question, it was realized that the accumulation of loop closures was posing greater computational challenges as the map grows, more so than with one single robot. We came to the conclusion that more needs to be done to exploit the fundamental structure of the SLAM problem, starting with a single robot system. In our opinion, some structural aspects of the problem (detailed later) have been fully exploited in current SLAM back-ends. Such aspects relate in part to unexploited pattern of independences, and in part to a more astute definition of the unknowns to be inferred. Our contribution starts from the current view of SLAM and propose a new framework to address SLAM using directed graphs and the fundamental notion of causality.

But a proper explanation of this shift requires reconsidering the problem of SLAM from the beginning, or rather, from before the invention of common tools used to solve it. In the first chapter, we investigate what we think is an apropos historical example, from the viewpoint of SLAM practitioners: the geodesic triangulations. We identify one specific geodesic operation in the late 18^{th} century. Surprisingly, the scientific challenge at the time, not only holds many similarities with the SLAM structure but also propelled fundamental theoretical developments in probability theory. To our knowledge, this historical precedent has not been investigated in the context of SLAM research. The analysis presented thereafter could be considered (wrongly) out of scope of a contribution in robotics. Yet, lessons learned from that investigation have significant implications in addressing the advocated new approach to SLAM.

Chapter 1

Pre-Robotics Adjustment: the Errors to be Feared

This chapter is mainly concerned with the analysis of adjustment problems in geodesy, for which we identify commonalities with SLAM. "SLAM is essentially a geographic surveying problem" Thrun et al. (2005, §10.5) had noted, thus studying the evolution of this practice can bring forth crucial insights to develop alternative views of SLAM. Indeed, the historical contributions made around these adjustment problems are still considered monumental. They include advances in probability theory and inferential methods, but as will be seen, do not limit to that. Through these historical lenses, and by comparison with the popular adjustment technique used in modern SLAM, we will come to question whether a new SLAM statement is worth being studied. This chapter is furthermore useful to put forward the interpretative framework for our contribution in robotics, presented next.

In section 1.1, we briefly present the context and motivations that led to the measure of the meridian arc length of the Dunkerque-Paris-Barcelona meridian (from northern France to Spain) from 1792 to 1799. We review the sources used to approach this study. Then, the experimental process of geodesic triangulation is described (section 1.2); we explain how this particular process differs from seemingly similar operations conducted prior to that period. The corresponding novel (at the time) scientific challenges raised by this problem are developed in section 1.3. We point out that new mathematical tools were designed in that regard to exploit observed data. In section 1.4, the important contributions of Delambre, Legendre, Laplace, Gauss, and others, are discussed from primary sources. An analysis is provided in section 1.5 on the far-reaching ramifications implied by those contributions on our approach. Finally, we discuss in section 1.6 what the historical analysis advocates in terms of new approaches to SLAM.

The events are presented chronologically while keeping track of similarities with modern SLAM problems. In a trade-off between clarity and historical rigor, we sometimes use anachronistic terms like Gaussian, normal distribution, Bayes law, central limit theorem even though they acquired those names much later.

1.1 The Broader Context in 1792-1799

During the 1789 political turmoil occurring in the French Kingdom, which marks the beginning of the *Révolution* period, a popular and recurring demand was for the authorities to end the uncontrolled proliferation of weight and measure systems. These systems were arbitrarily defined by local seigniory. They varied not only city to city, but also from a corporation to another. Sometimes they held the same names but had different values (Guedj, 2000). The resulting general confusion represented an obvious hindrance to trade, and was cause for fraud. Previous attempts to standardize the weights and measures failed due to political inertia.

Between 1789 and 1792, several "Weights and Measures" commissions of the *Académie Royale des Sciences* (Academy of Sciences) were formed to lay the groundwork for the definition of a new coherent system. The aim was to dramatically rationalize the issue, and, hopefully, to also export abroad the ideas of the new system. The end product would be known as the *Metric System*. Some of the members of these commissions are distinguished names, often mentioned in modern scientific activities: Laplace (1749-1827), Lagrange (1736-1813), Coulomb (1736-1806), Legendre (1752-1833), Lavoisier (1743-1794) etc.

In order to maximize the chances that the new system be adopted nationwide (and perhaps internationally), and to unsure that it would not easily be replaced, it was decided to define the length value of 1 meter using an objective measurement rather than an arbitrary one. Basing the *meter* as a fraction of the earth meridian[1] was after debates the final choice[2]. However, no reliable value for the meridian arc-length was available. Indeed, using an existing, but unreliable, estimate of the meridian length would put the meter at risk of significant update in its value in the future, hence threatening the much desired stability of this new weights and measures system. On the other hand, conducting new measurements of the meridian, with sufficiently small bounds around the uncertainty of the result, would improve the chances of success of the metric system and make it future-proof.

The measurement of the meridian arc length was covered independently in two books published by historians Denis Guedj and Kenneth Alder. Their work mostly overlap: Guedj (Guedj, 2000) describes in more details the initial debates to motivate the choice of meridian as a basis for the new system, while Alder (Alder, 2003) puts more emphasis on the subsequent change of paradigm in scholars, in particular concerning the processing of observation errors. The extension of the arc meridian a few years later was addressed by Pierre Bayart (Bayart, 2007). Historian of statistics Stephen Stigler primarily focuses on the development of the theory from early 17^{th} century to 1900 (Stigler, 1986, 2002). He goes into great details into the advent of the probabilistic concepts, the early struggles, the collaboration/competition of minds. His work has been essential to understand and to refine where the problem of the arc length meridian fits into this bigger picture. Former geodesist Levallois (Levallois, 1988) provides an analysis of the practices of French geodesy over three centuries. The successes and failures of adjustment operations carries, in our assessment, valuable lessons for SLAM. In her PhD thesis, Jozeau (1997) investigates the difference in evolution between French and German practice of geodesy in 19^{th} century. Notably, the difference in the significance given to the method of least squares is addressed. As one might expect, the historians of the

[1] A meter was defined such that the arc-length of the quarter of a meridian be 10 thousands kilometers.

[2] Detailing the reasons and the back and forth discussions leading to that choice (over e.g., the pendulum) is interesting but would be an unnecessary digression in this document. We refer the reader to the works of historians such as D. Guedj or K. Alder (Guedj, 2000; Alder, 2003).

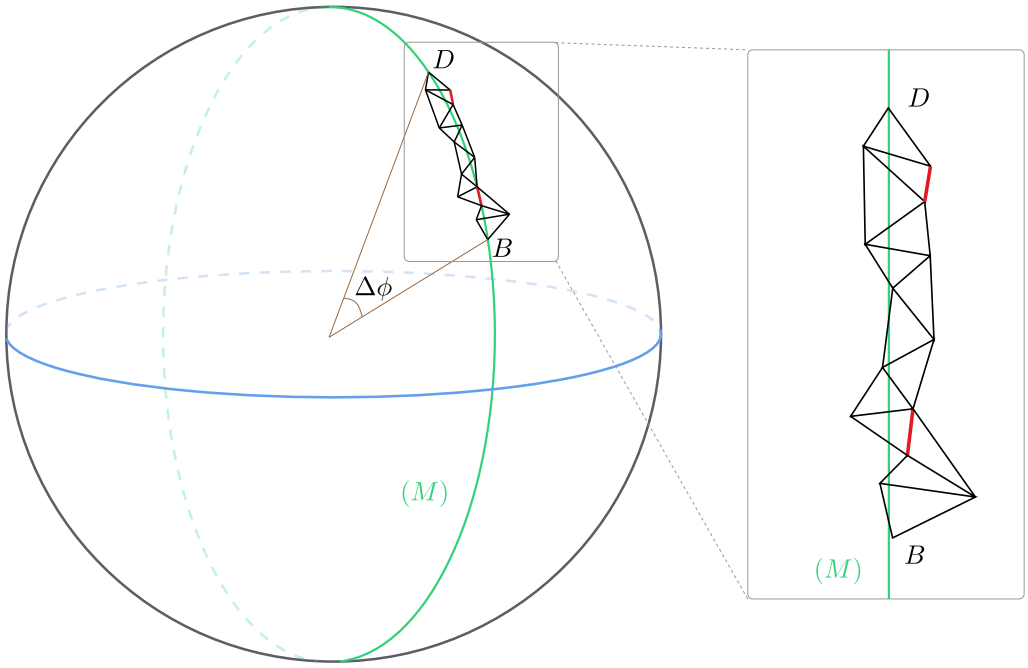


Figure 1.1: Sketch of a simplified arc-length measurement between points D and B. The distance between these points is deduced from the angular measurements in the chain of triangles and the length measurements of the baselines (red). The full meridian arc-length is then extrapolated thanks to latitude measurements, yielding $\Delta\phi$. The fitting parameters of ellipse (M) can then be deduced from $\Delta\phi$ and the chain of triangles.

aforementioned period previously are not invested into the intricacies of the SLAM problem (to our knowledge). Thus, in order to push our investigation further and to produce this synthesis, their accounts have been supplemented by a direct study of the source material written by scholars such as, in no particular order, Delambre, Méchain, Legendre, Laplace, Arago, Gauss (we used Bertrand (1855) translation from Latin to French), La Condamine, Puissant, Tardi, etc. Most sources are available online via official documentation (links given in the references), although often only in French.

1.2 Geodesic Triangulation

Figure 1.1 shows a simplified example of the measurement of a portion of a meridian at the surface of the earth. Since measuring the length of the full meridian ((M) in fig 1.1) is impractical, the standard practice was to measure only a limited portion of the arc-length (Dunkerque-Paris-Barcelona). Then, thanks to latitude observations, the length can be straightforwardly extrapolated to ninety degrees. Despite the limited portion considered, measuring the Dunkerque-Barcelona distance with rulers put together back-to-back would still be impractical. Thus, the well-known process of triangulation was employed. This process was invented by W. Snellius back in 1617.

The concept of geodesic triangulation is fairly simple: instead of measuring directly the distance, the relative angles between landmarks are measured. These landmarks (chosen

among mountain tops, bell towers...) form a chain of triangles between D and B (fig 1.1). Only one triangle side in the chain, the so-called *baseline* (in red), must be measured directly on site, with precise rulers and lots of care. Then, the remaining of the spatial information can easily be deduced from basic trigonometry, i.e., the law of sinus. A second *baseline* is commonly measured at the end of the chain of triangles (also in red). The motivation for the measurement of second baseline is to *verify* if the deduced length from the first baseline *plus* the angular measurements agree with the direct measure on-site. Of interest to us is how close these meridian operations match the pattern of a typical SLAM problem. As mentioned before, the principles of triangulation for geodesy were not new at the time. But this particular problem has specificities that compelled scientists to analyze the results in the following years, and to come up with novel concepts.

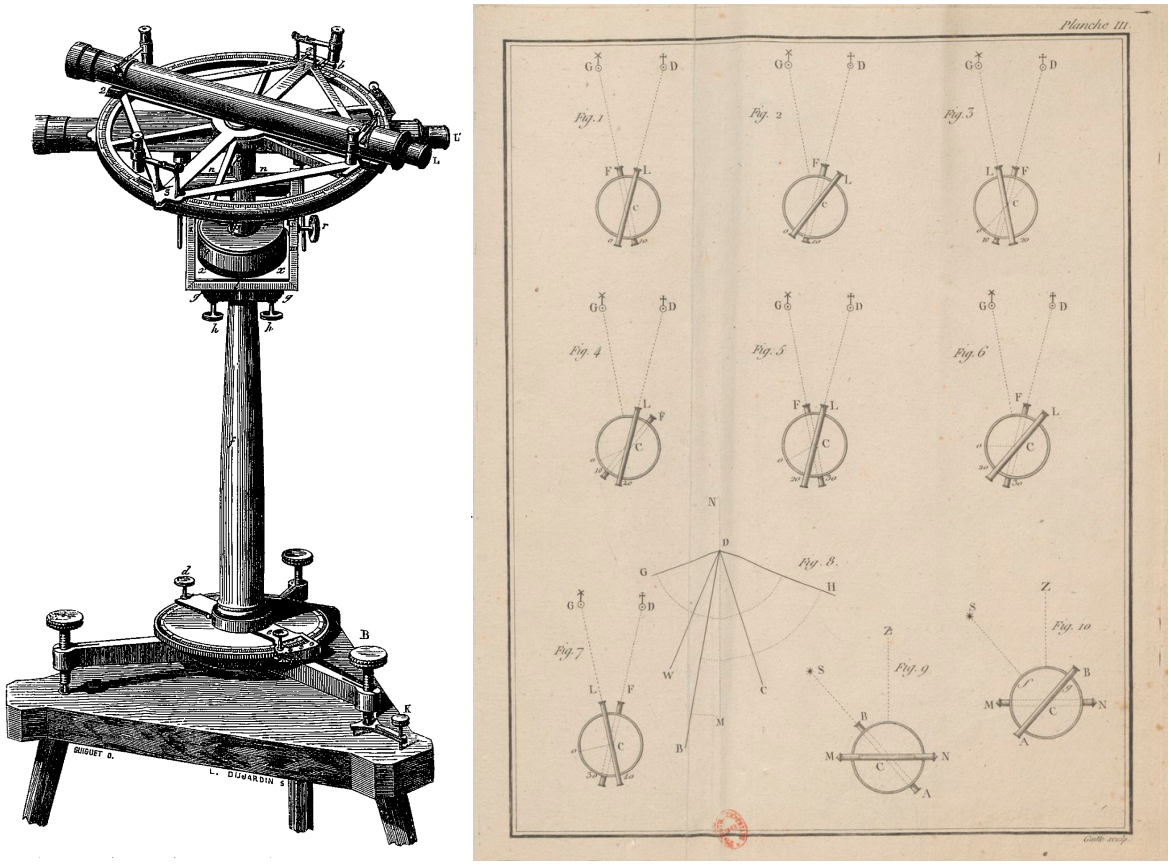


Figure 1.2: Borda's repeating circle. On the left, the perspective view of the instrument. Reproduced from (Arago, 1857, fig 250). The schematics on the right are reproduced from Cassini et al. (1787), and show the principle of the repeating procedure.

Those main specificities of this operation, compared to previous ones, can be distinguished in two aspects. Firstly, the *repeating circle*, designed by experimental physicist J.-C. Borda (1733-1799), was a new instrument suited to the measurement of relative angles between two distant objects, see figure 1.2 and box 1. The "repeating" aspect of the circle comes from the fact that a sequence of sightings could be repeated an arbitrary number of times, though reading only once the sum of those angles on the graduated circle. This meant that, for a skilled user, the expected measurement error could be one to two orders of magnitude lesser

Box 1: Borda's repeating circle

The "repeating circle" is an instrument, based on a concept from Tobias Mayer, designed by Borda and crafted by Lenoir. The geodesic instrument was used to measure relative angles and conduct astronomical observations. The first recorded use of the instrument is from 1787 to join the Paris and Greenwich meridians (Cassini et al., 1787). The innovation of this instrument comes from the fact that an arbitrary number of measurements could be made without returning the scopes to the origin of the scale (Alder, 2003, chap 2). The ensuing angular readings, divided by the number of measurements, was the mean of all those measurements. As we will see in section 1.4.2, Laplace would later identify an application of the central limit theorem (proven in 1810), hence legitimizing the usage of the normal distribution to describe the error curves associated with the angular value.

than with former instruments (Arago, 1857). The hardware could also be configured to conduct astronomical observations (e.g., to determine the astronomical latitude). A description of its manipulation process is available in (Alder, 2003, fig 7). Secondly, the requirements in terms of rigor, precision and transparency were new for this kind geodesic operations. The goal was to inspire confidence in the metric system (Guedj, 2000; Alder, 2003). This could only be achieved if sufficient trust could be placed in the results of the operations. Despite rigorous process, measurement errors were ultimately unavoidable; it was thus necessary to argue that the ultimate error on the meridian arc length was likely bounded to an acceptable interval. Some fundamental tools in probability theory were unfortunately not yet conceived (but as we will see, they were around the corner).

Astronomers J.-B. Delambre (1749-1822) and P. Méchain (1744-1804), considered as excellent observers, were entrusted to conduct those operations starting from 1792, aided by the mathematician A.-M. Legendre who would help to post-process the data. Delambre and Méchain were equipped with one Borda repeating circle each. Their task would turn out to be extremely challenging due to the conjecture[3]. Eventually, they completed their operations in 1799, i.e., seven years later. The number of triangles forming the chain between Dunkerque and Barcelona amounted to 115 (see figures 1.3 and 1.4). Astronomical latitudes [4] were carefully measured at 5 points along the meridian, in north to south order: Dunkerque, Paris, Evaux, Carcassone, Barcelona. The results were verified the same year (1799) by an international commission, and the value of the unit-meter established[5]. Remarkably, the mismatch between the measured length of second baseline, and its value deduced from the first baseline plus the chain of triangles, was only 31 cm. To get a sense of the proportion, the first and second baselines are approximately 12 km in size (Vincent, 1998; Valette, 2007) and are

[3]Due to the proximity of conflict with other Europeans powers, Delambre and Méchain were in some instances suspected by local authority to be spies/royalists; their budget was given in hyper-inflated currency; Delambre was temporarily fired from the job for supporting Lavoisier (who was guillotined); Méchain was in a coma following an accident; villagers were superstitious about their strange instruments. See Alder (2003); Guedj (2000) books for more.

[4]Astronomical latitude were computed by observing the declination of the stars w.r.t. the local horizontal plane. This local horizontal plan was defined by a pending lead ball attached to the repeating circle via a wire. Thus, any irregularity in the local gravitational field affected the result (it was akin to today's accelerometer bias).

[5]The "correct" value of 1 meter is off by 0.2 *millimeters* from the definition established then.

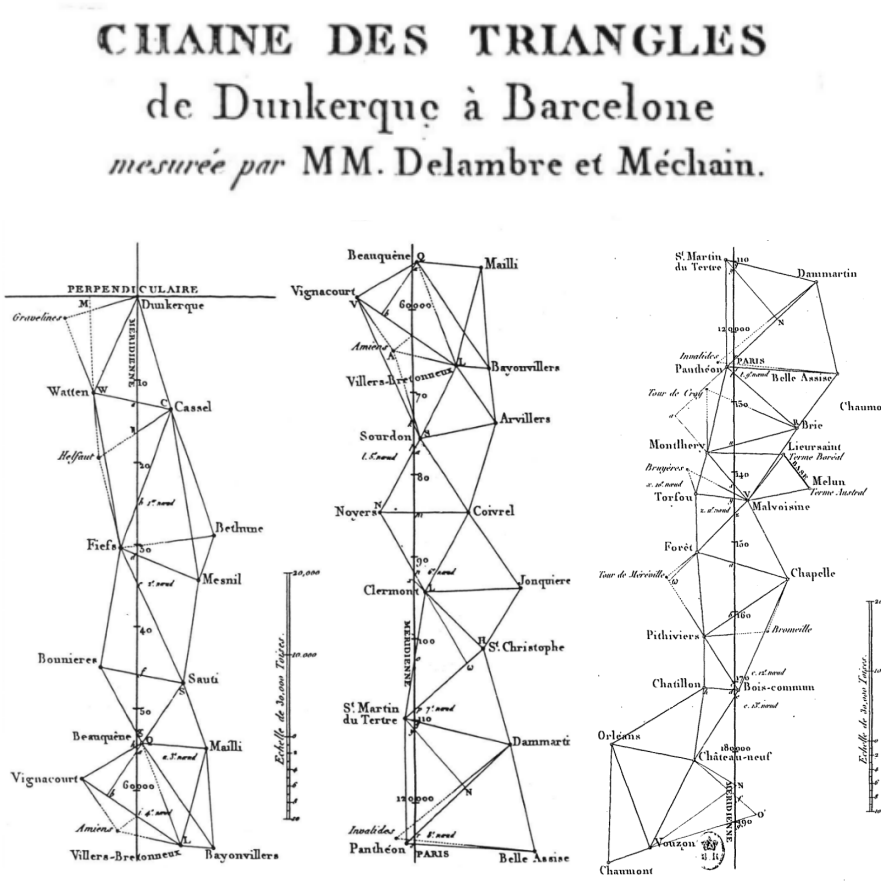


Figure 1.3: Northernmost portion of the chain of triangles, between Dunkerque and Paris. Reproduced and assembled from Delambre (1810, p706-7). The measurements related to this northern portion of the meridian were done by Delambre and his team. Their job was made difficult due to an unstable political situation. The first baseline was located between Melun and Lieusaint (rightmost schematics). Contemporary GPS-aided investigations have estimated the error on the first baseline to be just under 12 *centimeters* on a 11 607*meters* length (Vincent, 1998).

located more than 650 *km* apart. However, inconsistencies in the data of the astronomical latitudes would slightly shorten the initial value of the meter[6]. The latitudes were indeed necessary to estimate the portion of the meridian covered by the chain of triangles. In other words, the angle $\Delta\phi$ in figure 1.1 was compromised.

[6]The potential causes of astronomical latitude inconsistencies were known to be: imprecision of the observers, the irregular form of earth (even at sea levels) and local irregularities of the gravitational field (a.k.a. the geoid). A recent study in the journal of Geodesy, by Vaníček and Foroughi (2019), picked interest in the issue and concluded that 95% of the error can be attributed to the bias in gravitational field, 3% to the form of earth and only 2% to the imprecision of the observers.

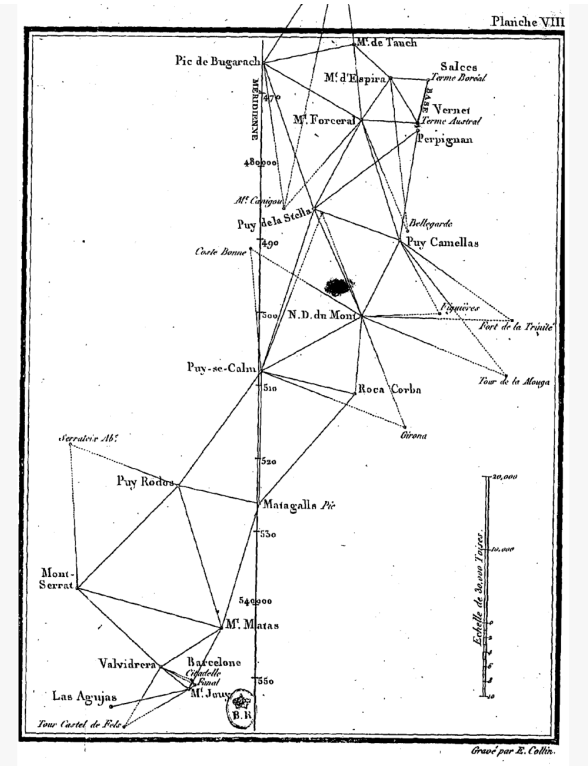


Figure 1.4: Southernmost portion of the chain of triangles, between Perpignan (France) and Barcelona (Spain). The measurements related to this portion of the meridian were done by Méchain and his team. The second baseline was located between Perpignan and Salses (top right part of the figure). The task of Méchain and his team was made difficult in part due to an untimely conflict between both countries, while working on the triangles in the border area (Alder, 2003, chap 3). Above all encountered difficulties, Méchain was tormented by unexpectedly large inconsistencies in the determination of astronomical latitudes in Barcelona. He would return to the area in 1803 to extend the chain of triangles to the Balearic Islands where, unfortunately, he would perish from illness.

1.3 From the meridian to the theory of errors

In support to the results of their geodesic operations, Delambre and Legendre (the geometer, in charge of analysis) jointly wrote a report detailing their method (Delambre and Legendre, 1799). The specificities of this operation, that we mentioned previously, lead the authors to criticize previous approaches and to explain why the processing of the 1792-1799 meridian operations required new analytical tools (our translation):

> Those methods [employed in previous geodesic operations] were purely approximate. They may have seemed sufficient w.r.t. the precision of the instruments then used. The errors induced from calculations were usually far smaller than the ones induced from observations, and it was thus legitimate, at the time, to not factor in too much calculations correctness that would only be illusory.

The previous geodesic operations Delambre and Legendre referred to are ones conducted fifty years early (mid-18th) by several expeditions in Peru (at the equator), Lapland (Finland) and South-Africa. The objective of each expedition was to measure the arc length of their

meridian, at their respective latitude. Then, by comparing those arc length, a scientific controversy would be solved: is the shape of earth is more flattened at the poles, according to Newton's theory, or at the equator, as suggested by the Cassini cartographers ?

But this was only a binary question and thus less rigor in the analysis was expected at the time. Each scientist had different approaches to deal with errors. Maupertuis, after the 1736 Lapland expedition, calculated "his" arc-length covered by 9 triangles and 1 baseline. He did so via 6 unspecified methods and chose as a final value the arithmetic mean between 2 of them (Levallois, 1988, p34). In the larger 1735 Peru/Ecuador expedition, a chain of 34 triangles and 3 baselines was measured. A dispute erupted concerning the processing of errors between the lead scientists. One of them, Bouguer, a mathematician, refrained at first from adjusting the measured values. He considered this practice cheating (Ycart, 2001), but eventually corrected his angles according to vague considerations (Jozeau, 1997, p33). In some sense, for many, adjusting observations a posteriori was a secret sauce, only permitted so long as it did not alter the conclusions of their research question. La Condamine, a more expedient man, argued in favor of taking the arithmetic mean (a.k.a. Cotes rules since 1722) of several results. Contrary to many others, he openly wrote about his approach in his report Condamine (1745, art25-26 p455-63). La Condamine and Bouguer split ways and went back to France separately, where their dispute would continue. Gradually, the notion that corrections should be applied started to be accepted in later part of the 18th century. The research of best manner for combining of equations of observations would later be called by Gauss *one of the most important problems of natural philosophy* (Bertrand, 1855, preface).

Boscovich, a Dalmatian Jesuit scientist, proposed in 1755 that the sum of corrections applied after a geodesic operation should be minimal in absolute value (Stigler, 1986, chap 4). The idea was to set a criterion so that to which the corrections $c_0, c_1, \ldots, c_n$ should minimize $\sum_{i=0}^{n} |c_i|$. Laplace, in the 1780s, tried to improve on the idea (only given in textual form by Boscovich) by incorporating the notion of weights for each observation. The notion of weights was a surrogate for how to quantify the precision of measurements. He attempted to use probabilities but was also unsuccessful in expressing a general methodology.

But for the purpose of the metric system, none of these approaches seemed satisfactory for Delambre and Legendre (1799). They continue from previous quote:

> Novel instruments, by providing us a far greater precision in angular measurements, compelled us to research more rigorous and exact methods for our calculations.
>
> Delambre & Legendre

What is referred by *more rigorous and exact methods* ? In the document, Delambre and Legendre would provide novel methods related to spherical trigonometry and also methods to account for height differences between geodesic stations (triangle vertices in figures 1.1, 1.3 and 1.4). One would hope, by reading the quote, that a system for managing uncertainty should be detailed. However, a formal language for dealing with the randomness of errors was not yet available. As such, some arbitrary decisions to combine unavoidable mismatches (even though they were small in amplitude) between measurements were made. For instance, any dimension on the northern portion of the chain of triangles was tied to the first *baseline* (near Paris), while the remainder were tied to the second *baseline* (near Perpignan) (Delambre, 1810). The motive will be addressed in section 1.4. Without the probability theory framework, it was not

guaranteed that the result stemming from this arbitrary inference method would be the most probable, nor could the result be attached with a quantifiable confidence interval (what we today know as variance, or standard deviation). This was problematic given the metrological objectives of the metric system.

Relevantly, what we recognize today as major names in early probability theory developed a strong interest in the problem. Firstly, Laplace was actually a member of that same 1799 commission which verified the arc meridian results and established the value of the meter (Delambre, 1806, p94). Secondly, leading German mathematician and astronomer C.F. Gauss (1777-1855) had access to the dataset (which was distributed internationally) the same year. He would also be interested in geodesic problems later. It is historian K. Alder's case that the operations of the arc meridian paved the way to the formalization of the *theory of errors*, precursor of statistics (Alder, 2003, chap 11). Alder observed that P. Méchain was tormented by his measurement errors. Méchain considered their manifestation as the expression of a moral fault by the observer (lack of skill, careless manipulation of instruments, etc.). As the commonly used expression at the time proclaimed, the errors were *to be feared* ! Others, like his colleague Delambre, started to consider the errors unavoidable, not necessarily the fault of the observer but that they were partly due to ignorance of natural phenomena (form of the earth, atmospheric refraction, irregular gravitational field, wear of instrument). He was correct in that assessment, see footnote 6. It can only be surmised that, for Delambre, as long as the errors were managed transparently, rigorously, the result of an experiment affected by heterogeneous uncertainties (e.g., uncertainty in angular measurement, base length measurement, etc.) could astutely be most likely confined within acceptable bounds.

In that spirit, Delambre would write *Base du système métrique décimal*, "Basis for the metric decimal system" (*Basis* for short), a 2400-page report that aimed to reproduce relevant documents and address all questions that could arise in relation to the definition of the metric system. The third and last tome would only be achieved only in 1810 (Delambre, 1806, 1807, 1810).

Not waiting for the full justification, the value of the unit meter was set according to the conclusion of the 1799 commission. These developments were followed outside of France. The German journal *Allgemeine Geographische Ephemeriden* published a summary of the dataset which was proofread by renown mathematician C.F. Gauss in the summer of 1799. Later that year, back in France, Laplace was named Minister of the Interior (equiv. of Home Secretary) with the task to establish the metric system in the French society[7].

1.3.1 Legendre's new data fitting technique: the least squares

Concerning Legendre, the 1799 quote reproduced above seems premonitory. In 1805, he published the famous least squares method, future workhorse of statistics[8]. Though the publication title refers to the determination of the trajectory of comets, the least squares method is only presented in appendix and its application relates to the arc meridian dataset (Legendre, 1805, p72-80). Legendre processed the distances (deduced from triangulations and treated as error-free constants) between the 5 latitude stations (from north to south: Dunkerque, Paris, Evaux, Carcassone, Barcelona) to infer the flattening coefficient of an elliptic earth model (figure 1.5).

[7] After a few weeks, he was expelled from that position. For Napoléon, Laplace *sought subtleties everywhere, conceived only problems, and finally carried the spirit of infinitesimals into the administration.*

[8] And, we may add: a popular (if not the most popular) method for modern SLAM solvers.

The ensuing minimization gave him a value of 1/148, which seemed suspiciously low compared to the consensual flattening coefficient of around 1/320 to 1/334. By substituting the inferred value into the equations, he noted that this value would imply large posterior errors (residuals) on some of the astronomical latitude measurements. Those errors were judged too large to rest only the observer's inaccuracies. He postulated that gravitational irregularities were mostly to be blamed, for they had affected the leveling to the horizontal of the repeating circle (see again Vaníček and Foroughi (2019) for recent confirmation). This phenomenon is also known as the deflection of the vertical in geodesy (Torge, 2023; Hofmann-Wellenhof and Moritz, 2006). In the last sentence, remarkably, the mathematician concluded that a pendulum experiment would have been a more proper way to define the value of the meter, thus discrediting the whole project of his colleagues to define the unit meter based on a geodesic operation.

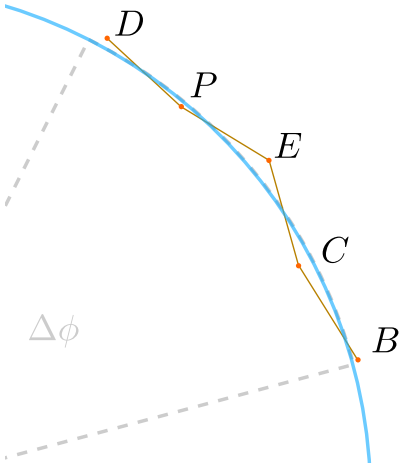


Figure 1.5: Legendre presented the method of the least squares of errors on the arc meridian dataset that was used to define the metric system.

1.3.2 Gauss' probabilistic view of least-squares and the simultaneous corrections

Legendre noted that the least squares method is very convenient for data fitting. Indeed, so far no method for the combination of "different observational equations" could be considered general when there were more equations than unknowns (Stigler, 1986). Ad-hoc, case-by-case procedures were the norm. The method of least squares was yet another ad-hoc construction, but a straightforward one. Hence, this was a significant step, but no link was made so far with probabilities. Gauss (1809) would later contest Legendre's priority on this discovery by claiming that he already used the (unnamed) least squares routinely, before 1805 and had communicated it to three other astronomers.

Remark 1.1. *On the Gauss-Legendre dispute.*

Though Gauss claim has merit for good reasons, no record has been found establishing unambiguously that he used this method before Legendre. Interestingly, there are debates that Gauss used a form of least squares in 1799 on precisely the same arc meridian dataset ! We mentioned previously that Gauss had read an issue of the Allgemeine Geographische Ephemeriden*. He detected a printer's error and sent a small letter to the editor, proposing new values for earth eccentricity and the arc length. He did not elaborate on the nature of the method he used. Stigler (1981); Celmiņš (1998) investigated whether this method was the least squares. The eccentricity value he got, of* 1/187*, differs from Legendre estimate of* 1/148*. Legendre, even if one supposes that he discovered the method years after Gauss, caught on the value of the method and considered it worthy of public communication. This is the conclusion of Stigler (1981): "If there was any single scientist who first put the method within the reach of the common man, it was Legendre, in 1805". Gauss should rather be praised for his in-depth, probabilistic analysis.*

In any case, by 1809, the scientific impact of the least squares method was considered worth of a priority claim by the involved actors.

But aside from the dispute, Gauss would provide in 1809 an essential interpretation of the least squares method by framing it as a tool for probabilistic inference, in particular in conjunction with the normal distribution (a.k.a. the *Gaussian* distribution).

His idea, now well-known[9], consisted in wrapping the *equations of conditions* into error curves $\varphi(.)$ functions (a.k.a. today as the likelihood functions), which were axiomatically of Gaussian form, and consider their product.

Denoting $V, V', V'', \dots$ functions[10] of the sought-for parameters, and $M, M', M'', \dots$ the observations, he postulated that the product:

$$\varphi(V - M)\varphi(V' - M')\varphi(V'' - M'')\cdots = \Omega, \tag{1.1}$$

should be maximized. In Gauss words (Davis, 1857a; Gauss, 1809; Bertrand, 1855):

> This product expresses the expectation or probability that all these values will be produced simultaneously [*simul* in Latin] from the observations.
>
> C. F. Gauss

Applying the negative log on the product of Gaussians, the problem translated into the minimization of the sum of squares. He analyzed correctly that the h constant in the form $e^{-h^2\Delta^2}$ can be seen as a measure of the precision of the observations (Bertrand, 1855; Gauss, 1809, p119,§4), and used an "elegant theorem by Laplace" for the closed-form value of the integral

$$\int_{-\infty}^{\infty} e^{-h^2\Delta^2} d\Delta = \frac{\sqrt{\pi}}{h}. \tag{1.2}$$

Eventually, he stated that his method ("principium nostrum", i.e., "our principle", which infuriated Legendre) is more convenient, even if one does not accept the axioms leading to the probabilistic framework. From a calculation standpoint, it is still more economical than minimizing the sum of power of 4, or 6, etc. In particular, Gauss criticizes Boscovich's and Laplace's approaches (Bertrand, 1855; Gauss, 1809, p133, §12), which minimized the sum of absolute values, what amounts to ignore some observations to the benefit of others. This argument is made clear in Noël and Tilleuil (2005) analysis.

In our view, these papers about the Legendre-Gauss priority dispute and the quote of Delambre and Legendre (1799) we unearthed, seem to reinforce further Alder's point that the least squares method may be as much the byproduct of the arc meridian problem as the (independent) discoveries of Legendre and Gauss[11]: "Both men worked on the same geodesic problem. [...] Both read the same authors, in particular Laplace." (Alder, 2003, chap11). A third less known mathematician, Robert Adrain in the United States, also discovered in 1808

[9] Today, in SLAM, the Gaussian characterization of uncertainty is dominant. For instance, in Davison and Ortiz (2022) words, "Almost all serious, scalable probabilistic estimation is based on the core assumption of Gaussianity in 'most' measurement distributions and 'most' posterior variable distributions, 'most' of the time". Furthermore, eq (1.1) is the well-known factored joint distribution of observations, which can be represented in a factor graph.

[10] The discrimination between variables and functions is sometimes tricky in the old papers.

[11] This should not be understood as taking away the fact that astronomy, as a discipline, played also a driving role. However, it is already a well known fact that those renown scientists were invested in astronomy. On the other hand, the geodesic aspect tends to be less known and holds many similarities to SLAM (see next section).

(thus, officially one year before Gauss) how to link Legendre's method (since that is the one he had read), with the normal distribution (Dutka, 1990, 1995). Adrain was a surveyor, so his problem was also one of adjustment.

In the interest of SLAM research, we extend Alder's point. By pointing out the differences and similarities of the arc meridian with a typical contemporary SLAM problem, we also wish to highlight the role of this problem in further developments of probability theory. The algorithmic treatment to correct the measurement errors in the chain of triangles around the arc meridian reveals, in our view, patterns of thinking worthy of the consideration of SLAM experts. It is henceforth investigated in the next section. We will continue to analyze important 1820s contributions of Gauss further in this chapter.

1.4 Geodetic adjustment via the theory of errors

By 'geodetic adjustment', we designate the process by which geometric corrections are applied to the chain of triangles after accounting for any mismatch between some measurements and their predicted values obtained by calculations from other measurements. Our objective is now to investigate the patterns involved in geodesic adjustment so that in the next sections it is analyzed against the SLAM paradigm.

First, it should be explained why we only consider geodetic adjustment from the period of the exploitation of this arc meridian dataset, and later. As mentioned, Snellius invented the process of geodesic triangulations in the 17th century, and there were previous expeditions to measure degrees of meridian in the mid 18^{th} century (section 1.3). In that regard, let us expand on some points addressed the previous section:

- New instruments were available. Namely Borda's repeating circles. Observers Delambre and Méchain could achieve greater precision in the individual angular measurements by averaging an arbitrary large number of observations.

- Delambre noted that the process of correction of triangles in previous operations was often not explained. In the absence of least squares and an established probability theory, both the management of errors and the combinations of equations of observation were vague concepts. Geometrically, in order to build the chain of triangles from the baseline, only two angular measurements were necessary per triangle. The rest could be deduced using basic trigonometry. But one could make a third angular measurement to apply some corrections so that the sum of triangles equals π. This was already known before. The issue is that, without proper concepts, the corrections were done differently one from analyst to the other.

- The second baseline (sometimes third, fourth, etc.) was primarily used to verify, not *correct* previous values. The second baseline length could be predicted from the recursive application of trigonometry starting from the first baseline. If that predicted value was not too different from the measured length on-site, the verification was successful. The main characteristic of the second baseline was to offer reflection on the skill of the observer(s). However, it was not yet realized that the length mismatch (between prediction and actual measurement) could be systematically exploited to correct the chain[12].

[12]Or rather, if it was realized, no argument, no favored calculus existed to apply the corrections. Delambre

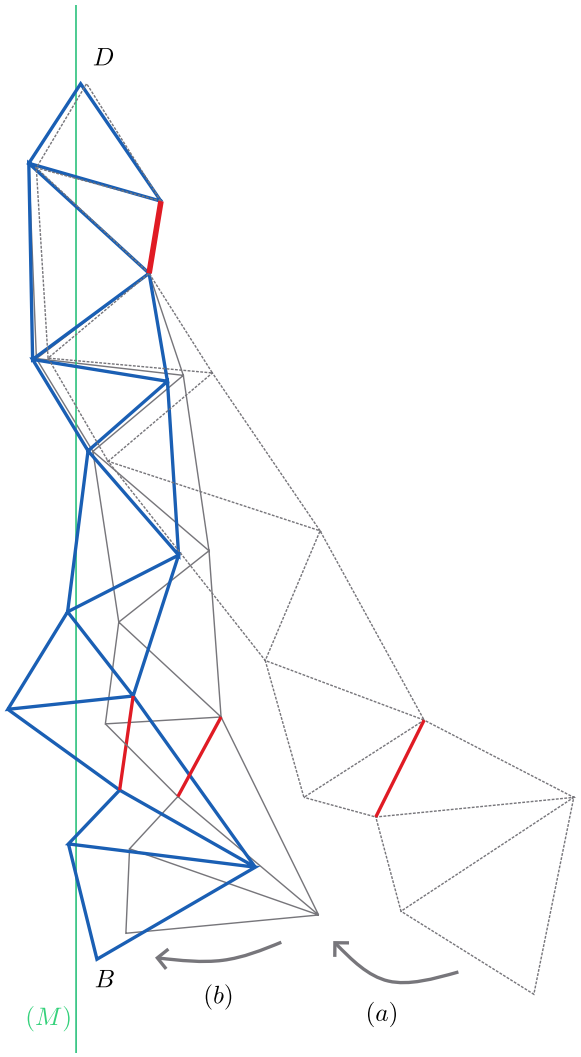


Figure 1.6: Made-up geodetic adjustment of the simplified arc meridian problem of fig 1.1. Dash Gray (no correction), solid gray (partial corrections), and blue chains (all corrections) are different phases of the geodetic adjustment problem. Baselines are in red. Step (a) shows putative corrections following the closure of triangles. Step (b) shows putative corrections following closure of the second baseline.

- The requirements in terms of experimental rigor in the 1792-99 geodetic operations were so far unmatched. There was no confidence that the results of previous operations did fall within acceptable range to satisfy the definition of the meter. Having accessed the archives, Alder (2003) describes how Delambre kept consistent logs of the observations he made in books with numbered pages, dated and with signatures from him and his assistants. Observations values were regularly communicated back to the Académie des Sciences (or, from 1795 onwards, to the Bureau des Longitudes) in Paris so as to prevent any unstated alterations (Delambre, 1810, p698-704).

We will use the terms *triangle closure* and *baseline closure* to refer to the two types of corrections. Those terms are chosen on purpose, as they resemble to the well-known SLAM phenomenon of *loop closures* that is henceforth investigated.

To support our arguments, we will make use of the simplified problem of the chain of triangles presented figure 1.1. Figure 1.6 shows a made-up (exaggerated) geodesic adjustment.

mentioned a correction of 1 angular value made by LaCaille in 1740 to improve the baselines mismatch in the Dunkerque-Colioure arc meridian (Delambre, 1807): but as no explanation on the whys and hows of the process, that seems more arbitrary than thought through.

In essence, we found out that the follow-up works on subject by Delambre, Laplace, Gauss (and colleagues) are remarkable and deserve to be known and interpreted in the context of SLAM. To our best knowledge, most of the work in French, except Laplace's, is not translated in English. Some Latin texts of Gauss have translated in French by Bertrand (1855) and English by Davis (1857b). Nonetheless, their writings are easily available online.

Before the next subsection §1.4.1, let us summarize the chronology of the events in the late 18th and early 19th century.

First, Delambre and Méchain geodesic on-site operations occurred between 1792 and 1799. Among the leading mathematicians, Laplace, Legendre and Gauss are provably involved in the problem in 1799. The first two were members of the special 1799 commission, while Gauss, not involved directly, nonetheless got access to the dataset (Delambre and Legendre, 1799; Stigler, 1981). The contribution of Méchain stopped in 1804 as he would try to extend the survey from the south of Barcelona to the Balearic Islands, perishing from illness. Legendre communicated the least-squares method known in 1805. Two other observers, J.-B. Biot (1774-1862) and F. Arago (1786-1853), were sent in Spain in 1806 to take over Méchain's project (Bayart, 2007; Arago, 1854). Gauss, in 1809, deduced the normal distribution from specifications (Noël and Tilleuil, 2005; Stahl, 2006), and tied it to the least squares method: whenever the errors distribute normally, the most probable results are the minimizers of the sum of squares. We can start to see and appreciate how the geodetic adjustment was one of the fundamental problems that accompanied closely, if not motivated, major contributions in probability theory and inference techniques.

1.4.1 Delambre's view: harmony should be restored

When Delambre released *Basis* (Delambre, 1806, 1807, 1810), he addressed all operations, calculations and choices made in relation to the design of the metric system. Among the addressed topics, our interest concerns the adjustment of the chain of triangles owing to the second baseline measurement, i.e., the baseline closure. We mentioned that the difference between the measured length on the second baseline and the length calculated from the chain of the triangles (spanning 650 kilometers) was less than a third of a meter (on an approximately 11.7 kilometers portion). This is a small error in amplitude but the 1799 weight and measure commission still felt compelled to discuss it and imagine the corrections that could be made from that. We present the initial discussion of the pattern and then show how Delambre would later improve it.

In the third tome of *Basis*, Delambre (1810, p417) reproduce the document dated 1799 (fig 1.7), where the members of 1799 commission (Van-Swinden *et al*) initially addressed the cause of the mismatch (our translation):

> The verification baseline, measured near Perpignan, is linked to the one of Melun by a chain of 53 triangles. We can therefore deduce its length by [trigonometric] calculations, and then compare this length to the measured one. [...] One just has to know [geodesic] operations of this kind to be convinced of the following: the difference of the deduced length of the verification baseline [the second baseline] from its actual measurement, purely comes from the slight unavoidable [angular measurement] errors in each triangle, that can, in a chain as expanded as this one, either compensate or accumulate themselves.
>
> Van-Swinden, Trallès, Laplace, Legendre, Méchain, Delambre, Ciscar

RAPPORT

Sur la détermination de la grandeur de l'arc du méridien compris entre les parallèles de Dunkerque et Barcelone, et sur la longueur du mètre qu'on en déduit.

Fait à la commission des poids et mesures le 11 floréal an 7.

La base de Perpignan, déduite par une chaîne de cinquante-trois triangles de celle de Melun réduite en arc, se trouve, par le calcul, de . . . . . 6006·089 demi-modules.

La base de Perpignan réduite en arc, se trouve, par la mesure actuelle, de . . . . . 6006·249

La différence est de . . . . . 0·160

Fait au Palais national des sciences et des arts, le 11 floréal an 7 de la République française, tous les membres de la commission spéciale étant présens.

Signé, J.-H. Van-Swinden, Trallès, Laplace, Legendre, Méchain, Delambre, Ciscar.

Lu à la commission générale ledit jour, et approuvé par elle.

Signé, Lagrange, Pedrayes, Fabbroni, Coulomb, Vassalli, Mascheroni, Multedo, H. Aeneae, Lefèvre-Gineau, Darcet.

Figure 1.7: *Report on the determination of the meridian arc length between the parallels of Dunkerque and Barcelona, and on the deduced length of the meter*, by the members of special commission of weights and measures, is a significant document in the history of the metric system. Extracts are reproduced from *Basis* (Delambre, 1810, p415,418,433), Gallica - Bibliothèque Nationale de France. The first definitive value of the unit meter is given. The topic dealing with baseline closure is also first addressed in this document. The authors and the field observers conceded that errors are unavoidable, but did not propose an adjustment method to remove the inconsistency.

The authors tone in the full paragraph appears defensive, and argues in favor of not measuring again the baselines. It does so by pointing out that the difference is small in amplitude, and by explaining the causes of the mismatch. A baseline length measurement was a forty-five to fifty-day long tedious and delicate operation. Whenever the suspicion emerged of an external event affecting the stillness of the rulers, Delambre decided to throw away all measurements from that work-day, and start again the next day from the point of a previous valid day[13]. GPS measurements conducted in 1998 have found that the first baseline length differs approximately 12 centimeters from Delambre's measurement (a 10^{-5} relative difference) (Vincent, 1998). Having directed both angular measurements and baseline length measurements, Delambre and Méchain, using their appreciation of the relative accuracy of different types of measurements, concluded with the special commission that the errors did not come from their baseline length measurements, but rather from the accumulation of small errors in the chain of 53 triangles *between* the first and second baseline. More subtly, remark that the angular errors in the remainder triangles, which are as much unavoidable as the ones in the 53 intermediary triangles, had no effect in this argument. Indeed, recall that there were between 107 and 115 triangles in the full Dunkerque-Barcelona chain (depending on how they are counted), the commission members thus indicated that they knew that no errors in the triangles north of the first baseline (north of Lieusaint-Melun in fig 1.3) and south of the second baseline (south of Perpignan in fig 1.4) had any influence on the mismatch. That point is a logical conclusion from the 1799 text, but is not explicitly stated; the authors must have found that assessment intuitive. In probabilistic terms, this is a belief propagation statement, whereby the new data (second baseline measurement) creates new dependencies in one subset of the chain of triangles, while other parts of the chain are not affected.

The intuitive notion that the influence of errors are guaranteed to be limited to only a portion of the problem (no matter the amplitude of the errors) is a subtlety somewhat lost in modern representations of SLAM. Modern and popular views of SLAM focus on exploiting computations to the detriment of structural astuteness and basic intuition. This will be discussed in the next chapters, and more broadly, our insistence on the exploitation of a priori statistical independences between decision variables will be a recurring theme throughout this document. Just as it was recognized before that triangle closure could be used to provide corrections using geometry, practitioners the baseline closure should then also provide additional corrections.

But then, how could these corrections be made ? The 1799 solution was an *ad-hoc* one: to split the chain of triangles, the northern half the chain of triangles was dimensioned by the first baseline, while the southern part of the chain is dimensioned by the second baseline.

However, Delambre would note, in the second tome of *Basis* (Delambre, 1807, p704-5), that there was a smarter way to proceed. Delambre pointed out that the initial decision in 1799 did not remove the mismatch, it only moved it to the middle of the triangle chain (at the town of Evaux). In his view, the smallest corrections possible should be applied in the 53 triangles between the two baselines in order to *restore harmony* (translation ours) :

> To restore harmony, I thought I would be allowed to bring the angles between Melun and Perpignan [i.e., between the baselines] very slight changes which would make both baselines match. [...] These changes are imperceptible in themselves, but acting always in the same direction, they produce the effect that I had in

[13] Two revealing anecdotes: one time wild dogs played too close to the rulers, while another time suspicions of perturbations from gusts of wind also triggered a redo (Guedj, 2000; Alder, 2003).

> mind, which is to make the difference disappear between the two baselines, for I believe their lengths are more certain than our angles.

Delambre qualitative approach is close to the numerical behavior of SLAM solvers : different types of measurements have unequal certainty/confidence attached to them. The constraint of geometric "harmony" is restored by making changes to where he considered the error could be the greatest, i.e., the angles. The fact that the changes made were the smallest possible (under 0.1 seconds of degree) shows his understanding that there was a penalty for unnecessary large corrections. Unfortunately, Delambre did not attempt to make use of his colleague Legendre's 1805 least squares best fit criterion to combine the trigonometric equations. As such, his baseline closure corrections are not presented in a formal way, the amplitude of corrections are tailored to the values of this particular arc meridian dataset. Perhaps he considered the least squares method to be too cutting edge for *Basis*, for it was only one way of combining equations of observations (a.k.a. equations of condition) among many other ways. Legendre's work showed convenience but did not tell how to input prior appreciation of uncertainty. Neither did he claim that the results obtained by this method should be more probable than with other methods. Another difference with SLAM is that the baseline measured lengths were treated as error free. Their predicted errors were just a priori considered smaller than the cumulative effect of angular measurement errors, but no attempt was made to quantify relatively one to the other. This leads to the remark that, although he was closer than the 1799 Special Commission to a sound approach of corrections, an obvious missing central component at the time of Delambre explanation is (again) a mature framework of probabilities. No assessment on the evolution of uncertainty was given to the arc meridian result *a posteriori* of the corrections emanating from both triangle and baseline closures. One could think that this was quite untimely (1807) given significant advances on probability theory appearing in 1809 by Gauss (refer to previous section), and from 1810 onwards by Laplace (as we will come to see in the next subsection). However, reading historian of statistics Stigler, and given the mentioned facts that both Gauss and Laplace were interested in geodesy, we should also consider whether advances in probabilities were not partly due to geodetic adjustment. Indeed, in what he calls the *Gauss-Laplace synthesis* period, Stigler (1986, chap 4) writes (emphasis ours):

> The Gauss-Laplace synthesis brought together two well-developed lines —one the combination of observations through the aggregation of linearized equations of condition, the other the use of mathematical probability to assess uncertainty and make inferences into a coherent whole. In many respects, it was one of the *major success stories in the history of science*. Yet it also poses a puzzle, for the applications of this marvelous statistical technology were widespread only geographically; to an amazing degree they remained confined to the *narrow set of disciplines that spawned them*. They became commonplace in astronomical and *geodetic* work while remaining unknown in the social sciences [...].

Is the problem of *geodetic adjustment*, that has commonalities with SLAM, included in *the narrow set of disciplines* that spawned *one of the major success stories* of science ? Another supporting argument in favor of a positive answer can be found in Laplace 1812's treatise, *Théorie Analytique des probabilités* (Analytical theory of probabilities), one of his most famous work.

1.4.2 Laplace's framework and the theory of corrections

Having worked for decades on probability (Dhombres, 2012), since 1774, when he rediscovered 1763 Bayes law (under a different formulation, as Bayes work was not widespread yet), Laplace motivation to write a treaty on probabilities may have been magnified by Legendre and Gauss developments mentioned earlier, in the 1805-1809 period. One issue Laplace and colleagues did have, before 1809-10, is the specification of the error curve of observations. There are indeed many functions that are unimodal, symmetric and whose density is more or less concentrated about zero. But nothing could justify that an error curve adhering to these axioms should be preferred over any other, besides the convenience provided during the calculations. For instance, in the 1774-1786 period, Laplace used the functions $\frac{m}{2}e^{-m|x|}$ and $\frac{1}{2a}\log(\frac{a}{|x|})$, $|x| \leq a$, see Stigler (1986, chap 3). These are obviously quite unpleasant forms to work with. Yet, ironically, not only he had encountered the form $e^{-h^2x^2}$ before, but he is recognized to be the first to derive the value of the *Gaussian* integral as early as 1774 in another problem (Stigler, 1986, footnote 8, chap 3). Only in 1810, he proved the central limit theorem, which gave legitimacy to the normal distribution when the number of measurements is large. Then, later that year, reading Gauss' 1809 work, Laplace supplemented his work by propagating that legitimacy to the least squares method[14]. Furthermore, he showed using Bayes law, that Gauss's "most probable" answer was also the maximization of the posterior, which could also be quantified (the "weight of result", now known as precision).

Advances were then compiled in the 1812 book *Théorie Analytique des probabilités*, of which we have access to the fourth edition (Laplace, 1820), completed in 1820.

The first edition (1812) presents all his results in two books (Livre I, Livre II). In the end of fourth chapter of *Livre II*, Laplace (1820, p353) acknowledges Gauss and Legendre significant role as there was no fixed rule for combining different (linear) equations of observations, before the least squares method. All existing rules (e.g., Cotes' rule or Boscovitch's rule) were arbitrary and based on vague considerations; sometimes measurements were dropped at the observer discretion. Legendre and Gauss were thus praised for having managed to *avoid these gropings*.

In 1814, Laplace supplemented his 1812 treatise with *Essai philosophique sur les probabilités* (*Philosophical essay on probabilities*), a text for the general public explaining the application of probabilities in various fields, e.g., game of chances, judgement of courts, life expectancy, marriage, astronomy, and of course, in interest to us, geodetic adjustment.

Laplace (1814b, english translation of 1902) writes in the chapter dedicated to natural philosophy (emphasis ours):

> This method [rendering a minimum the sum of the squares of the errors of observations] may be employed again with success in geodetic operations. We determine the length of the great arc on the surface of the earth by a chain of triangles which

[14] Laplace's position was to believe Gauss' priority by recognizing that he had corresponded with some other astronomers about the method (Laplace, 1820, p353). However, we can remark a facet of the Gauss-Laplace competition, as he only says in the text that Gauss *attempted* to link the method with probability theory, implying he did not fully succeed for his taste. For Stigler (1986, chap 4, Reenter Laplace): *Laplace may have said, Gauss's derivation was nonsense, but he, Laplace, already had an alternative in hand that was not [i.e., the CLT]*. Gauss, born in 1777, was a generation younger than Legendre and Laplace, he would instead provide a better justification in 1823 (Grcar, 2011). Finally, the way Laplace divided the credit between Legendre and Gauss for the least squares pipeline might have superior motives, as it makes his contribution more important (in relative terms).

> rest upon a base[line] measured with exactitude. *But whatever precision may be brought to the measure of the angles, the inevitable errors can, by accumulating, cause the value of the arc concluded from a great number of triangles to deviate appreciably from the truth.* We recognize this value, then, only imperfectly unless the probability that its error is comprised within given limits can be assigned. The error of a geodetic result is a function of the errors of the angles of each triangle.

The sentence we emphasize above is structurally similar to the explanation of drifting in SLAM given in the beginning of this document, and is shown in dash gray on figure 1.6. Without corrections, unavoidable accumulation of errors leads to drift. This drifting aspect, an intuitive/qualitative consideration, was evidently identified long ago in SLAM, e.g., in one of the early papers by Chatila and Laumond (1985, §1.3):

> World models are generally built gradually by the robot itself as it moves and discovers new areas. As long as only new parts of space are discovered, there is no possibility to find out and correct any errors, in the model being built, that are due to robot position drifts. But when the robot is to perceive again already known and modeled regions, inconsistencies may occur, and this correction becomes necessary, and sometimes possible.

Note that this drifting aspect was not universally well understood among the practitioners of geodesic operations. For instance, when he came back from the 1735 Peru/Ecuador expedition, Condamine (1751, p. 2) wrote that it would be "against all likelihood to suppose that the errors [...] accumulate instead of compensating each other [...], the amplitude of the overall error should not be feared from a bigger chain than from a small chain". Despite this incorrect reasoning, the chain of triangles in Peru/Ecuador did have a second baseline (in case that "strange hypothesis" of error accumulation were true, as La Condamine puts it).

In geodetic adjustment, the first possibility to bring correction is via triangle closure, i.e., by measuring the angular distance between the 2 previous stations within each triangle (where flags were generally left to be sighted from afar) from the third station (the third angle) of the triangle (Laplace, 1814b, *Application of the Calculus of Probabilities to Natural Philosophy*):

> There is then a great advantage in observing the three angles of each triangle and in correcting them as we have just said [by correcting a third of the difference between the sum of the three angles and 180 degrees]. This advantage can be foreseen by basic common sense; but the calculation of probabilities alone is able to appreciate it and to render apparent that by this correction it becomes the greatest possible [i.e., the most probable values].

Triangle closure by measurement of the third angle was, historically, straightforward enough (*basic common sense*) to not leave room for ad-hoc fuzzy judgements. Assuming that the errors distribute similarly in the three angles of the triangles, which is reasonable, each angle could be corrected by a third of the excess summation of the angles from π. Laplace here merely points out that only a probabilistic framework can prove that this is the best approach.

However, what was previously vague and confusing (including for Laplace in 1799, see figure 1.7) is the correcting mechanism associated to the baseline closure. There could be an *infinity of ways* to proceed, but the most advantageous corrections were given by the analysis of probabilities. Laplace (1814b) continues in the same chapter:

> In order to ensure oneself of the exactitude of the value of a great arc which rests upon a base[line] measured at one of its extremities, one measures a second base[line] toward the other extremity; and one concludes from one of these base[line]s the length of the other. If this length varies very little from the observation, then there is all reason to believe that the chain of triangles which unites these base[line]s is very nearly exact and so is the value of the large arc which results from it. One corrects, then, this value by modifying the angles of the triangles in such a manner that the base[line] is calculated according to the base[line]s measured. But this may be done in an infinity of ways, among which is preferred that of which the geodetic result has the greatest weight, inasmuch as the same error becomes less probable. The analysis of probabilities gives formulae for directly obtaining the most advantageous corrections which results from the measurements of the several base[line]s and the laws of probability which the multiplicity of the base[line]s makes; laws which become very rapidly decreasing by this multiplicity.

Hence, the *most advantageous corrections* are those which give the posterior result the greatest "weight". Laplace said that his new tools allowed him to come up with an unambiguous approach to the baseline closure correction. Furthermore, even though only two baselines existed for this arc meridian operation, he could contemplate how multiple baselines would improve the "weight" of the result. Similarly, multiple loop closures in SLAM generally[15] steer the *maximum a posteriori* (MAP) state closer to the truth. The mathematical details of this text are given in the second and third supplements of *Théorie Analytique des Probabilités* (Laplace, 1820). Those supplements were appended in 1818 and 1820[16]. From that point, Laplace was able to answer a rich set of questions that could arise from the structure of a geodetic adjustment problem. Let us identify the ones that are fairly close to SLAM (other ones relate to the determination of certain aspect of the shape of the earth).

In the introduction of the second supplement, the points mentioned above are repeated from 1814's *Philosophical Essay on probabilities* and expanded upon. Laplace then points out that the angular errors distribute normally. This is justified by the process of the instruments, namely Borda's repeating circles (see figure 1.2). Indeed, as the reading given by the instrument is the sum of an arbitrary number of *independent* measurements, a direct application of the central limit theorem is identified. Thus, denoting α the angular error, we have the distribution $\alpha \sim e^{-h\alpha^2}$ (h not given though). Even though an individual sighting has unknown distribution, the CLT gives the distribution of the average.

In fact, upon closer inspection, we postulate that at the time, Laplace's CLT was most suited to application on data from the Borda's repeating circle. Indeed, the name of the chapter (Laplace, 1820, chap 4) is "Results averaged from a large number of observations: on the probabilities of their errors, and on the most advantageous mean results". Although a bit longer than the 20th century branding "central limit theorem", this title made clear that he was concerned with observation errors, and not with game of chances like in the preceding chapter (Laplace, 1820, chap. 3) like Moivre before him. In this fourth chapter, Laplace first considered the errors to be discrete integer values $-n, -n+1, \ldots, -1, 0, 1, \ldots, n-1, n$ (Laplace, 1820, p309) which would correspond to the readings of a graduated instrument,

[15] Excluding outliers from incorrect loop closure association.

[16] We can find translations of Laplace work in Pulskamp's website: http://www.probabilityandfinance.com/pulskamp/Laplace/index.html.

before extending it to the continuous case. Finally, the feasibility of getting many observations, under similar conditions, necessitates (a) the availability of an instrument that allows repetitions such that one measurement was not affected by the previous (independence condition), and (b) a time invariance of the conditions of observations between the first and last observation. As mentioned earlier, Borda's repeating circle (figure 1) could be used both for astronomical observations and angular measurements between two points at the surface of the earth. However, in the case of astronomy, due to the persistent relative motion of celestial bodies (especially comets), by the time the observer moves the scope and watches again (between half a minute to a minute), the relative angulation from the observer to the object significantly changes. This makes the problem of geodetic adjustment, again, one of the most eminent of its time as both a recipient and a catalyzer of fundamental scientific progress.

As mentioned, a condition for the CLT to apply is the independence of additive sources of measurement: Laplace stresses that all observations should be reported, i.e., no guess work to get rid of some of them. This echoes Delambre's approach and his critics of the attitude of former observers.

Remark 1.2. *Delambre critics were not just directed towards the conduct of mid-18th century geodesic operations. He also ended up disapproving of the attitude of his former colleague Méchain. He eventually realized that Méchain was concealing latitude measurements that were inconsistent: "Méchain had an unfortunate resemblance to Bouguer[one of the scientist in the 1735 Peruvian operations], i.e., the weakness leading him to conceal anything that could diminish the reputation he aspired for: to become an observer whose dexterity and exactitude are above any other astronomer" (Delambre, 1817, p283).*

The first section of the second supplement deals with the processing triangle closure Laplace (1820, p531). Laplace estimates, via least-squares optimization, the corrected value $\bar{\alpha}$ as a function of the sum of errors and the weight h. The value of h in $e^{-h\alpha^2}$ is inferred by using the sum of squares, over all n triangles in the chain, of the three observed angles, minus π. In the second section, the expression of the error on the arc length is deduced as a function of all the angular errors, and so is its weight. In the third section, Laplace addresses the question of baseline closure by determining the corrections to be applied to all the angular errors. These angular errors are denoted $\alpha, \alpha^{(1)}, \ldots, \alpha^{(n-1)}, \gamma, \gamma^{(1)}, \ldots, \gamma^{(n-1)}$ in the text, where superscript (i) indexes the ith triangle. The estimation is again done by maximizing an exponential that holds a sum of squares. The resulting corrections depend on the mismatch λ between the predicted length of the second baseline and its measurement. Laplace shows that the "weight" of the resulting arc-length is increased following the baseline closure: in other words, not only the error terms are subjected to corrections, but their uncertainty decrease. The reasoning generalizes to multiple baseline closures. Note that in the virtual chain of n triangles considered, the first and second baselines are assumed to be located in the extreme ends, which was not the case in the real Dunkerque-Barcelona measurements (figures 1.3 and 1.4).

The third supplement (Laplace, 1820, p581), probably added in 1820, deals with questions raised by the extension of the meridian. To contextualize, in the new 1806-1807 expedition in Spain, Biot and Arago added more triangles south of Barcelona along the coast up to Denia, in a southwesterly direction. Then, forming an "L" shape (figure 1.8), the last triangles go eastwards to the Balearic Islands of Ibiza, Formentera and Mallorca, thus covering again the original Dunkerque-Barcelona meridian (Bayart, 2007). Unfortunately, due to an(other) erupting conflict between France and Spain, no additional baseline length could be measured

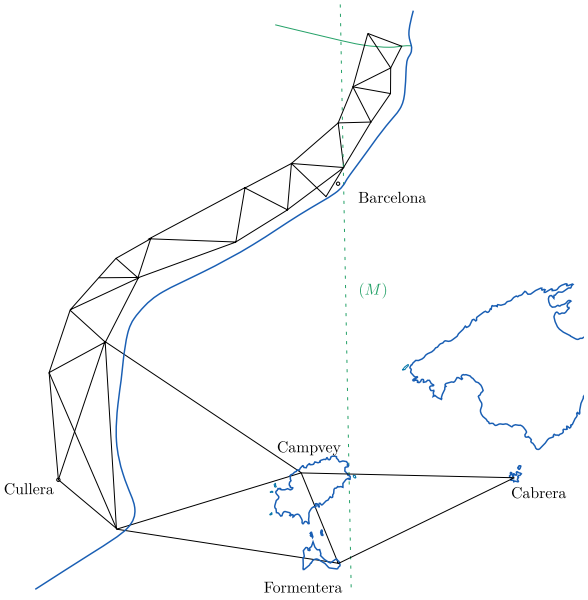


Figure 1.8: Approximate reproduction of the extension of the Dunkerque-Paris-Barcelona meridian by Biot and Arago, 1807. Loosely based on Bayart (2007) and Levallois (1988, fig 20).

on site to provide closure for this part of the arc meridian[17]. A chain of 26 triangles was open-ended from the second baseline (Perpignan, figure 1.4), extending approximately over 466 km. Laplace determined the uncertainty associated with this predicted value. In his formulation, it can be bet 1 to 1 that this error lies within ± 8.0757 meters. The same style of description of confidence interval is also employed for astronomical examples in Essay (1814). This is an enrichment of vocabulary compared to the second supplement: Laplace does not just speak of "weight" anymore. Additionally, we identify the following questions Laplace answered in this third supplement: what would be those limits if instruments from previous generations were used ? How much will those limits change if, as contemplated, 2 more triangles were added which include a new measured baseline ? Then, the topic of baseline closure is addressed again in the third section of third supplement (Laplace, 1820, p589). Laplace calculates the "probability of *simultaneous* existence" of the errors $\alpha, \gamma, \alpha^{(1)}, \gamma^{(1)}, \ldots$ resulting from the processing of the mismatch λ between the second baseline measured length *vs* the one deduced from the chain of triangles. Then, the resulting effect on the distribution of the error of the full arc length is derived. It is then predicted that the measure of the length of a new baseline in Formentera would yield a mismatch of ± 0.3423 meters. The main difference in the approach with the second supplement is that the prior (to corrections) distribution of errors are not assumed normal[18]. Laplace (1820, p590-1) also admits that the approach for the baseline closure he and colleagues used in 1799 (see figure 1.3) was incorrect,

[17] The situation of Arago, in particular, became extremely precarious: almost lynched by a mob, he had to turn himself prisoner in Palma for his safety, he was then announced hanged by local newspaper, he evaded in a boat to Alger, was attacked by corsairs on the way back to France, ended up detained in continental Spain, was released but a boat he was in drifted to North Africa due to a sea storm, found itself once more in Alger, where he was stuck for other reasons (Arago, 1854) etc. Nonetheless, he came back in Paris in 1809 with the observations. The aggregate perils and somewhat amusing anecdotes of Arago, Delambre, Méchain later inspired a novel by Jules Verne in which protagonists conduct a geodesic operation (Verne, 1872).

[18] In our interpretation, this still shows that Laplace was not confident on using the normal distribution in certain situation.

"in the ignorance that we were with respect to the true *theory of corrections*"[19].

1.4.3 Gauss' geodetic adjustment

In the 1826 supplement (*Supplementum*) of his 1823 paper (Bertrand, 1855; Gauss, 1826, p70-112), it is Gauss who introduces the term *adjustment of the observations* (*observationum compensationem*), which was used by neither Legendre, Delambre nor Laplace. In sections 19 and 20, Gauss acknowledges that calculating the least squares for large systems can be discouraging. He proposes to divide the problem into "groups" and to proceed iteratively by: calculating the adjustment for one group, ignoring the second group, then using the results of the first group to adjust the second group, which would in turn gives new starting values for adjusting the first group, and so on, until convergence.

> Whenever the number of conditional equations is very large, determination [...] by direct elimination can be very costly. Then it will often be advantageous to complete the compensation by successive approximations [...]. [Later, on the same page] After a few iterations we will arrive at stable numbers.

That description, surprisingly, evokes the notion of message passing by loopy belief propagation (LBP) in graphical systems. LBP appears in probabilistic graphical models (Koller and Friedman, 2009), with a variant called Gaussian belief propagation. Ironically, the latter was not named after Gauss, but rather on the fact that it is a variant of LBP with Gaussian distributions. For examples of applications in SLAM, see loopySAM (Ranganathan et al., 2007) and FutureMapping (Ortiz et al., 2021; Davison and Ortiz, 2022). Further historical investigations could be done on that particular topic. In this text, Gauss did not clarify whether this procedure was applicable to circular/loopy clusters. In sections 21 to 25 (Bertrand, 1855; Gauss, 1826, 101-112), Gauss applies his framework to some small sections of the geodetic adjustment of Holland (section 22) and Hanover (section 23-25), see figures 1.9 and 1.10. How different is his treatment from Laplace's, given that both used least squares ? Both seek to use the least squares to correct the observations, but Gauss does not pose what he called "the equations of condition" (constraints) in the same way. He states that his system can be populated by 3 types of equations of condition which we will name horizon closures, triangle closures (not new), and chain closure:

- (**I**) the horizon closures, whereby the sum of the angles around one station should be 2π. This requires the exterior angles of the triangle to be observed;
- (**II**) the classical triangle closures, whereby the sum of the angles in one triangle should be π ;
- (**III**) the chain closures, which concerns a chain of triangles that join itself (like a circuit). In any triangle, the ratio of the sides lengths is the ratio of sines of the opposing angles. Type (**III**) equations consist, for each triangle, in considering sides adjacent to both the previous and the next triangle in the chain. The product of all those ratios is 1, since the chain forms a circuit. See figure 1.11 for an example.

[19] This statement may not just a scientific judgement: the early metric system had been scraped in 1815 by the new royal authorities. Laplace tended to adapt very well to the numerous regime changes of the period.

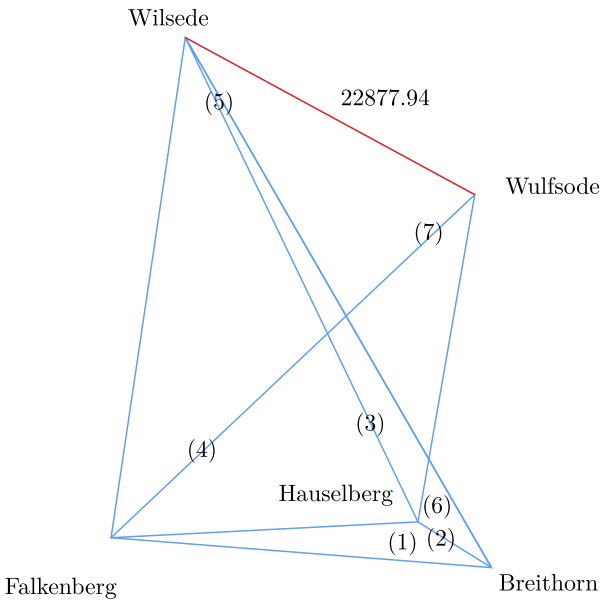


Figure 1.9: Reconstitution of the toy problem addressed by Gauss in 1826 in *Supplementum* (Bertrand, 1855, p106-112) which is sampled from the triangulation of Hanover. There are 7 triangles and all 3 angles of each triangle are measured. A baseline has been measured between Wilsede and Wulfsode (red line).

First note that Laplace did not address the third case, as he only dealt with a chain of triangles. In that regard, Gauss' approach applies to a larger set of topologies.

Let us investigate how Gauss tackles the small Hanover problem (figure 1.9). There are 27 angular measurements, and thus 21 corrections to make[20]. Gauss builds a system with 7 equations of type (**II**) equations of condition (i.e., the triangle closures), and 8 equations of type-(**III**). There are no first type-(**I**) equations considered in the Hanover problem. Gauss then raises an important caveat. Given the topology considered whereby some triangles can be contained with a larger one, only 5 out of 7 equations of type (**II**) are mutually independent[21]. Concerning the 8 possible ways to pose equations of type (**III**), in other words among the 8 possible ways to consider a circuit of triangles for in figure 1.9, only 2 remained independent. Consequently, there are only in total 7 equations of condition for 27 decision variables, which makes the system under-determined (as also corroborated by Grcar (2011, §3.3)). This raises a completeness issue whereby it cannot be guaranteed that all the information of the problem have been used to their fullest extent.

Concerning the portion of the Holland network of 9 triangles, see (Grcar, 2011, fig 6) for illustration, not only the same problem occurs, but also the choice of equation of conditions is different: 13 equations of condition: 2 type (**I**) equations, 9 type (**II**) and 2 type (**III**) equations.

Thus, the main difficulty in Gauss' adjustment framework, apart from the under determi-

[20] Gauss does not detail in his memoir whether the measurement of the baseline is considered perfect or not. We assume he does, like Delambre and Laplace did. Nonetheless, it does not affect our analysis.

[21] In the linear algebra language that did not exist back then, the information matrix was not full rank.

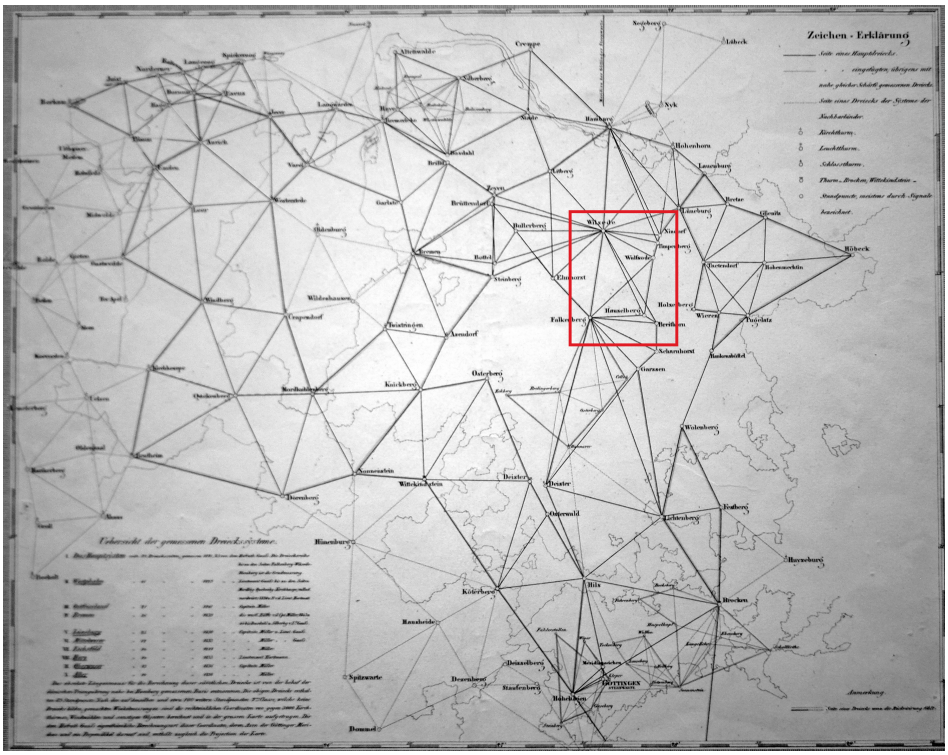


Figure 1.10: Geodesic network over Hanover which Gauss worked on between 1821 and 1825, reproduced from Wikipedia Commons (2023). The red box we appended corresponds to the triangles studied by Gauss in *Supplementum* (Bertrand, 1855; Gauss, 1826, p106-112) and shown in figure 1.9.

nation aspect, is the difficulty to pick the adequate equations while ensuring independences, i.e., so that no equation comes as a linear combination of others. Otherwise, the same equation of condition may artificially be double counted in the product of probabilistic factors (1.1). Besides, each different network of triangles requires separate careful considerations to deal with this complication. Specifying all assumptions (in the form of equation of conditions) Gauss made for a geodetic adjustment was thus not straightforward. We can see that for large networks, such as figure 1.10, the selection of equations can become non trivial.

1.4.4 The variation of coordinates

In his treatise of geodesy (Tardi, 1934, chap 10), cartographer P. Tardi confirms the difficulty to build a Gauss adjustment system for large maps, although he does not provide a thorough analysis. He argues in favor of using another method: the variation of coordinates. Instead of correcting the errors, this method proposes to formulate the problem on the geographical coordinates of the geodesic stations, i.e., on the vertices of the triangles. In the previous formulations, which concerned the corrections of the measurement errors, the geographical coordinates[22] were only the byproduct of the adjusted observations. The coordinates of these stations are now the decision variables, and the 'factors' involved are azimuthal relations between observed stations. The adjustment, still conducted via least squares, consists in

[22]Those coordinates could be understood as either: plan coordinates $\mathbb{R}^2$, longitude/latitude, or more involved projections like Lambert, Mercator etc.

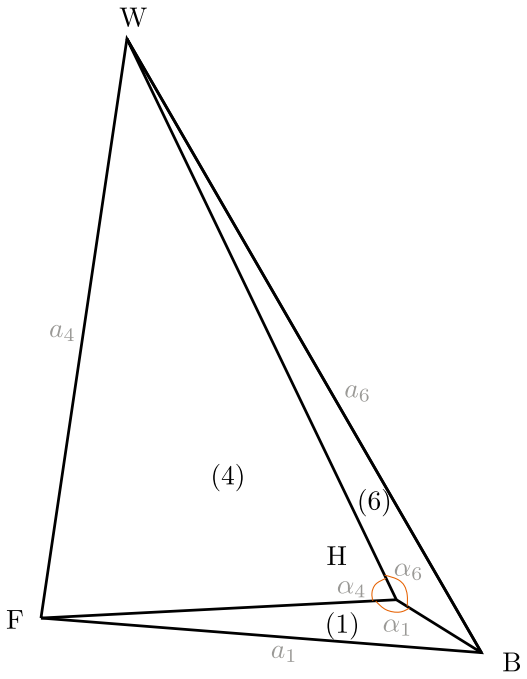


Figure 1.11: Portion of the triangles from figure 1.9 illustrating the principle type (**III**) of chain closure. Let $\alpha_i, \beta_i, \gamma_i$ denote the angles of triangle (i), in clockwise order (only α is shown for clarity). Let a_i, b_i, c_i denote the lengths of their respective opposing side. Consider the circuit for the ratio of adjacent sides $(1) \to (4) \to (6) \to (1)$, the product ratio $\frac{c_1}{b_1} \cdot \frac{c_4}{b_4} \cdot \frac{c_6}{b_6}$ amounts to 1 since $b_1 = c_4$, $b_4 = c_6$ and $b_6 = c_1$. Using the law of sines, this leads to the equation of condition: $\frac{\sin \gamma_1}{\sin \beta_1} \cdot \frac{\sin \gamma_4}{\sin \beta_4} \cdot \frac{\sin \gamma_6}{\sin \beta_6} = 1$, that Gauss employed in a null sum of logs form (Bertrand, 1855; Gauss, 1826, p109, §24).

finding the coordinates maximizing the product of factors (same as in equation (1.1)).

The motivation for such systems is the simplicity of setting the equations and the fact that there are fewer stations than measurements. Indeed, consider the Hanover example figure 1.9. Under the method of variation of coordinates, there are only 7 unidimensional variables (the stations coordinates)[23] and 18 azimuthal relations. This is deduced from Bertrand (1855); Gauss (1826, data p106, §24). Tardi mentions that the 1830 triangulation of France (figure 1.12), which was built around the Dunkerque-Paris-Barcelona meridian, used this type of adjustment. However, according to Levallois (1988, p96-7), who was himself a successor of Tardi, the 1830 map of France was never properly adjusted due to the underestimation of the time allocated to conduct those calculations. For Jozeau (1997), this period, which roughly starts after Delambre, Laplace and Legendre deaths (resp., 1822, 1827, 1833), corresponds to a relative decline in French geodesy, where the various methods of least squares did not receive much consideration compared to the advances made by Europeans neighbors. Thus, it is likely that the method of variations of coordinates was used initially just for small scale

[23]There are 5 stations which imply for a state of size 10 (assuming two coordinates per station). However, since we are in a projected space, one station should be fixed as the starting point, while at least one other station should serve to orient the network (by local latitude observation) via one of its component. The station which orients the network is commonly known as the "Laplace station" in geodesy. Thus, there are only 7 decision variables left.

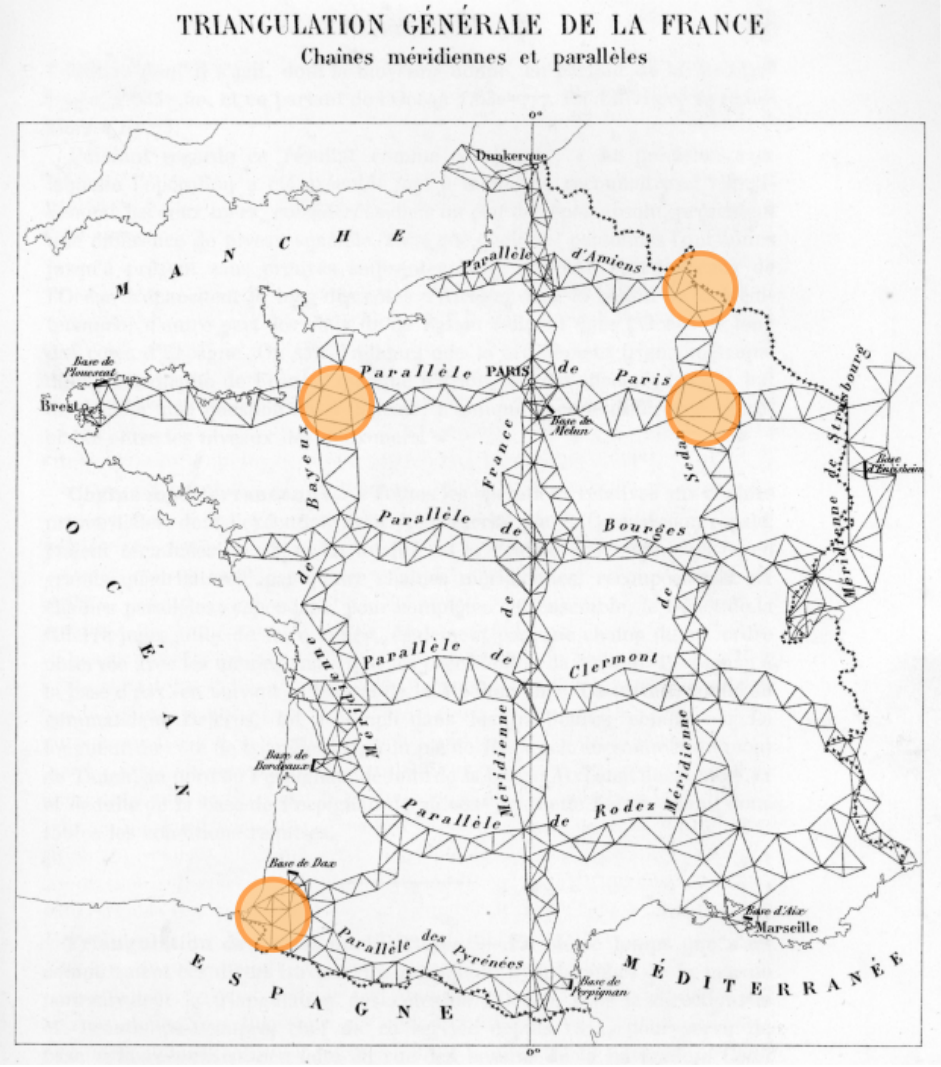


Figure 1.12: First order triangulation done for the map of France between 1817 and 1830. Reproduced from Berthaut (1898b, p24-5). The orange circles show the areas with the greatest mismatch due to the absence of global adjustment.

adjustments (see caption of figure 1.12).

We have not found out when, where or by whom precisely the method of variations was invented, or around what time it overtook methods that considered the errors as decision variables. Presumably, it was in the late 19th century when additional type of measurements where introduced such as the telegraphic communication to deduce the longitude at a station. In his treatise, Liagre (1852, 3rd part) only describes the method of errors. But once the variations of coordinates took over, there are no signs that the dominance of these methods were seriously contested. See Golub and Plemmons (1980) for a late 20th century approach to geodetic adjustment.

1.4.5 A least squares winter in France

Jozeau (1997) investigated on the surprising loss of interest in France for geodesy and ad-

Box 2: The 1830 triangulation of France, and its problems.

Figure 1.12 shows the first order triangulation of France done by the cartographers of the "Dépôt". This triangulation builds on the French portion of the Dunkerque-Barcelona meridian measured between 1792-1799 (0th degree in figure 1.12). The term "first order" comes from the fact that this network was the main frame of a more detailed 1 : 80000 French map, published in 279 sheets. Smaller triangulations and surveying activities could then be attached to this first order triangulation. Hence, in principle, this is the highest order of a hierarchical map. The shape of this first order network, which roughly cover 3 meridians and 6 parallels, was initially outlined by Laplace and Delambre in 1816.
Unfortunately, as noted by Levallois (1988, p96-7), the adjustment due to the closure at the junctions of the various meridians and parallels were not calculated due to both: (a) the scale of computation; (b) insufficient manpower for the task. Relating to (a), Levallois, himself an accomplished 20th century geodesist, estimates that a proper adjustment would have taken 30 to 40 years to be completed by one calculator: "even with today's computers [1980s], it would demand careful planning and preparation" (Levallois, 1988, p96-7). Relating to (b), Levallois points out the absence of a bureau of calculus: the observers (a relatively small team of specialists) had the task of calculating the adjustment during their "rest time" in winter, which only provoked delays and concessions due excessive workload (Levallois, 1988, p100). At the junctions between several chains of triangles, the mismatch could reach several tens of meters (orange areas in figure 1.12). Jozeau (1997) points to the absence of scientific research (& review) in this topic in the 1830-1855 period. Consequently, small size adjustments were made within each sheet of the French map, which moved the mismatch in the overlap of adjacent sheets. "In the country of Legendre, we could have done better" Levallois (1988, p100) regrets.

justment methods, in particular for the least squares technique. The Borda's repeating circle instrument, victim of its success, was used until the 1860s, when it was finally replaced by more modern hardware such as the theodolite. The method of least squares was taught by Arago between 1817 and 1836, but without relying on probability theory, which at the time was considered too abstract (Jozeau, 1997, p46). Even worse, the teaching of least squares in France was almost stopped by prominent scientists like Le Verrier and Cauchy in 1850 (Jozeau, 1997, p47). Le Verrier considered that the method of least squares leads to excessive calculation costs and that, instead, the practitioner should use his best conjectural judgement to find his/her own method to adequately combine observations. On the surface it seems that ad-hoc and vague considerations were back from the 1750s. But there might be more to it, we will come back to that issue in a later development.

Belgian geodesist Liagre, who was aware of the difference of interest in the subject between France and Prussia, noted in the preface of his treaty Liagre (1852) that the employment of the calculus of probabilities "has raised against it, from the outset, numerous prejudices which have not yet been entirely dissipated".

As discussed in section 1.2, missions and post analysis of important geodesic triangulation were given to high ranking scientists, arguably often the best of their time and discipline. But starting from 1816, the bulk of geodesist practitioners were attached to a military body, the

"Dépôt de Guerre" (the War Repository, roughly translated). But this organization was suboptimal from the viewpoint of scientific progress, it was later even criticized from insiders. Indeed, it mainly affected the cartographers to unending geodesic campaigns, with the main focus set on finishing triangulations of France and colonial possessions. The members of this unit followed strict procedures, and little room was left for critical thinking towards the established methods, which were considered complete (Levallois, 1988, p127). The practice of French geodesy was practically cut from foreign scientific advances, which themselves were initially inspired by successes of the 1792-1799 arc meridian (Jozeau, 1997).

The situation would nonetheless improve in the late 19th century, when the technical and scientific advances made by scientists such as Gauss (see 1.4.3), Bessel, Struve, Everest finally pierced, notably thanks to the treatise of Belgian geodesist Liagre (1852) and Bertrand (1855) translation of Gauss' contributions to the theory of errors.

But even in the beginning of the 20th century, the methods of least squares, and more widely the concept of adjustment, still had its sceptics in France. For instance, Henri Bouasse, physics professor of the University of Toulouse, published a book on geodesy in 1919 (Bouasse, 1919). Our esteemed colleague from a century ago was known for not concealing his strong views whenever he considered the methods of his peers to be inadequate. In the opening of his book "Géographie Mathématiques", Bouasse immediately berates geodesy as a field. For him, there is not much left that the methods in this field can offer of value (Bouasse, 1919, p4):

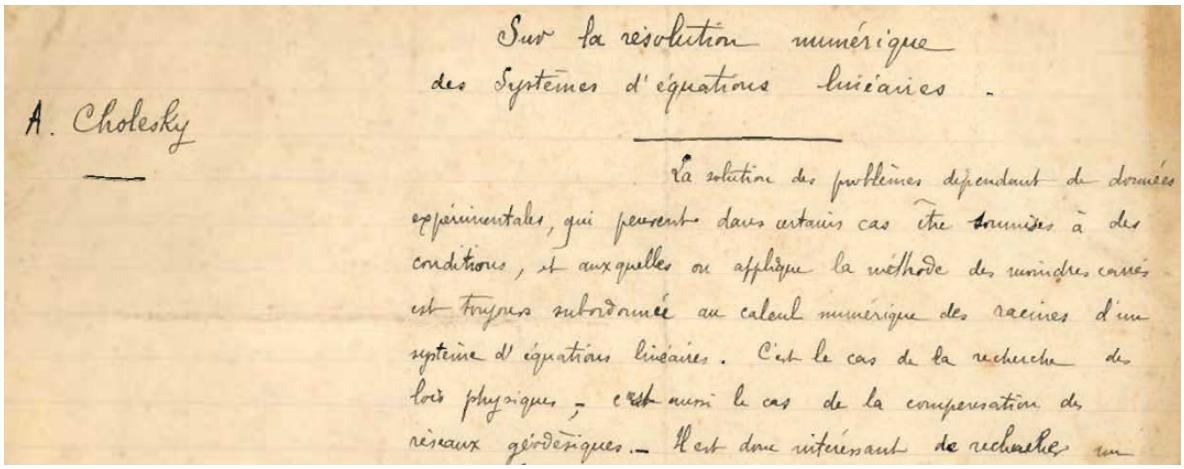
> The methods [in geodesy] have been established for a long time; we only hope for technical improvements in detail, interesting for sure, but presenting the same order of interest as a more convenient corkscrew or a more economical stove.
>
> H. Bouasse

It is surprising, if not revealing, that in this 500-page book, Bouasse barely addresses the concept of geodetic adjustment. While in contrast Gauss (1826, §22) referred to geodetic adjustment as "geodaesiam sublimiorem" (superior/sublime geodesy) in *Supplementum*, Bouasse only covered the topic in less than a page. Mentioning the corrections made in *Basis* for the 1792-1799 arc meridian dataset (see 1.4.1), Bouasse (1919, p244) expresses dissatisfaction with Delambre's explanations a century ago. He writes:

> The process of propagating the mismatch between a calculated and a measured baseline throughout the chain [of triangles] : this is what I call tinkering with the results.

Evidently, the concept according to which the errors in different observations compensate, not multiply, still provoked mistrust.

Nonetheless, during the same period, new impactful techniques would emerge from the field. The Cholesky decomposition is a numerical technique sometimes used in SLAM that originates from geodetic adjustment. A.L. Cholesky was a cartographer in the military body mentioned above. He specialized in the processing of geodesic operations (Brezinski and Gross-Cholesky, 2005). He would come up in 1910 with an algorithm to speed up calculations, helped by a mechanical calculator. But it would only be published in 1924, years after


Sur la résolution numérique
des Systèmes d'équations linéaires .

A. Cholesky

La solution des problèmes dépendant de données
expérimentales, qui peuvent dans certains cas être données à des
conditions, et auxquelles on applique la méthode des moindres carrés
est toujours subordonnée au calcul numérique des racines d'un
système d'équations linéaires. C'est le cas de la recherche des
lois physiques ; c'est aussi le cas de la compensation des
réseaux géodésiques. — Il est donc intéressant de rechercher un

Figure 1.13: *On the numerical resolution of systems of linear equations.* Draft written by Cholesky signed and dated Decembre 2nd, 1910. It was found relatively recently, in 2005. The 8-page document details the efficient linear systems decomposition the author is known for. The motivating problem was the numerical application of the adjustment of geodesic networks (*compensation des réseaux géodésiques*), whereby the "system of linear equations" (today $Ax = b$) is given by the method of least squares. As an aside, and of interest to computer scientists, Cholesky points out that the technique allowed him to use the capabilities of his mechanical calculator with "maximum efficiency" (page 4). Reproduced from the Sabix library (Cholesky, 1910). For further analysis, see (Mansuy, 2008; Brezinski, 2018).

Cholesky death towards the end of the World War I (1918). The well known Cholesky decomposition[24] is considered a major tool in scientific computing. Its sparse variant, implemented by Davis (2004); Davis and Hager (2009); Chen et al. (2008), is notably made available in many software packages through their dependencies on *SuiteSparse*. An extract of his draft (Cholesky, 1910) is reproduced on figure 1.13.

Relevant on our discussion, the application of the technique (often noted today as $LL^\top = A$, L lower triangular) concerned a geodetic adjustment of 56 equations to "compensate [i.e., adjust] the altitudes of the primary[25] chains of triangles of Algeria" (Cholesky, 1910, p8). This is another case of a useful algorithm coming from geodetic adjustment.

We now make some important points for the remainder of the document, which will be valuable in view of our ultimate goal: an approach of the SLAM problem through causality and the explicit representation of measurement errors.

1.5 Analysis and Ramifications

In this section, we seek to categorize the similarities and dissimilarities of the problem described above, formalized around 1820, in comparison to the modern view of SLAM.

We highlighted the fact that describing geodetic adjustment in terms of the least squares technique is not sufficient. A system that employs least square needs to be fed equations of

[24] Although the Cholesky decomposition is now understood and implemented in the matrix form, the original draft only deals with system of linear equations.

[25] The word "primary" may refer to the geodesic stations, i.e. the vertices of the triangles of the primary chain. The primary chain of triangulation was the main frame of a map, to which local secondary chains attached themselves to. It has the same meaning to "first order" used for the map figure 1.12.

Approach	Year	Limitations	Historical Example
Laplace corrections	1812	Presented in limited scope (only open chain of triangles)	Dunkerque-Paris-Formentera arc meridian (1806)
Gauss corrections	1826	Difficulty in picking independent equations; under-determined system of equations	Hanover (1825), British Isles (1880)
Coordinates Variation	*Not found*	Additional considerations are required due to the projection system; higher than expected computational cost	NT France (1959), US National Geodetic Survey

Table 1.1: Summary of the various approaches to state the geodetic adjustment. Note that once the problem is posed, each approach has been historically treated with least squares.

condition, a.k.a. likelihood functions, and decision variables. In other words, even if the least squares became the estimator of choice, the problem of geodetic adjustment still required the non-trivial step of specifying the functional to be estimated, a.k.a. the estimand. Regarding which variables can be picked as decision variables, we identified two school of thoughts: one aims to correct directly the measurement errors, i.e., the theory of errors of Gauss and Laplace, while the other aims to estimate the projected coordinates in the map given the measurements, i.e., the method of variations of coordinates.

The method of variation of coordinates appeared mainly due to the limitations, or perceived limitations, of the preceding theory of errors[26].

1.5.1 Comparison with SLAM

In no particular order, the commonalities are listed as follows:

A spatial inference problem Even though geodetic adjustment is not concerned with the locomotion of a robot like SLAM, it is a problem of spatial inference. After all, geodesy is the study of the shape of the earth, which involves mapping. Measurements are plagued by uncertainty coming from imperfect observations. In arc meridian problems, the trajectory could be considered as the meridian curb joining both ends of the chain of triangles, although, obviously this line which has not been travelled.

Drifting The phenomenons of drift by error accumulation and adjustment by closure affect both SLAM and geodetic adjustment. Any approach of such problems consisting solely on the minimization of the observations errors is doomed, giving rise to the need of closure mechanisms and concepts enabling the combination of equations of condition/likelihood functions.

Hidden systematic errors in some measurements Local biases affect some instruments. It is the case with the repeating circle, when set in astronomical configuration, as well

[26] In part of the literature, the method of variation of coordinates is considered as part of the "theory of errors". We chose not to include it to mark the difference with the approach of Laplace and Gauss that is described in §1.4.2 and §1.4.3.

as with modern inertial units (typically both accelerometers and gyroscopes are affected by bias).

Probabilistic formulation Estimation queries are answered based on a probabilistic formulation: the joint pdf is a product of the densities of the observations. We have seen that there are several ways to pose the estimand (see table 1.1). In the modern version of equation (1.1), the elements $\varphi(V - M)$ are replaced by the conditionals $p(z_k|\mathcal{X}_k)$ where z_k denotes the k-th measurement and $\mathcal{X}_k$ denotes a subset of the hidden parameters $\mathcal{X}$. Finding the maximum likelihood estimate amounts to maximizing the product:

$$\hat{\mathcal{X}} = \underset{\mathcal{X}}{\operatorname{argmax}} \prod_{k=0}^{M} p(z_k|\mathcal{X}_k). \tag{1.3}$$

This is similar to modern probabilistic SLAM formulations, as we will see in section 2.1.

Modelling of uncertainty by additive Gaussian noise The peculiar process by which measurements where made with the repeating circle licensed the application of the central limit theorem. The modelling of error curves as Gaussians distributions became justified. Eventually, even for small samples, the assumption that random variable are normally distributed became a very popular one, including in SLAM, for it leads to easy mathematical derivations and manipulations (e.g., the simple mechanics of marginalizations and conditioning with covariance and information matrices, see Schön and Lindsten (2011)).

Least Squares A consequence (or motivation) of last point assumption, is that $\hat{\mathcal{X}}$ can be obtained by nonlinear least squares optimization.

Poses/Positions as decision variables The last approach of geodetic adjustment we reported in 1.4.4 proposes to solve directly on the position of stations of interest, while previous methods tried to correct the errors. This simplifies the formulation of the joint pdf compared to Gauss' approach, as each measurement simply becomes a factor in the joint pdf. This view is inline with well established SLAM techniques of maximum likelihood estimation such as factor graphs (Dellaert and Kaess, 2017), GraphSLAM (Grisetti et al., 2020), etc.

On the other hand, from the view point of established least squares SLAM algorithms, we can point out some archaisms due to the young age probabilistic theory. We noticed that Laplace gave confidence intervals only in one dimension. Absent is the notion of correlation between variables. The topic is avoided by making frequent use of marginalisations in order to reason on only 1 variable. Whenever a correction affects many decision variables (e.g., due to baseline closure), the results are given as if marginalizations were done concurrently, not in a modern joint probabilistic way. Other difference with modern SLAM (and modern data fusion in general), the measured lengths of the baselines were considered exact by Delambre (and then by Laplace), which was not the case (Vincent, 1998) as unavoidable errors also accumulated during this delicate process. A modern and more rigorous view should attach some uncertainties to those measurements. To adhere to Delambre judgement about the relative precision of measurements (see quote in 1.4.1), the precision of the baseline lengths should be superior to the precision of the angles. More broadly, a single adjustment system that encompasses the baselines, the triangles, the astronomical latitudes (with biases) and the

ellipse parameters of the meridian is not proposed. They are treated separately by Legendre and Laplace.

1.5.2 Facets of early geodetic adjustment missing in SLAM

At this point, it may be tempting to establish a parenting link between geodetic adjustment and the modern prevailing view of SLAM. The scientific achievements of the early 19$^{\text{th}}$ century celebrated so far in this chapter would just be an inspiring perspective in that regard. Potentially, this could also evoke further thoughts on the shift from other approaches of state estimation, like the Kalman filters or the Rao-Blackwellized particle filters[27], back to the "initial" approach, i.e., the one involving the method of least squares.

However, this is not our view. We believe there are other fundamental differences between the 1820 geodetic adjustments and the established view of SLAM, and that actionable lessons can be learned here.

Firstly, let's recall the identification early in 1799 that the baseline closure only affects a part of the triangles (quote in 1.4.1), in this case those between the first and second baselines. This has been shown in the 1799 special commission document, and reiterated by Delambre in Basis, see subsection 1.4.1. The same aspect is known in SLAM, as quoted in the same section 1.4.2 (Chatila and Laumond, 1985), but as we will see in the next chapter those qualitative considerations have progressively become unused.

Secondly, the geometrical aspects of the adjustment problem lead Laplace to consider the cases of mismatch (in triangle and baseline closures) as hard constraints. It can be stated that Delambre shared this view since he spoke of restoring *harmony* of the geometrical laws, see 1.4.1. In addition, to solve those constraints, it was necessary to know where their violations come from, i.e., what are their causes. The causes were the unavoidable measurement errors. Therefore, the decision variables were the errors attached to the observations. Which is why this early probabilistic framework is often referred as the *theory of errors*, and Laplace even mentioned a *theory of corrections*, as explained in subsection 1.4.2. Any additional geometrical construct, like the distance from one extremity of the chain to the other (arc length), is considered to be a function of the errors. This is made clear in the quoted passage from *Essay* (Laplace, 1814a). Conceptually, and to our surprise, the angles of the triangle, the lengths of the sides of triangle, the geographical coordinates of the stations, and any other geometrical construct are not decision variables in the inference step. These elements are instead byproducts of the measurements errors, completed by structural knowledge of geometry. The uncertainty of the deduced geometric elements is also a function of the *a posteriori* precision of (some of) the error variables.

This view profoundly differs from the inference strategies employed in modern SLAM and in the method of variation of coordinates, mentioned in subsection 1.4.4. The decision variables, i.e., the object of inference, are generally the poses and the landmarks, not the errors emanating from the sensor imperfections. In the modern and celebrated view of SLAM, the measurements account for noisy geometric associations between decision variables: these are soft constraints that appear as factors terms in a joint distribution (similar to equation 1.1). In contrast, the overall Laplacian approach of geodetic adjustment appears to be of a quasi-deterministic nature: most functions are deterministic (trigonometric relations) decorated with probabilities, rather than a largely probabilistic one. To show that this claim has merit,

[27] Especially since filtering techniques are more recent inference concepts than the method of least squares.

and understand why the quasi-deterministic approach is a viable alternative, we propose a deeper investigation of Laplace conception in the next paragraphs.

1.5.3 The Laplacian doctrine: systems, causality and probability

Remark 1.3 (Disclaimer)**.** *The purpose of the subsection is to provide explanations about the Laplacian approach to concepts causal determinism, probabilities and uncertainty in general. The subject is not to agree or disagree with Laplace on the problems of existential aspects of the universe, on ontological questions etc. The goal here is to identify his way of thinking, in particular how the concepts mentioned are organized. This is done in order to see how it influences his approach to geodetic adjustment, and finally, by virtue of the similarities that we mentioned above between geodetic adjustment and SLAM, why all this is relevant in robotics.*

The philosophical formalization of Laplace's concepts on probability theory is given in the first words of *Philosophical essay on probabilities* (Laplace, 1814a). The text is only a few pages long and is no doubt the most discussed aspect of Laplacian thinking (far beyond geodesy). In this text, a causal deterministic view of the world is proposed. As a thought experiment, it is postulated that if an entity, at any point in time, possesses complete knowledge of (a) the state of all particles in the world at some point, and (b) all the forces in nature, then, nothing would be uncertain in the past nor in the future, provided that infinite computations (a.k.a. Laplace's computation) are made available to conduct the analysis. An entity, in the text, is an *intellect*. This idea is presented at the first pages of the essay, and is not mentioned later on. Nonetheless, it seems that most subsequent comments which often mention it as the Laplace *demon*, ignore the remainder of the essay, for a *philosophical essay on probabilities* should mostly talk about (relatively) dull considerations on probabilities. Besides, the deterministic view of the world was not a Laplacian specificity. Other scholars had already proposed the thought experiment. According to Van Strien (2014), "this seems to have been an idea that was widespread around the time that Laplace first expressed it in 1773, particularly in France". Van Strien points out that variations of the calculating intellect/demon argument were already given by D'Holbach, Condorcet, Diderot, Maupertuis and Boscovich. The last two names have already been mentioned in this document in regard to geodesic triangulations in the mid-18th century. Belgian geodesist Liagre also made, later, a similar statement when introducing probability calculus (Liagre, 1852, §1). Kožnjak (2015, 2022) shows that the first of such statement was actually given by Boscovich, decades before anyone. The real question to ask, in the case of Laplace's statement, is *what purpose serves this proclamation of causal determinism in a book on probabilities ?* Contextually, Laplace intellect/demon statement appears to be well-placed catch phrases in the beginning of an essay for the general public. But besides the opportunity of capturing the reader's attention and imagination, the main point of these first pages is precisely to pinpoint where the concept of probability lies. The word "probability" appears in the main text only once the words "cause", "knowledge", "ignorance", "system" have occurred, respectively, 6, 3, 4 and 3 times.

One might wonder how these words organize themselves with the notion of probabilities into a coherent framework. We believe it is now worthwhile to digress on the Laplacian mindset as: (a) it will be the one eventually used for practical purposes in this document, (b) it differs from the dominant so-called Bayesian perspective. The following explanation of Laplace views can be reasonably inferred by relying on historical context and other writings from the famous geometer.

Laplace, born in 1749, was deeply impressed by advances in understanding made in the prior century[28]. For him, the 17th century had been deemed as the one that "gives the most honor to human spirit" (Laplace, 1814a, p218). This suggests that the objective was to discover the foundational principles (e.g., Newton's laws, Kepler's laws), then use them to build systems to explain nature, packed with deterministic notions. In his book on the arc meridian measure, Guedj (2000, chap 11) calls it the "systematic spirit". Besides, the regularity of motions observed in astronomy lead to a cause-effect temporal mechanism, i.e., this is the causal physical determinism that was stated by many before him (Van Strien, 2014; Kožnjak, 2015, 2022). As a man of the late 18$^{\text{th}}$ century[29], Laplace moderates these expectations by declaring the full comprehension of all cause-effects relations in nature (the one achieved by the *intellect*) to remain off-limit for the human mind (*infinitely far*). However, this does not justify the abandonment of the causal deterministic view of the world. Probabilistic tools have tremendously helped Laplace in astronomy, particularly by disentangling slight deviations from models being caused either by observation errors or by new phenomena. Laplace's achievements in astronomy notably included a first order proof of the stability of the solar system, and a refined prediction of the motion of the moon (Laplace, 1796). He even derived from the perturbations of the motion of the moon a precise estimate (for the time) of earth flattening coefficient: 1/305 (Laplace, 1796, livre IV,chap 5), later refined to 1/301 (Laplace, 1820, p362). For him, the regularity observed in the motion celestial bodies, "without a doubt", happens in all things. Hence, systems (by interpreting, causal systems) should be built to account for the state of understanding for a given topic. This is the "spirit of systems", as opposed to the "systematic spirit". The "spirit of systems" was prevalent in the later 18th century period (Guedj, 2000)[30]. Those systems are quasi-deterministic as opposed to fully deterministic. This is precisely where his conception of probabilities comes: "they are related partly to our knowledge, partly to our ignorance". Unavoidable errors may appear once data (observations, events) are provided, as they induce a mismatch between the deterministic relations in the system (input from knowledge) and the data (e.g., the closures in geodetic adjustment). While errors are indeed to be feared, probabilistic tools can be pushed forward to bet against the errors, e.g., "*[...] the mass of Saturn is 3,512th part of that of the Sun. I find that there are 11,000 to bet against 1, that the error of this result is not a hundredth of its value*" (Laplace, 1814a, chap 8). Thus, the role of probabilities is to mathematically capture what is not deterministic, i.e. the exogeneous perturbations we ignore, or we chose to ignore in our model, and to offer quantitative analysis (most probable results, confidence bounds, etc.). The calculus of probabilities considers the errors as its primary variables of interest (hence the *theory of errors*), while the uncertainty affecting other variables can be deduced downstream, once the errors have been adjusted, using the deterministic relations (e.g., the interval confidence bounds for the length of a meridian).

The implication of that view is that the causal model is assumed, initially, from assumptions derived from domain knowledge, not from the data. In computer science, this paradigm

[28] His historical review of astronomy occupies a large part of his other book published: "System of the World" (Laplace, 1796).

[29] D'Alembert, mathematician and figurehead of the enlightenment, was a mentor of Laplace.

[30] Some other members of the *Académie des Sciences* took advantage of the Révolution power shift to organize society in systems (education, civil rights, territories, etc.). A political leader opposed to those costly projects, Marat (1791), complained about the system-mania of the scientists in a violent letter entitled *Charlatanisme Académique*, hence providing corroboration to the argument made.

has been championed and further developed by (Pearl, 2009), author of *Causality*. The data can only be interpreted throughout either an already established model, or a provisional one.

1.5.4 Causal assessment in geodetic adjustment and lack of formalization

Recall, from subsection 1.4.2, that the application of probability theory to geodesic operations by Laplace is addressed in two documents: the "Philosophical Essay on Probabilities" (Laplace, 1814a) and in the second and third supplements of "Théorie Analytique des Probabilités". In both instances, the description of the problem starts with the same sentence (Laplace, 1820, p531):

Causal Statement 1.1. *We determine the length of the great arc on the surface of the earth by a chain of triangles which rest upon a base[line] measured with exactitude.*

Given that causal statement 1.1 is the starting sentence of the supplement, this corroborates the fact that the model is described first. Furthermore, note that the statement is directional, i.e., baseline $\rightarrow$ (triangle 1 $\rightarrow$ triangle 2 $\rightarrow$ ... $\rightarrow$ triangle 115) $\rightarrow$ length of a great arc. This aspect is important: it is a causal statement that describes the structure of the geometry. It does not just declare that the baseline, the triangles and the arc length covary (or are linked) with each other. In other words, the statement is not associative. Neither is it a chronological statement. Indeed, recall from section 1.1 that the geodesic operations started in 1792 with both Delambre and Méchain covering a distinct portion of the meridian. The first baseline, Melun-Lieusaint, was only measured in 1798 (Guedj, 2000; Alder, 2003; Vincent, 1998), after most of the triangles were completed. In the beginning of "Philosophical Essay on Probabilities", analyzed in the previous subsection, Laplace speaks of cause and effect only in terms of chronology, using astronomy as the motivating example. When it comes to geodetic adjustment, it is crucial to note that causality is expressed in the order necessary for geometric construction. Indeed, without at least one baseline (pre-1798), the only data available were the angular measurements which are insufficient to construct any of the triangles.

In contrast, modern SLAM does not represent the construction of the environment in a causal way, rather in an associative one. Although it is known that, for a typical SLAM setting, a robot pose, denoted $\mathbf{x_i}$, is caused by the robot pose $\mathbf{x_{i-1}}$ at the previous time (plus some relative motion), the probabilistic tools used to described that relation only says that $\mathbf{x_{i-1}}$ and $\mathbf{x_i}$ correlate with each other[31]. Besides, the errors are not explicitly introduced as variables. In that regard, modern probabilistic SLAM shares similarities with the method of variation of coordinates described in 1.4.4. The whole starting point of the SLAM problem is in fact often presented as a joint distribution, whose variables (poses and landmarks) are sparsely connected. It has not always been like that, but we can explain why and other views are possible, as we will come to see in the next chapters.

Nonetheless, a reason why the mindset presented for geodetic adjustment, which includes causality, has not been popular in SLAM is the lack of formalization. Indeed, Laplace 'only' proposed a calculus for probabilities. How the notion of probabilities fits together with the wider system and causality viewpoint, as explained in subsection 1.5.3, is only explained in the text, for left to be parsed by the attentive reader. This is quite unfortunate, since, if we want this mindset to be an inspiration for SLAM, we need to be able to program the

[31] This remark encompasses several techniques such as Kalman filters, the Rao-Blackwellized, or Smoothing-and-Mapping and GraphSLAM techniques

full paradigm, not just the aspects that relate to probabilities. That was not explained by Laplace, nor was it, to our knowledge, by his immediate successors[32]. For a more actionable framework to approach a causal representation of the SLAM problem, we will eventually turn to Pearl's structural causal model (SCM) in this document (Pearl, 2009).

Thus, we can only surmise that since (a) enticing techniques for probabilistic inference were proposed, and (b) conversely, a mathematical language for causal thinking was not available, then the early tools of probabilistic inference eventually thrived and were built upon for estimation tasks and later on for statistics. However, there was no highlight of the importance of setting probabilistic calculus within a prior causal structure.

Mathematical frameworks to express and manipulate causation are now more advanced and can, presumably, be employed for a Laplacian quasi-deterministic view of SLAM, just like there was a Laplacian view of geodetic adjustment.

1.6 Discussion

In this first chapter, we have analyzed the methods for processing geodesic operations. Notably, a specific challenge concerns the determination of the arc length of the Paris meridian (1792-1799). It was related to the advent of the metric system and helped to push forward probability theory. Thanks to an investigation which included primary sources, the first contribution of this study is to synthesize and join two reflections together: (1) about the common patterns between this problem and SLAM (in particular the geodetic adjustment problem that deals with the closures), and (2) about how this problem accompanied fundamental theoretical development in probability theory. On (1), it is inspiring that Laplace (1820)'s "Théorie Analytique des Probabilités", spends a large part of the book on a SLAM-like problem. On (2), solid and enduring principles were laid out in a relative short period of time, thanks to the intertwined contributions, mainly, of Laplace, Legendre, Gauss and Delambre. In particular, Laplace's approach to geodetic adjustment was the most advanced at the time partly thanks to his mastery of the new mathematical tools, and partly thanks to his conception of probabilities within a quasi-deterministic causal conception of the world. Identifying the significance of this facet is the second contribution of this chapter.

The aspects related to causality, which Laplace did not attempt to formalize, are as underappreciated for the problem of geodetic adjustment as they are today for the problem of SLAM, if not outright non-existent. After 1820, it seems that no serious effort was made on that topic. As a result, the theory of errors, although it was partly developed in France, slowly declined despite dire needs for theoretical improvements, as made clear in the case of the French map of 1830 (figure 1.12). Without the causal doctrine backing the theory of errors, even the method of least squares became threatened at some point (as per §1.4.5). One can consider that the notions of causal reasoning are too metaphysical in nature and that they do not need to be formally represented, that these just concern human cognition[33].

[32] The *de-facto* reference in questions of geodesy after Laplace's death was Louis Puissant (Levallois, 1988; Berthaut, 1898a), a seasoned practitioner of geodesic operations, but not a contributor to probability theory. Puissant (1842, p410-414) tried to generalize the application of Laplace's results, given for a chain of triangles, to a network (the 1830 French map, see figure 1.12). However, this generalization was unwarranted, as chains and networks have inherently different causal models due to the latter containing loops.

[33] Recall that prior to the "Gauss-Laplace" synthesis, the definition of error curves and the combinations of equations was also non-formal (Alder, 2003; Stigler, 1986).

However, for the task of emulating spatial understanding close to human level, these notions may well be necessary, if only for being able to implement them in a computer which do not possess, a priori, these cognitive capabilities.

We will address the ramifications for SLAM progressively in the next chapters. An analysis of the issues of modern SLAM is first given (chapter 2). Notions of graphical representations are introduced, for our approach to SLAM relies heavily on them to reason about conditional independences between variables. In chapter 3, we use the structure of Causal Bayesian Networks to improve upon existing SLAM algorithms in order to exploit causal relationships along the robot trajectory. This is our first attempt to insert causal knowledge in a SLAM structure. In chapter 4, we decide to reformulate the SLAM problem using the paradigm of SCM. We also give some of our perspectives on the problem of spatial understanding.

In the 'epilogue' (chapter 5), we come back to the problem of geodetic adjustment, for which we propose a causal representation, using the tools developed in chapter 4.

Chapter 2

Successes & Anomalies of Modern SLAM

In the concluding sentences of the previous chapter, we brought up the delineation between the ability to solve the spatial state estimation problem and the wider notion of spatial understanding. Solving the spatial state estimation problem, for a robot, consists in inferring a consistent state of its close environment given local observations of it. In this chapter, we review the major solutions proposed to the state estimation problem, and we challenge the currently well-established mathematical representation of SLAM. In particular, we seek to investigate whether it is capable of transcribing common-sense considerations about SLAM. This common mathematical representation is the joint probability density function (pdf), whereby the hidden state is the set of robot poses and landmarks, and the evidence is the set of observations. There are other formulations, arguably less formal, that we do not include in this review. They are more tuned for specialized use cases while we try here to remain general. Such examples notably include variants of volumetric structure (occupancy maps, etc.).

The following principle is hypothesized: human beings and animals possess basic intuitions about their situation and evolution within their surroundings, e.g., to give just two generally agreed upon examples: one's current position is his/her/its past position plus the relative motion undergone within the subsequent time interval; if lost in a city or in nature, re-observing a recognizable landmark informs about the path travelled since the last observation of the same landmark; etc. Without doubt, these collections of judgements would presumably be useful in the emulation of human-level spatial understanding. But, even if we limit ourselves to the scope of state estimation, we expect these judgements to be incorporated in some ways, so that they participate in the adjustment process.

Thus, by investigating the suitability of the common formulation of SLAM in regard to these considerations, we will shape an answer about whether (a) we should use the existing mature formulation and build upon it, or (b) we should instead propose distinct SLAM statements in the next chapters.

The rest of the chapter is organized as follows: section 2.1 presents the existing approaches of SLAM, their fundamental assumptions, their possible graphical representation, and a method to the maximum likelihood estimate (MLE) of the robot trajectory and landmark positions. Similarities between (a) the MLE SLAM and (b) both Gauss formulation and the method of variations in geodetic adjustment problems will surely be noticed. We

then summarize, in broad outlines, the successes achieved by this paradigm. In section 2.3, we analyze the adequacy of this formulation along three axis: features, faithfulness and complexity. Finally, the last section discusses the takeaway points in order to articulate the work of the next chapters.

2.1 Principles of probabilistic SLAM

In the period named as the "classical age" (1986-2004) by Cadena et al. (2016), the first probabilistic formulations of SLAM were given: the (large) family of Kalman filters (KF) (Sola, 2014; Bourmaud et al., 2015; Ćesić et al., 2017, e.g.,), the Rao-Blackwellized Particle filters (RBPF) and maximum likelihood estimators (MLE). Under the MLE terminology, we broadly gather approaches such as graph-based SLAM (Lu and Milios, 1997; Thrun et al., 2005), Smoothing-And-Mapping (Dellaert and Kaess, 2006; Kaess et al., 2008, 2011) etc. Arguably, among those, the most popular have become those based on MLE in view of their versatility. However, historically, it can be considered that KF and RBPF reached maturity before MLE. Durrant-Whyte and Bailey (2006) mainly reviewed KF and RBPF approaches. A pivotal reference that presents these 3 approaches (and much more) is "Probabilistic Robotics" (Thrun et al., 2005).

The MLE approach is now considered well established (Rosen et al., 2021). Note however, that these approaches share the same fundamental conceptual basis. Indeed, the main difference between MLE and filtering techniques is the process by which the joint pdf is reduced[1]. For instance, KF systems typically use successive marginalizations so that only the most recent pose is kept, while the previous poses are integrated out. The starting point of the formulation, the full joint pdf, is however similar. As with geodetic adjustment, we emphasized three different probabilistic formulations (Table 1.1), but they were different. Two of them solve for the errors while the remaining one solves for the geographical coordinates (variation of coordinates). For SLAM, however, all other things being equal, the initial probabilistic formulation is almost similar for KF, RBPF and MLE. We will develop this formulation and its implications in the next subsection.

We mainly focus on the MLE framework because it is the more mature, i.e., algorithmically analyzed[2]. Overall, it is the most promising, for wider perception systems often build upon it (e.g., see Hughes et al. (2022)). This framework highlighted the key role of sparsity for performant SLAM and propelled the development of highly efficient libraries (Grisetti et al., 2020; Ila et al., 2017; Wang et al., 2018; Dellaert, 2012; Rosen et al., 2019; Kummerle et al., 2011).

2.1.1 Probabilistic formulation

It is not unusual to present the SLAM estimation problem directly in terms of a joint pdf, or sometimes, more expediently, as the underlying optimization problem. But as our purpose is to discuss the adequacy of the mature joint pdf formulation, it is necessary to collate it with a verbose expression of a SLAM problem. Therefore, throughout this chapter, we give

[1] In the R.A. Fisher statistical sense of the term reduction (Fisher, 1922): "the object of statistics is the reduction of data".

[2] This is again a reference on the term used in Cadena et al. (2016), where "algorithmic-analysis" age denotes the period after the "classical age".

verbose statements describing several small sized scenarios (toy problems) and we investigate the suitability of the targeted statement.

Box 3: SLAM Scenario 1

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot is fed sequentially 5 control inputs $\mathcal{U} \triangleq \{u_0, u_1, u_2, u_3, u_4\}$. A first landmark, denoted l_a is discovered (no prior knowledge) by a robot exteroceptive sensor at time t_1 and re-observed at t_4. A second landmark l_b is discovered at t_2 and re-observed at t_3. Robot motion and sensor measurements $\mathcal{Z} \triangleq \{z_0, z_1, z_2, z_3\}$ are affected by noise. We assume no adverse effect due to the robot inertia, false sightings, outliers, etc.

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}\}$ and the landmarks position $\mathcal{L} \triangleq \{l_a, l_b\}$.

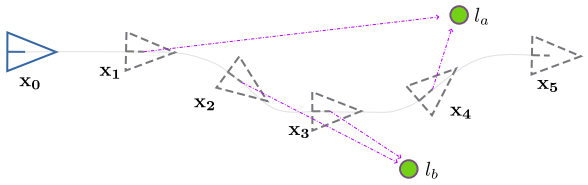


Figure 2.1: Illustration of the SLAM Scenario 1 presented in box (3).

In box 3 (figure 2.1), we consider toy scenario 1, which we will rely on to present the established probabilistic SLAM principles. For now, we abstract away the types of the elements in $\mathcal{X}$ and in $\mathcal{L}$ (e.g., resp. SE(2) and $\mathbb{R}^2$). $\mathcal{X}$ and $\mathcal{L}$ can be called the hidden state (estimation context), the decision variables (in optimization context), or the parameters.

The first step of the probabilistic SLAM approach is to translate the stated objective in formal probabilistic terms.

Assumption 2.1 (Probabilistic SLAM)**.** *The inference problem is expressed as the estimand* $p(\mathcal{X}, \mathcal{L}|\mathcal{U}, \mathcal{Z}, \mathbf{x_0})$*, referred as the* joint posterior pdf*.*

We are mainly interested in the mode of $p(\mathcal{X}, \mathcal{L}|\mathcal{U}, \mathcal{Z}, \mathbf{x_0})$, i.e., the most probable values of $\mathcal{X}$ and $\mathcal{L}$ given all known variables $\mathcal{U}$ and $\mathcal{Z}$. This is referred as the *maximum a posteriori* (MAP). The marginal posterior covariance over a subset of the decision variables may also be part of a query, depending on the task.

In order to detail the expression joint pdf for toy scenario 1, statistical assumptions must be formulated.

Assumption 2.2 (Probabilistic SLAM)**.** *The joint pdf is a product of estimand which involved a subset of the decision variables, they are named 'factors', or 'potentials'. Each potential establishes a statistical relation between the local hidden geometric entities (poses, landmarks) involved at the time of the perception or the control event. The noises affecting the measurements in each factor are considered mutually and individually independent.*

In toy scenario 1, the factors emanating from controls are $p(\mathbf{x_5}|\mathbf{x_4}, u_4)$, $p(\mathbf{x_4}|\mathbf{x_3}, u_3)$, $p(\mathbf{x_3}|\mathbf{x_2}, u_2)$, $p(\mathbf{x_2}|\mathbf{x_1}, u_1)$, $p(\mathbf{x_1}|\mathbf{x_0}, u_0)$; and the landmark relative observations give $p(z_0|l_a, \mathbf{x_1})$,

$p(z_1|l_b, \mathbf{x_2})$, $p(z_2|l_b, \mathbf{x_3})$, $p(z_3|l_a, \mathbf{x_4})$. This may also be referred, respectively, as the motion model and the observation model (Thrun et al., 2005).

The Bayesian perspective is adopted to link up the factors to the joint pdf $p(\mathcal{X}, \mathcal{L}|\mathcal{U}, \mathcal{Z}, \mathbf{x_0})$:

$$p(\mathcal{X}, \mathcal{L}|\mathcal{U}, \mathcal{Z}, \mathbf{x_0}) \propto p(\mathcal{U}, \mathcal{Z}|\mathcal{X}, \mathcal{L}, \mathbf{x_0}) p(\mathcal{X}, \mathcal{L}|\mathbf{x_0}), \tag{2.1}$$

where the term $p(\mathcal{U}, \mathcal{Z}|\mathcal{X}, \mathcal{L}, \mathbf{x_0})$ denotes the likelihood function, i.e. the probability of measurements and controls given an assignment $(\mathcal{X}, \mathcal{L})$, and the term $p(\mathcal{X}, \mathcal{L}|\mathbf{x_0})$ is the prior probability which captures prior knowledge. Assumption 2.2 gives the license to decompose the likelihood term $p(\mathcal{U}, \mathcal{Z}|\mathcal{X}, \mathcal{L}, \mathbf{x_0})$ into a product of factors.

Let $\varphi_k(\mathcal{X}_k)$ denote a factor which involves a subset of the decision variables $\mathcal{X}$. This subset is denoted as $\mathcal{X}_k$ and called *the scope of factor* φ_k. The likelihood term rewrites as:

$$\begin{aligned} p(\mathcal{U}, \mathcal{Z}|\mathcal{X}, \mathcal{L}, \mathbf{x_0}) \propto \varphi_0(\mathbf{x_1})\varphi_1(\mathbf{x_1}, l_a)\varphi_2(\mathbf{x_1}, \mathbf{x_2})\varphi_3(\mathbf{x_2}, l_b)\varphi_4(\mathbf{x_2}, \mathbf{x_3}) \\ \times \varphi_5(\mathbf{x_3}, l_b)\varphi_6(\mathbf{x_3}, \mathbf{x_4})\varphi_7(\mathbf{x_4}, l_a)\varphi_8(\mathbf{x_4}, \mathbf{x_5}) \ . \end{aligned} \tag{2.2}$$

In equation (2.2), a correspondence was made between the motion and observation models mentioned earlier and the factors: for any k, $\varphi(\mathbf{x_k}, l_{k'}) \propto p(z_{k''}|\mathbf{x_k}, l_{k'})$ and $\varphi(\mathbf{x_{k-1}}, \mathbf{x_k}) \propto p(\mathbf{x_k}|\mathbf{x_{k-1}}, u_{k-1})$ (Dellaert, 2021, §2.1).

As there is no other prior knowledge, $p(\mathcal{X}, \mathcal{L}|\mathbf{x_0})$ is flat and the MAP estimation problem of equation (2.1) also corresponds to the MLE (Cadena et al., 2016). This is the case for toy scenario 1 (Box 3). But even in problems that contain other prior knowledge on $\mathcal{X}$ and $\mathcal{L}$, these terms $p(\mathcal{X}, \mathcal{L}|\mathbf{x_0})$ can be viewed as additional likelihood factors (see Cadena et al., 2016, §2), and their causes of existence (measurement, prior, etc.) are irrelevant in the subsequent steps of the estimation process[3].

Notation simplification: Following the standard adopted in the literature, we repurpose the symbol $\mathcal{X}$ to cover all decision variables, i.e., both poses and landmarks. Since, in the factor graph paradigm, the distinction between controls and measurements has become nonessential, both of them are gathered in z (u does not appear anymore) labeled by the index k of the factor they are attached with.

All in all, defining the full set of factors as Φ, and its cardinality as $|\Phi|$, the following holds:

$$\phi(\mathcal{X}) = \prod_{k=0}^{|\Phi|-1} \varphi_k(\mathcal{X}_k). \tag{2.3}$$

In summary, the conjunction of assumptions 2.1 and 2.2 lead to the representation of probabilistic SLAM problems as a product of factors, aggregating terms stemming from measurements, control and prior knowledge. Simply put, we now have $p(\mathcal{X}, \mathcal{L}|\mathcal{U}, \mathcal{Z}, \mathbf{x_0}) \propto \phi(\mathcal{X})$, so that $\phi(\mathcal{X})$ is the mathematical object of choice to continue the estimation process. Typically, the MAP (or MLE) is found by maximizing the product of factors:

[3] When prior knowledge on the landmarks do exist (or more rigorously, on the landmark given the first pose $\mathbf{x_0}$), it is generally artificially reconsidered as a measurement, and subsequently treated as a factor. If there are no correlations between the prior knowledge of several landmarks, then each landmark receive a unary factor. For instance, see Dellaert and Kaess (2017, fig 1.2).

$$\mathcal{X}_{\text{MAP}} = \arg\max_{\mathcal{X}} \phi(\mathcal{X}). \tag{2.4}$$

According to Thrun et al. (2005, §11.10), the seminal introduction of this paradigm was done by Lu and Milios (1997). The pose-graph problem was addressed, i.e., a SLAM problem restricted to the robot trajectory, and where the factors typically come from alignment of laser range scans and/or proprioceptive sensors.

Remark 2.1. *As mentioned before, alternative approaches such as KF and RBPF have also used the joint pdf as the cornerstone to SLAM. For instance, rather than starting by a prior-likelihood factorization, as in (2.1), FastSLAM (Montemerlo et al., 2002, eq 4) uses Bayes rule to factorize the landmarks with the last pose, before decomposing the landmarks (using the assumption that the landmarks become independent conditioned on the trajectory and the measures). In contrast to the MLE technique, where virtually any information is transformed into a likelihood factor, these approaches maintain a richer representation of the joint pdf by preserving other types of conditional probability distributions (CPDs). Typically, the motion models of the form $p(\mathbf{x_i}|\mathbf{x_{i-1}}, u_{i-1})$ are not cast as likelihoods. But fundamentally, the key assumptions 2.1 and 2.2 of probabilistic SLAM remain valid for those approaches. Therefore, although the arguments developed on the next sections mainly discuss the MLE framework, the key assumptions of probabilistic SLAM are ultimately questioned.*

The next step consists in solving equation (2.4), i.e., how the MAP is inferred. However, we first give an overview on the possible graphical representations of this view of the SLAM problem. A valuable feature of graphical representation of joint distributions is their ability to visually convey conditional independences (Dawid, 1979; Pearl and Paz, 1985), revealing what assumptions are being made when posing the SLAM as a joint pdf (or what additional implications are entailed). This will be useful for the remainder of the document.

2.1.2 On the graphical representation of probabilistic SLAM

We split the existing graphical representations of joint pdfs into two classes: directed and undirected. As the latter can be inferred from the former, the directed graphical is presented first.

Definition 2.1 (Bayesian Network (BN))**.** *From (Koller and Friedman, 2009, def 3.5), (Darwiche, 2009, def 4.1).*

A BN for variables $V_1, \ldots, V_n$ is a pair $(\mathcal{G}, P)$, where:

- *$\mathcal{G}$ is a directed acyclic graph (DAG) over the nodes (vertices) $V_1, \ldots, V_n$.*
- *The joint pdf $P(V_1, \ldots, V_n)$ is expressed as a product of conditional probability distributions (CPDs), one on each variable in $V_1, \ldots, V_n$, i.e.,*

$$P(V_1, \ldots, V_n) = \prod_{i=1}^{n} p(V_i \,|\, \text{Pa}^{\mathcal{G}}_{V_i}), \tag{2.5}$$

where $\text{Pa}^{\mathcal{G}}_{V_i}$ is the set of V_i's parents in the DAG $\mathcal{G}$.

For nodes without parents (a.k.a. source nodes), $\mathrm{Pa}_{V_i}^{\mathcal{G}} = \emptyset$.

The interest of DAGs is that the independences in the underlying joint pdfs can be visually identified using a graphical criterion, called *directional separation*, or *d-separation* for short. This is what motivated the development of Bayesian Networks (Pearl, 1985, 2022).

Definition 2.2 (d-separation)**.** *, from Pearl et al. (2021, def 2.4.1).*

We say that a path between two nodes A and C of $\mathcal{G}$ is blocked by a set of nodes V_z if and only if:

1. *the path contains a chain of nodes $A \to B \to C$, or $C \to B \to A$, or a fork $A \leftarrow B \to C$ such that the middle node $B \in V_z$,* ***or****,*
2. *the path contains a collider $A \to B \leftarrow C$ such that neither the collision node B nor any of its descendants belong to V_z.*

If V_z blocks every path between two nodes A and C, then A and C are said to be d-separated *conditioned on V_z.*

If two variables A and C are d-separated in a graph $\mathcal{G}$ conditioned on a set $V_z = \{B\}$, then A and C are conditionally independent, given the set V_z, in the corresponding distribution P (Darwiche, 2009, theorem 1.2.4). We will note:

$$(A \perp\!\!\!\perp C|B)_{\mathcal{G}} \implies (A \perp\!\!\!\perp C|B)_P \tag{2.6}$$

The set of blocked nodes V_z will henceforth be colored in gray (e.g., nodes B in figure 2.3). In the probabilistic SLAM problem, the gray nodes generally correspond to the available measurements and controls ($\mathcal{U}$ and $\mathcal{Z}$). Figures 2.2 and 2.3 shows different patterns (chains, forks, colliders) and how d-separation applies in those cases.

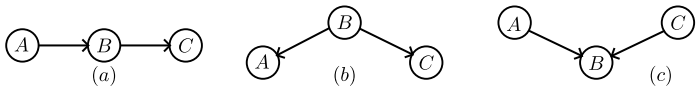


Figure 2.2: (a) chain, (b) fork and (c) collider patterns in Bayesian Networks. The set of blocked nodes is empty, $V_z = \emptyset$, therefore, given definition 2.2, $(A \perp\!\!\!\perp C|V_z)_{\mathcal{G}}$ is true only for the collider (c).

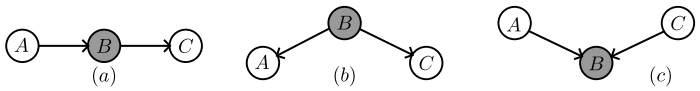


Figure 2.3: (a) chain, (b) fork and (c) collider patterns in Bayesian Networks, where the set of blocked nodes is $V_z = \{B\}$. The d-separation criterion leads to $(A \perp\!\!\!\perp C|V_z)_{\mathcal{G}}$ for cases (a), (b), but not for (c).

Remark 2.2. *Different authors attach different names to the graphical concepts presented above. A chain, a fork and a collider are called, resp., a causal/evidential trail (depending on the direction), a common cause and a v-structure in Koller and Friedman (2009) while, in Darwiche (2009) they are called a sequential valve, a divergent valve and a convergent valve. A path that is blocked (or not) is called, resp., an active/inactive trail in Koller and Friedman (2009), and a closed path (or not) in Darwiche (2009). In this document, for consistency, we use Pearl's denominations:*

- *"open colliders", such as node B on figure 2.2-(c), do not create a new path of influences between parents,*
- *"closed colliders", such as node B on figure 2.3-(c), do create a new path of influences between parents.*

Figure 2.4 shows the BN of toy scenario formed by the set of CPDs (the motion models and the observation models). This BN representation is often found to support KF and RBPF techniques for SLAM: see, e.g., Stachniss et al. (2016, fig 46.1), Montemerlo et al. (2002, fig 1), Durrant-Whyte and Bailey (2006, fig 4).

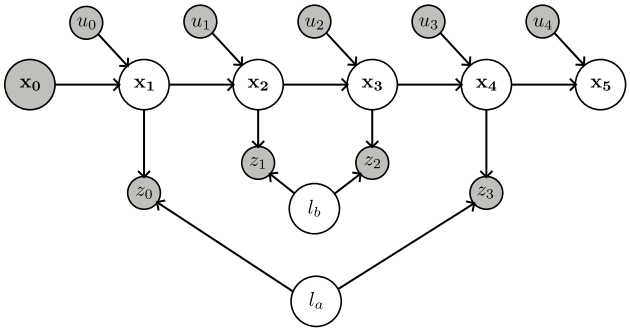


Figure 2.4: Bayesian Network of toy scenario 1 described in box (3). Grey nodes denote the observed variables.

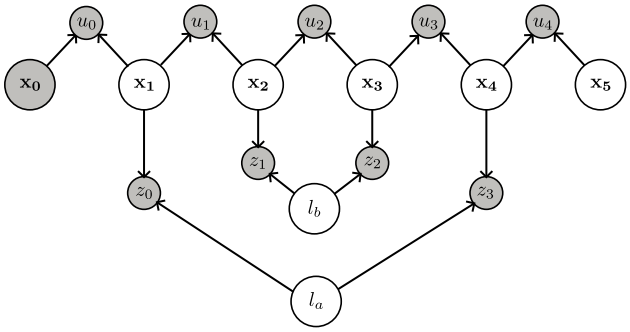


Figure 2.5: Bayesian Network of toy scenario 1 described in box (3) and where all CPDs are likelihood functions.

As mentioned before, in the MLE approach, all CPDs are cast into likelihood functions. This is represented by the BN of figure 2.5. Remark that, in this paradigm, there are no open colliders, and the latent variables are (sparsely) linked with each other through blocked colliders. Conceptually, the BN representation has become over-featured for this type of joint distribution. Furthermore, since any prior knowledge has also become a likelihood function, the latent variables have implicitly a flat prior, which is arguably not elegant in a BN. Consequently, from an aesthetic and visually effective standpoint, the factor graphs (bipartite graph in figure 2.6) are a more appropriate choice to represent the product of factors (2.3) for

MLE SLAM than a BN is (figure 2.5). The factor graphs, as the name suggests, are bipartite graphs involving nodes for variables (white circles) and nodes (black dots) for every factor uniting them in the joint pdf (2.1). This representation has been championed by (Dellaert and Kaess, 2006, 2017) and co-authors. The notion of 'blocked path' for factor graphs (as well as for other undirected representations such as Markov Random Fields) becomes simplified: a path between two nodes is blocked if a node in that path is blocked. The qualification of d-separation does not change: i.e., two nodes of the factor graphs are d-separated if all paths between them are blocked.

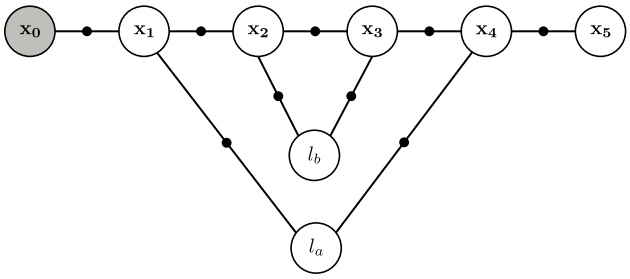


Figure 2.6: Factor graph representation of the toy scenario 1 described in box (3) and equation (2.2). Grey nodes denote the observed variables. Note that the observed variable $\mathbf{x_0}$ does not have to be represented explicitly. The factor between $\mathbf{x_0}$ and $\mathbf{x_1}$ would be replaced by a unary factor on $\mathbf{x_1}$.

Remark 2.3 (Are BNs causal ?)**.** *Although Bayesian Network are often built on the basis of a causal interpretation of a given problem, it remains nonetheless that BNs expose independence assumptions, not causal ones. This is made clear, e.g., in Pearl (2009, § 1.3) and Darwiche (2009, § 4.2). Suppose that a distribution P factorizes over the graph ($\mathcal{G}_1 : \quad A \rightarrow B$) and that $\mathcal{G}_1$ has moreover a sensible causal interpretation (e.g., A corresponds to temperature and B corresponds to thermometer). Then P also factorizes over ($\mathcal{G}_2 : \quad B \rightarrow A$) by application of the Bayes rule on the expression of P. In fact, $\mathcal{G}_1$ and $\mathcal{G}_2$ are equally valid in the context of BNs, although the latter goes against causal preconception.*

However, since building a BN based on causal judgements is more accessible, graphical structures that carry causal assumption (in addition to conditional independences) also exist, still in the form of DAGs: these are the Causal Bayesian Network and causal graphs, which will be addressed in subsequent chapters.

2.1.3 MLE estimator

To solve the MAP estimation problem in (2.4) as well as in geodetic adjustment, the standard way to proceed is to assume Gaussian additive dynamic and measurement noises, so that

$$\forall k \in (0, \dots, |\Phi| - 1), \quad \varphi_k(\mathcal{X}_k) \propto e^{-\frac{1}{2} h_k^\top (\mathcal{X}_k; z_k) \Omega_k h_k (\mathcal{X}_k; z_k)}, \tag{2.7}$$

where: h_k is the model function linking the decision variables to the observations, (similar to equations of conditions mentioned in chapter 1); Ω_k is the information matrix of each

k-th additive Gaussian measurement noise. For convenience, we abstract away h_k and Ω_k by introducing the residual function r_k:

$$r_k(\mathcal{X}_k;\ z_k) \triangleq \rho_k^\top h_k(\mathcal{X}_k;\ z_k), \quad \text{with} \quad \rho_k \rho_k^\top = \Omega_k,$$

where ρ_k can be obtained from Ω_k via Cholesky decomposition. Equation (2.7) becomes:

$$\forall k \in (0, \ldots, |\Phi| - 1), \quad \varphi_k(\mathcal{X}_k) \propto e^{-\frac{1}{2} r_k^\top (\mathcal{X}_k; z_k) r_k(\mathcal{X}_k; z_k)} \ . \tag{2.8}$$

Remark 2.4. *The process by which the MAP in eq (2.2) is solved is akin to Gauss method for the* most plausible values of the parameters *in his equation (1.1) published in 1809, and applied to geodetic adjustment in 1821-1825 (Bertrand, 1855). However, to reiterate our previous point, the major conceptual difference is that Gauss (and more so, Laplace in "Théorie Analytique des Probabilités") would have solved for the measurement errors (and* did *it in the case of geodetic adjustment), neither for the poses nor for landmarks.*

In that regard, the standard MLE solution to SLAM is closest to the method of variation of coordinates (subsection 1.4.4), where the core assumption 2.2 of probabilistic SLAM is also relevant. Indeed, measuring an angle between two geodesic stations from a third one amounts to create a statistical relation between the three stations (modeled by a new equation of condition). Besides, it was also remarked that, since there are often less geodesic stations, i.e., triangle vertices, than measurements, fewer unknowns lead to a smaller system of equations (i.e., a smaller matrix system). It was thought that it would simplify the calculations when the number of chain closure is large (Tardi, 1934). Similarly, in SLAM, $|\mathcal{X}|$ is smaller than the number of factors $|\Phi|$, especially when the number of loop closures is large.

Taking the negative logarithm of (2.2) turns the maximization (2.4) into the least squares optimization:

$$\mathcal{X}_{\text{MAP}} = \arg\min_{\mathcal{X}} \sum_{k=0}^{|\Phi|-1} ||r_k(\mathcal{X}_k;\ z_k)||_2^2, \tag{2.9}$$

where $|\Phi| = 9$ in the toy scenario 1 Box 3.

In cases when the $r_k(\mathcal{X}_k;\ z_k)$ functions are linear in $\mathcal{X}_k$, the problem could be directly solved by linear algebra techniques. Unfortunately, the functions $r_k(\mathcal{X}_k;\ z_k)$ is nonlinear in most, if not all, practical applications. Notably, the pose of the robot is generally expressed in non-Euclidean manifolds.

To remedy this, a Taylor expansion is done/performed around a linearization point $\bar{\mathcal{X}}$, so that each factor is linearized around its linearization point $\bar{\mathcal{X}}_k \subseteq \bar{\mathcal{X}}$ along:

$$r_k(\bar{\mathcal{X}}_k \oplus \delta\mathcal{X}_k;\ z_k) \approx r_k(\bar{\mathcal{X}}_k;\ z_k) + \left.\frac{\partial r_k}{\partial \mathcal{X}_k}\right|_{\bar{\mathcal{X}}_k} \delta\mathcal{X}_k, \tag{2.10}$$

with $\delta\mathcal{X}_k$ a small increment expressed in the tangent space around $\bar{\mathcal{X}}_k$.
The matrix $\mathbf{J}_k \triangleq \left.\frac{\partial r_k}{\partial \mathcal{X}_k}\right|_{\bar{\mathcal{X}}_k}$ will be referred to as the Jacobian matrix. It can be broken down as a concatenation of blocks relating to the partial derivatives of r_k w.r.t. each element of $\mathcal{X}_k$, see example 2.1.1.

Example 2.1.1. *Let us instantiate the approximate residual for factor φ_1:*

$$r_1(\bar{\mathbf{x}}_\mathbf{1} \oplus \delta\mathbf{x}_\mathbf{1}, \bar{l}_a \oplus \delta l_a;\ z_1) \approx r_1(\bar{\mathbf{x}}_\mathbf{1}, \bar{l}_a;\ z_1) + \mathbf{J_1} \begin{bmatrix} \delta\mathbf{x}_\mathbf{1} \\ \delta l_a \end{bmatrix}.$$

The Jacobian matrix $\mathbf{J_1}$ can be broken down as follows:

$$\mathbf{J_1} = \begin{bmatrix} \mathbf{J}_{\mathbf{1},\mathbf{x_1}} & \mathbf{J}_{1,l_a} \end{bmatrix} \triangleq \begin{bmatrix} \frac{\partial r_1}{\partial \mathbf{x_1}}\Big|_{\bar{\mathbf{x}}_\mathbf{1}, \bar{l}_a} & \frac{\partial r_1}{\partial l_a}\Big|_{\bar{\mathbf{x}}_\mathbf{1}, \bar{l}_a} \end{bmatrix} \tag{2.11}$$

Commonly, such optimization problems are solved by iteratively updating the state $\bar{\mathcal{X}}$ after the optimal increment of $\delta\mathcal{X}$ is found in the local tangential space. There are several methods to update the iterative step between iterations[4] (e.g., gradient descent, Gauss-Newton, Levenberg-Marquardt, etc.). To expedite the presentation of the principles, we hereinafter place ourselves in a standard Gauss-Newton scheme.

The incorporation of the Taylor expansion (2.10) into (2.9) yields:

$$\delta\mathcal{X}_{\text{MAP}} = \arg\min_{\delta\mathcal{X}} \sum_{k=0}^{|\Phi|-1} ||r_k(\bar{\mathcal{X}}_k;\ z_k) + \mathbf{J_k}\ \delta\mathcal{X}_k||_2^2. \tag{2.12}$$

This least squares optimization problem (2.12) can be cast into the matrix form:

$$\delta\mathcal{X}_{\text{MAP}} = \arg\min_{\delta\mathcal{X}} \|A\delta\mathcal{X} - b\|_2^2, \qquad \text{with}\ \ b \triangleq [-r_0^\top, \dots, -r_{|\Phi|-1}^\top]^\top. \tag{2.13}$$

The matrix A is composed of blocks $\mathbf{J_k}$ and is generally rectangular sparse with $|\Phi|$ sparse block rows and $|\mathcal{X}|$ sparse block columns. The pattern of A and b for the scenario box 3 is given in figure (2.7).

$$A = \begin{bmatrix} \mathbf{J}_{0,\mathbf{x}_1} & & & & & & & \varphi_0 \\ \mathbf{J}_{1,\mathbf{x}_1} & \mathbf{J}_{1,\mathbf{x}_2} & & & & & & \varphi_1 \\ & \mathbf{J}_{2,\mathbf{x}_2} & \mathbf{J}_{2,\mathbf{x}_3} & & & & & \varphi_2 \\ & & \mathbf{J}_{3,\mathbf{x}_3} & \mathbf{J}_{3,\mathbf{x}_4} & & & & \varphi_3 \\ & & & \mathbf{J}_{4,\mathbf{x}_4} & \mathbf{J}_{4,\mathbf{x}_5} & & & \varphi_4 \\ \mathbf{J}_{5,\mathbf{x}_1} & & & & & \mathbf{J}_{5,\mathbf{l}_\mathbf{a}} & & \varphi_5 \\ & \mathbf{J}_{6,\mathbf{x}_2} & & & & & \mathbf{J}_{6,\mathbf{l}_\mathbf{b}} & \varphi_6 \\ & & \mathbf{J}_{7,\mathbf{x}_3} & & & & \mathbf{J}_{7,\mathbf{l}_\mathbf{b}} & \varphi_7 \\ & & & \mathbf{J}_{8,\mathbf{x}_4} & & \mathbf{J}_{8,\mathbf{l}_\mathbf{a}} & & \varphi_8 \end{bmatrix} \qquad b = \begin{bmatrix} -r_0 \\ -r_1 \\ -r_2 \\ -r_3 \\ -r_4 \\ -r_5 \\ -r_6 \\ -r_7 \\ -r_8 \end{bmatrix}$$

Figure 2.7: Pattern of the sparse matrix A and vector b of the toy scenario 3.

Problem (2.13) can be solved efficiently via sparse matrix implementations such as sparseQR or CHOLDMOD (sparse Cholesky). Important points should be raised on the implications of transforming the problem (2.9), where a global $\mathcal{X}_{\text{MAP}}$ is sought, to the problem of finding of local increment $\delta\mathcal{X}_{\text{MAP}}$.

[4] In Legendre (1805)'s original least squares publication (as per § 1.3.1), the geometer only explicitly linearized once.

First, this iterative process is applied until some convergence condition is reached. Secondly, the success of that operation is sensitive to the initialization point chosen to bootstrap the iterative least squares. One naive choice for the initialization point is to integrate the odometry measurements (dead reckoning). However, on large maps, the drifting effect leads to the initialization point being out of the basin of attraction of the true MAP. Carlone et al. (2015) show that solving a relaxed version of the optimization problem with the technique of "Chordal Initialization" (Martinec and Pajdla, 2007) leads to compelling results on pose graph problems. Thus, a popular approach consists in first solving the SLAM estimation on a relaxed problem, then is using that result to bootstrap the iterative least squares process.

This concludes our brief overview of the MLE approach for the SLAM problem. We first highlighted the fundamental assumptions that underpin probabilistic SLAM, which are in our view common to KF, RBPF and MLE SLAM. We then focused specifically on how to turn the probabilistic formulation and associated the joint posterior, to the BN and factor graph representation and consequent least squares optimization problem.

2.2 Achievements of probabilistic SLAM

Recall that the formalism describes in 2.1.1 was established in the "classical age" of SLAM. It has been dramatically enhanced in the subsequent in the "algorithmic-analysis" age, to stick to Cadena et al. (2016) terminology.

Many successes and significant improvements have since been achieved by MLE SLAM[5], supported by a large body of literature.

A first point concerns the improvements in solvers efficiency and their diversity: a number of libraries were developed that maximally exploits linear algebra subroutines in modern CPUs (Kummerle et al., 2011; Ila et al., 2017; Grisetti et al., 2010a, 2020; Hess et al., 2016). Most of these implementations rely under the hood on efficient and mature decompositions of sparse matrices (Chen et al., 2008; Davis, 2011) provided in the SuiteSparse package. As SLAM is fed by measurements coming sequentially over time, it benefits from implementing an incremental algorithm. Some approaches were proposed in that vein in order to update only a limited portion of the matrix systems as new measurements become available (Kaess et al., 2008; Polok et al., 2013; Wang et al., 2018). But matrix-based solvers have their alternatives. Additional insights are exploited by graphical model approach such as the Bayes tree (Kaess et al., 2010) for iSAM2 (Kaess et al., 2011), or the clique tree (Paskin, 2002; Pinies et al., 2012). These structures are useful tree-like graphical representation of the variable elimination algorithm (Koller and Friedman, 2009, chap. 9) of (2.3).

Remark 2.5. *As per §2.1.2, factor graphs and BN are a graphical representation of an underlying joint pdf, exhibiting the sparse statistical relationships between variables. In contrast, its corresponding clique tree (a.k.a. junction tree) represent graphically the exact inference process by which belief propagates between cliques (clusters) of variables (by so-called 'message passing'). The general framework of probabilistic graphical models provides algorithms and techniques to build clique trees from BN or factor graphs, and use them for inferential tasks. Next chapter 3 will be the occasion to describe various mechanics concerning clique tree structures: how to build them, how to pass messages, etc.*

[5]Some points in the list also applies well to other variants of probabilistic SLAM, such as EKF/UKF formulations (Ćesić et al., 2017; Bourmaud et al., 2015).

This graphical model approach to MLE SLAM solvers has extended use-cases, e.g., for predictive control (Mukadam et al., 2018; Yang et al., 2021). The library GTSAM (Dellaert, 2012) implements the iSAM1, iSAM2 approaches and extensions.

A major theoretical improvement compared to early SLAM solutions such as described in Thrun et al. (2005) (where positions and rotations where typically decoupled) concerns the widespread utilizations of Lie groups to describe the space of admissible robot poses. Indeed, SLAM decision variables can often be described as elements of special Euclidean Groups such as SE(2), SE(3). Optimization is then conducted on those smooth manifolds (Chirikjian, 2009; Grisetti et al., 2010b; Barfoot, 2017). Valuable tutorials have been produced to democratize these mathematical concepts to roboticists by Barfoot (2017); Solà et al. (2021); Deray and Solà (2020). Additionally, new Lie groups were proposed, in particular the group of "extended pose" (Barrau and Bonnabel, 2017) which gathers pose and instantaneous velocities into a single decision variable. This has proved useful notably for IMU processing (Brossard et al., 2022).

Deploying SLAM systems in real environment is challenging not only because of the sensor noise characteristics but also because of false sightings and the data association uncertainty (due, say, the imperfect capacity to recognize a previously visited place). Several algorithms to reject outliers were incorporated in the optimization process of SLAM systems, which can perform well even under many false loop-closures (Sunderhauf and Protzel, 2012, 2013; Agarwal et al., 2013; Bai et al., 2018; Yang et al., 2020). Robust data association schemes have also been proposed for factor graphs (Hsiao and Kaess, 2019; Doherty et al., 2019).

Previously mentioned challenges find themselves compounded in collaborative SLAM systems, where robots are supposed to work jointly, exchange information, to model their environment. However, remarkable solutions have been proposed in what seems, in our view, the most difficult experimental and algorithmic setting, see Cunningham et al. (2013); Paull et al. (2014); Cieslewski et al. (2018); Choudhary et al. (2016); Ortiz et al. (2022).

Finally, since SLAM systems may be deployed in life critical applications and highly regulated environment, there is a requirement to ensure proper behavior of the solvers in view of these challenges. Nonlinear least squares algorithms can be trapped in a local minimum, or can fail to converge. In response to this need, a set of certifiable algorithms have been proposed to offer global guarantees on the estimate result (Rosen et al., 2019; Holmes and Barfoot, 2022; Fan et al., 2020; Tian et al., 2021).

However, despite these achievements, probabilistic SLAM (including MLE SLAM) has profound issues, if not anomalies. In the next section, we identify these points in view of proposing improvements.

2.3 Analysis of current estimation issues

We now point out some problems in this currently dominant approach along three axes: reasoning features, faithfulness and complexity. The first axis presents several basic aspects of spatial reasoning not promoted by the probabilistic formulation. The second axis addresses inconsistencies and dilemma in the pattern of conditional independences when a loop-closure shows up. Finally, the third axis discusses general performance challenges faced by solvers of MLE SLAM.

2.3.1 Reasoning features

The first axis of issues we explore are the lack of *innate* spatial reasoning features of probabilistic SLAM when challenged with simple problems. We investigate the capabilities of MLE SLAM, through its graphical representations, to cope with what we consider to be important embedded mechanisms of spatial intelligence. These four features will be labeled by symbols $\mathfrak{f}1, \mathfrak{f}2, \mathfrak{f}3, \mathfrak{f}4$. We go through them one by one.

$\mathfrak{f}1$· **Explaining missing data** There is no built-in mechanism able to account for missing data. If a robot travels roughly in a loop shaped trajectory and overestimates the travelled path, then when it reaches the area where it is *supposed* to get a loop closure observation relating to a place already visited, no data enters the probabilistic SLAM. This is the scenario considered in box (4), illustrated in figure 2.8, and corresponding to the factor graph shown on figure 2.9.

Box 4: SLAM Scenario 2

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot is fed sequentially 6 control inputs $\mathcal{U} \triangleq \{u_0, u_1, u_2, u_3, u_4, u_5\}$. The path travelled corresponds to a loop. The robot is able to recognize a place already visited located in front of it. Dead-reckoning estimates the robot position to be at $\mathbf{x_6}$ close to $\mathbf{x_1}$. However, no place recognition measurement is made. We assume no adverse effect due to robot inertia, false sightings, outliers, etc.

Objective: The inference problem consists in estimating the robot pose sequence $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}, \mathbf{x_6}\}$.

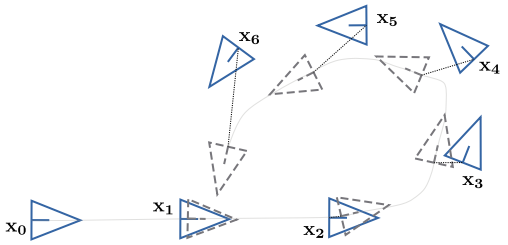


Figure 2.8: Toy Scenario 2 from Box (4). The state estimation has positioned the estimates of $\mathbf{x_6}$ and $\mathbf{x_1}$ close to each other, while their ground truth values (in blue) are still far away. It is suggested by the scenario that a loop closure would have been made by the place recognition module only if the distance $(\mathbf{x_1}, \mathbf{x_6})$ were as short as estimated. The missing data feature states that the absence of loop-closure measurement should be explained away to help update the state estimation.

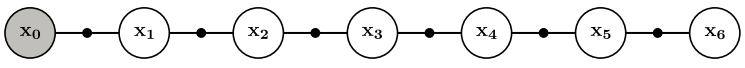


Figure 2.9: Factor graph representation of toy scenario 2 from box (4).

The robot does not get data because of either (a) a random pattern recognition or sensor failure; or (b) because of a tangential drift of the estimate forward while the place already visited ($\mathbf{x_1}$) has not yet been reached. A 'missing data' feature would draw upon a mechanism explaining the causes of missing data, in order to then correct the robot state estimation. Without this kind of feature, many opportunities for better adjustment are lost. For instance, in human cartography, if one asks what is the most significant "loop-closure" in History, in terms of the amount of map correction, it is certainly the circumnavigation of the Magellan expedition (or what was left of it) in 1519-1522. That famous loop-closure, to keep with the SLAM terminology, occurred when joining Asia from Europe by a longer than expected journey in the Pacific Ocean. Our interest with this specific example stems from the fact that cartographic corrections were made to explain away why the crew did not see any land while still navigating the ocean. It was a reaction to a missing data problem. Indeed, an analysis of the records of this expedition can show that the map were corrected days before making any land sighting. We provide the details and supporting material on this historical fact in appendix A.

ʃ2· **No backward influence of odometry or control** Consider toy scenario 1 (box (3)), with associated factor graph in figure 2.6. Suppose a control u_5 is sent to the robot, moving it to a new pose $\mathbf{x_6}$. No landmark re-observation is made at $\mathbf{x_6}$. Since the effect of inertia is willfully ignored, we expect no backward influence from the motion noise. In other words, our representation system should interpret that no updated adjustment is necessary, so no new path of statistical influence should be created. However, according the updated factor graph figure 2.10, and the rules of d-separation, the factor between x_6 and x_5 opens a path of influence to all other nodes in the factor graph. One might wish that the independence between those nodes could be exploited, if anywhere, right at the level of the graphical representation.

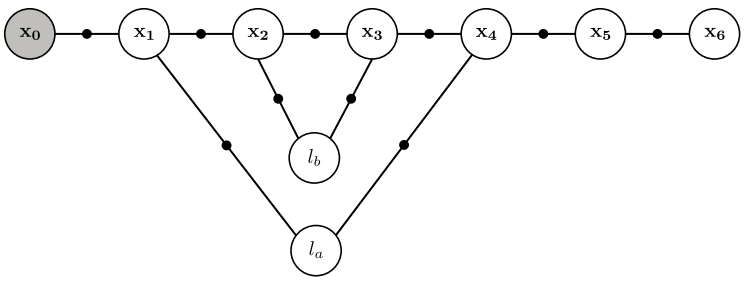


Figure 2.10: Factor graph of figure 2.6, scenario box (3), with a new pose $\mathbf{x_6}$.

Nonetheless, while the factor graph fails to exhibit this property, the updated Bayesian network figure 2.11 can cope with the fact that $u_5 \perp\!\!\!\perp \mathbf{x_5}$ due to node $\mathbf{x_6}$ being a collider. Consequently, the BN shows that the incorporation of $p(\mathbf{x_6}|u_5, \mathbf{x_5})$ in the joint pdf must not trigger a new adjustment.

However, that representational win of the Bayesian Network 2.11 over the factor graph 2.10 is short-lived. Suppose, with each control event u, that we also incorporate an odometry observation which measures the undergone motion. Intuitively, the feature of "No backward influence of odometry and/or control" should be maintained, but the

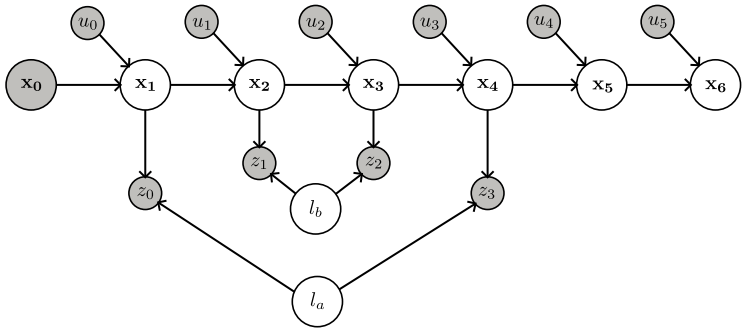


Figure 2.11: Bayesian network of figure 2.4, scenario box (3), with a new pose $\mathbf{x_6}$.

question remains: can we exhibit that with the Bayesian network ? One possibility is to consider the odometry z_{u_5} as a conditional pdf of the form $p(z_{u_5}|\mathbf{x_5}, \mathbf{x_6})$. Another possibility is to consider the odometry z_{u_5} as if it is an additional control $p(\mathbf{x_6}|z_{u_5}, \mathbf{x_5})$. Figure 2.12 shows the truncated Bayesian networks for the two options.

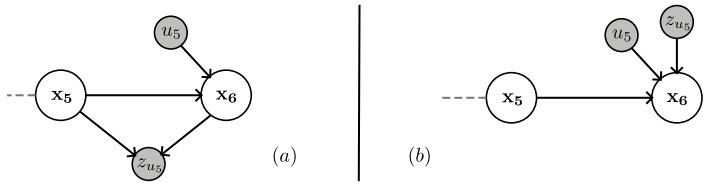


Figure 2.12: Design issues when a control u_5 and an odometry measurement z_{u_5} are both available. The rest of the Bayesian network (ancestors of $\mathbf{x_5}$) is not represented for space reasons (refer to figure 2.11). Sub-figure (a) presents a contemplated solution with the conditional pdf $p(z_{u_5}|\mathbf{x_5}, \mathbf{x_6})$, (b) presents a contemplated solution with the conditional pdf $p(\mathbf{x_6}|z_{u_5}, \mathbf{x_5})$.

However, these two options have problems. Figure 2.12-(a), corresponds to the incorporation of a likelihood function which is equivalent to a black dot in a factor graph, and we discussed above that it does not expose the independence we want. The second option, figure 2.12-(b), is not satisfying either, because it implies that the control u_5 and the odometry measurement z_{u_5} are mutually independent: $u_5 \perp\!\!\!\perp z_{u_5}$. That is obviously false. In fact, option (b) is worse than (a) since it creates an unruly independence in the representation.

ƒ3· **No influence of first landmark discovery** Whenever a new landmark for which there is no prior is discovered, the graphical representation should show that no correction is needed. Indeed, a new element in an unexplored part of the map simply does not inform about past decision variables. Furthermore, the noise characteristics in the landmark observation model are irrelevant. The reasoning behind this feature has been long acknowledged (Chatila and Laumond, 1985, §1.3):

> As long as only new parts of space are discovered, there is no possibility to find out and correct any errors, in the model being built, that are due to robot position drifts.

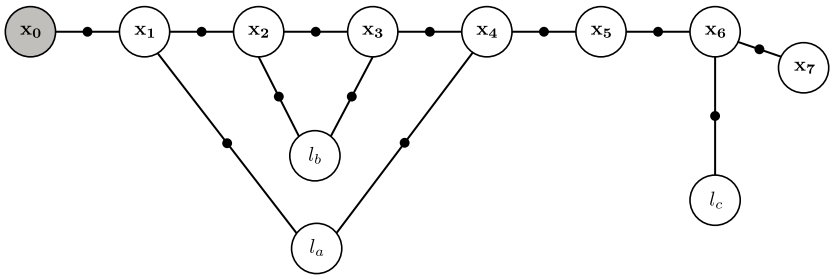


Figure 2.13: Addition of a new landmark

Expanding on the toy scenario box 3, a new landmark l_c is discovered, and the factor graph is shown on figure 2.13. Feature ƒ3 is not exhibited as the graph implies an active path of influence between l_c and $\mathbf{x_7}$, and between l_c and $\mathbf{x_5}$ (for instance). While it is expected that l_c be influenced by $\mathbf{x_5}$, the trivial qualitative consideration is that $\mathbf{x_5}$ (or $\mathbf{x_6}$, $\mathbf{x_7}$ etc.) should not be influenced by l_c. This report would be the similar for Bayesian network with the likelihood function $p(z_{l_c}|\mathbf{x_6}, l_c)$.

Since the potentials of landmark observation (i.e., the probabilistic observation model) are assumed of the form $p(z|\mathbf{x}, l)$, whether we are dealing with a discovery or a re-observation, Paskin (2002, §3.3) notes that we cannot resort to a graphical criterion to avoid computations. Therefore, the best strategy seems to be to detect the fact that a landmark is new, and via an ad-hoc *if-statement*, prevent any unnecessary process of adjustment. Another (no-)strategy consists in doing the moot computations anyway, which are inconsequential for the result, as proven by Paskin (2002, appendix C) for Gaussian densities. In our view, the issue might well be solved by, precisely, not considering the likelihood $p(z|\mathbf{x}, l)$ as the correct potential.

ƒ4· **No influence on the first variable in a loop** Another judgement concerns the scope of adjustment after loop closures. If the 'loop' is only made of relative measurements, then the scope of corrections can be limited to the variables within this loop. We have seen in chapter 1 that Delambre observed that the baseline closure only prompted to correct the triangle in-between the two baselines in the chain of triangles (see 1.4.1). See also Chatila and Laumond (1985, §3.5) for (perhaps) the first formulation of that axiom in SLAM. This judgement is not exhibited by the graphical models presented so far. A less known extension of that judgement concerns the first variable in a graphical loop. Upon careful examination, the first variable involved in a SLAM loop should not be influenced by the corrections carried out by the closure[6]. Although we consider this aspect to be intuitive once stated, we have not found any acknowledgement nor discussion on this specific point in the literature. Let us examine this on an example.

Continuing on scenario box 3 at the step of figure 2.13, we consider the addition of a new pose $\mathbf{x_8}$ and the re-observation of landmark l_c from $\mathbf{x_8}$. The corresponding factor graph is shown on figure 2.14.

This feature should vouch for the fact that $\mathbf{x_6}$, which is the first variable in the graphical loop $\mathbf{x_6}$—$\mathbf{x_7}$—$\mathbf{x_8}$—l_c, is not influenced by the loop-closure induced by the re-observation

[6]It is here still considered that there are no adverse effect of inertia, etc.

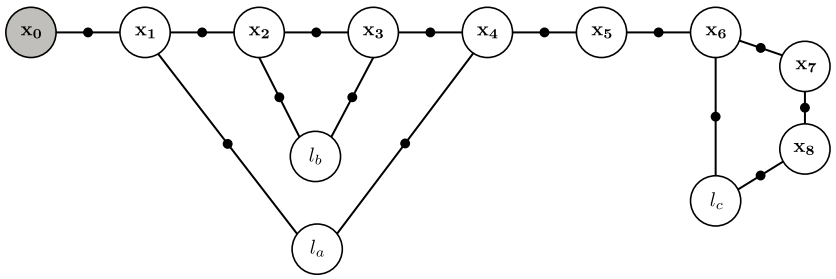


Figure 2.14: Incorporation of $\mathbf{x_8}$ and l_c re-observation to the toy scenario presented box 3, and continued from figure 2.13.

of l_c. The same reasoning should apply to similar situations in pose-graph problems.

On the motivation of the discussion

The reader might wonder, why bother mentioning these points? Indeed, as stated by property (2.6) a non-blocked path in the graph does not necessarily imply that the corresponding conditional independence is nonexistent in the distribution. First, as already stated, since the larger scientific goal is to emulate spatial understanding, it is reasonable to explore whether these basic notions could be implemented at the lowest level (representation), and not be a property exhibited at the optimization level (post-hoc). Second, although points f1, f2, f3 and f4 are not exhibited by the available graphical representation, they however imply distinct conditional independences which may jeopardize the faithfulness of those SLAM systems. This is explored in the next subsection.

2.3.2 Unfaithfulness

Except for feature f1 (missing data), the issues explored in the above subsection 2.3.1 fall in the category of unexploited statistical independences. In layman terms, for a given multivariate statistical problem, not exploiting a relation of (conditional) independence may be seen just as an inefficiency, or alternatively as a missed opportunity to refine such model. For the higher purpose of spatial intelligence, one might nonetheless start to see these issues as irrecoverable in that they impair the capacity to emulate reasoning about the environment. Perhaps this is correct, but this would require additional discussion and this is not our angle in this chapter. These issues above, by themselves, do not directly cast doubt on the whole estimation process. However, there is still a strange implication which we will refer as the 'loop-breaking' behavior.

Recall that probabilistic graphical models (such as Bayesian Networks, factor graphs, Markov Random Fields) depict patterns of conditional independences in the underlying joint pdf. Conceivably, if a probabilistic formulation supports (conditional) independences that do not exist in the 'real' problem it is supposed to represent, then that formulation becomes unfaithful.

We highlight two of such problems, denoted u1 and u2, below.

u1· **Unwarranted loop-breaking** The previous feature issue f4, which pertains to the non-backward influence on the first variable in the loop, is unfortunately not inconsequential. As a motivating example, consider the problem described in the toy scenario box (5),

whose factor graphs and BN representations are respectively shown on figure 2.6 and 2.4.

Box 5: SLAM Scenario 3 (pose-graph)

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot is fed sequentially 7 control inputs which drive it sequentially from $\mathbf{x_0}$ to $\mathbf{x_7}$. Consider that the last 5 controls move the robot about a loop. At $\mathbf{x_7}$, the place $\mathbf{x_1}$ is recognized, and the relative situation between both poses is measured. Robot motion and sensor measurements are corrupted by noise. We assume no adverse effect due to the robot inertia, false sightings, outliers, etc.

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}, \mathbf{x_6}, \mathbf{x_7}\}$.

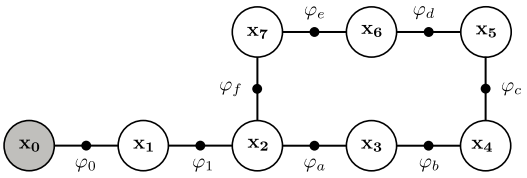


Figure 2.15: Factor graph representation of the third scenario presented in box (5).

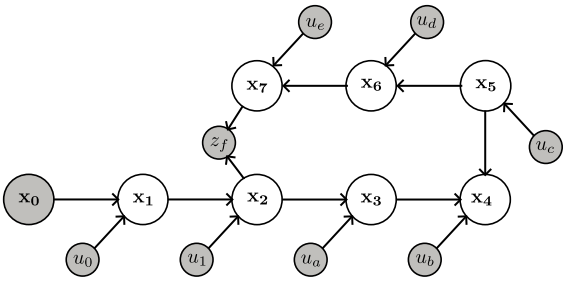


Figure 2.16: Bayesian network representation of the third scenario presented in box (5).

Let the measurements in the loop be denoted as $u_a, u_b, u_c, u_d, u_e, z_f$. Likewise, the corresponding potentials are denoted $\varphi_a, \varphi_b, \varphi_c, \varphi_d, \varphi_e, \varphi_f$. Suppose firstly that $\mathbf{x_5}$ were given to us. Node $\mathbf{x_5}$ becomes greyed in figures 2.15 and 2.16. This breaks the flow of influence in the loop, which is expected. The path of influence between $\mathbf{x_6}$ and $\mathbf{x_4}$ still exists, but cannot go 'through' $\mathbf{x_5}$, it is instead a simple chain ($\mathbf{x_6}$—$\mathbf{x_7}$—$\mathbf{x_2}$—$\mathbf{x_3}$—$\mathbf{x_4}$). Now, consider instead of $\mathbf{x_5}$, that the first variable in the loop $\mathbf{x_2}$ is conditioned on. The flow of influence between ($\mathbf{x_3}$—$\mathbf{x_4}$—$\mathbf{x_5}$—$\mathbf{x_6}$—$\mathbf{x_7}$) is, according to the rules of probabilistic graphical models, a simple chain. Recall that feature ƒ4 states that $\mathbf{x_2}$ should not be influenced by the statistics of the inner loop. In our view, $\mathbf{x_2}$ should not interfere with the statistical flow inside the rest of the loop ($\mathbf{x_2}$—$\mathbf{x_3}$—$\mathbf{x_4}$—$\mathbf{x_5}$—$\mathbf{x_6}$—$\mathbf{x_7}$) as well

(but yet it does in figures 2.15 & 2.16, as $\mathbf{x_2}$ **is** considered part of the loop). Expressed differently, we find it plausible, for instance, that:

$$\mathbf{x_7} \not\perp\!\!\!\perp \mathbf{x_3} \mid \mathbf{x_2}, \tag{2.14}$$

and that:

$$\mathbf{x_7} \not\perp\!\!\!\perp \mathbf{x_3} \mid \mathbf{x_5}. \tag{2.15}$$

However, it is difficult to understand why the following should hold:

$$\mathbf{x_7} \perp\!\!\!\perp \mathbf{x_3} \mid \mathbf{x_2}, \mathbf{x_5}. \tag{2.16}$$

Yet, this is what the statistical models of figures 2.15-2.16, derived from the key assumption 2.2, states. Consider the marginal posterior pdf of $p(\mathbf{x_2}|\mathcal{Z})$; whatever its mean and covariance may depict, it should not change the inner statistical relations *within* the 'submap' formed by $(\mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}, \mathbf{x_6}, \mathbf{x_7})$. Intuitively, the mean of $p(\mathbf{x_2}|\mathcal{Z})$ only offsets the sub-map by a similarity transform, without potentially blocking paths between variables of said sub-map, if it is conditioned on, like in (2.16). In other words, $\mathbf{x_2}$ should not influence the sub-map $(\mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}, \mathbf{x_6}, \mathbf{x_7})$ more than, say, $\mathbf{x_1}$ or $\mathbf{x_0}$ does.

u2· **Gauge-freedom dilemma** The gauge-freedom dilemma is a variation of the above issue, which we apply to the starting pose $\mathbf{x}_0$, instead of $\mathbf{x_2}$ the above scenario. Contrary to u1, this issue is somewhat acknowledged in the SLAM literature, especially in V-SLAM/bundle adjustment. However, existing works on this issue do not, in our view, go enough into details as to why this issue shows up, and what it reveals.

Box 6: SLAM Scenario 4 (pose-graph) - gauge freedom variation

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot is fed sequentially 5 control inputs, which drive it sequentially from $\mathbf{x_0}$ to $\mathbf{x_5}$. Consider that this sequence of controls move the robot about a loop. At $\mathbf{x_5}$, the **starting pose** $\mathbf{x_0}$ is recognized, and the relative situation between both poses is measured. Robot motion and sensor measurements are corrupted by noise. We assume no adverse effect due to the robot inertia, false sightings, outliers, etc.

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}\}$.

To introduce this issue we consider a new toy scenario in box (6). In addition, we consider the realizations of the measurement z_f and control inputs u_a, u_b, u_c, u_d, u_e to be *exactly the same as before*, in the scenario described box (5). Therefore, we should expect, after adjustment, that the robot trajectory within the loop be exactly the same as before, up to a similarity transform. However, the graphical model representation shows a dilemma between two competing ways of characterizing the starting pose $\mathbf{x_0}$.

Figure 2.17 shows the two possibilities. Since, according to the scenario in box (6), the starting pose $\mathbf{x_0}$ is fixed (or rather arbitrarily set, in general to the origin) it would seem that the right choice is (a). But the problem with (a) is that, as in previous point u1, it breaks the flow of influence in loop: it is as if the factors φ_a and φ_f were unary.

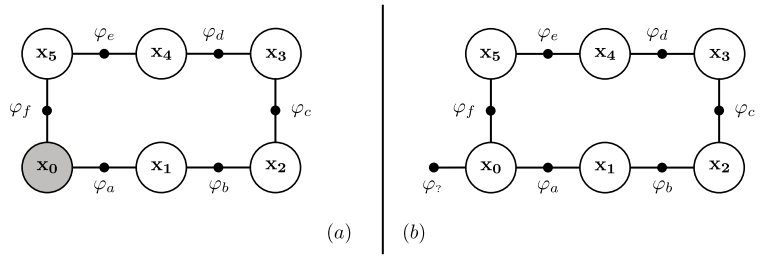


Figure 2.17: Factor representation of the gauge-freedom dilemma. There are two ways to consider the variable $\mathbf{x_0}$, and both have problems.

Contrary to our expectation, it then does not lead to the same loop as in the scenario of box (5) (though stated measurements are assumed the same). Then, consider solution (b), where we breach the letter of the scenario (Box 6) and make $\mathbf{x_0}$ a decision variable. This allows to reproduce the same loop as in figure 2.15. But besides being unfaithful to the scenario, it then raises the question on what to make of the factor $\varphi_?$? We noted this factor with an interrogative symbol because it does not correspond to any physical observation. If $\varphi_?$ is removed, then a singularity appears in the optimization problem, as there would only be relative measurements left in the problem. Conversely, if $\varphi_?$ is chosen with infinite information, then it is easy to see that it amounts to the solution (a). The last solution is probably the worst of both worlds, as it does not only breach the scenario, but also reproduces the problem of solution (a) noted above. In summary, this is a dilemma for which we see no proper way to proceed. Our suspicion is that it is commonly hoped that this choice is numerically inconsequential, especially on large problems. It still does raise the question of faithfulness, as a variant of problem `u1`.

The fact that this is a dilemma is corroborated by its various treatments in the literature. In bundle adjustment, Triggs et al. (2000) introduce the term of "gauge freedom", where freedom denotes the free choice of the designer to fix the datum (reference frame). Strasdat et al. (2010, 2012) confirm that the first frame is typically fixed to eliminate the gauge freedom which is present because measurements are purely relative. In Graph-SLAM, a pose-graph SLAM presented by Thrun et al. (2005), infinity terms are inserted in the diagonal of the first block of the information matrix. This is done in order to fix the first pose. Another approach is to set the information matrix of the prior on $\mathbf{x_0}$ as a matrix identity (Grisetti et al., 2010a). Solution (a) also exists, whereby $\mathbf{x_0}$ is openly considered a fixed variable (Grisetti et al., 2020). We also note that the question is often avoided. To our knowledge, Zhang et al. (2018) are the only ones to have specifically compared various approaches to the gauge freedom problem. It is shown experimentally that the choice slightly influences the estimation result, in particular the posterior covariance (Zhang et al., 2018, fig 5); and the prior weight affects the convergence time. Zhang et al. (2018)'s problem is considered from the optimization level, and it is proposed to mitigate the issue. The paper does not go into the *why*s of the issue, which are of interest to us.

While there is no doubt in anyone that the first pose should be somehow set, the probabilistic formulation of SLAM makes this seemingly simple step quite ambiguous and arbitrary.

Remark 2.6. *Anecdotally, the same problem can be identified in geodetic adjustment*

with the method of variation of geographical coordinates (see 1.4.4). Tardi (1934, chap10) explains that the Panthéon building (in Paris) was chosen as the fixed origin for the adjustment of the NT France network.

As a final point on u2, note that the issue can manifest itself on a small a three-variable problem, see the scenario described in box (7) and figure 2.18.

Box 7: SLAM Scenario 5 (pose-graph) - 3 variables dilemma

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot is fed sequentially 2 control inputs which drive it sequentially from $\mathbf{x_0}$ to $\mathbf{x_1}$, and then from $\mathbf{x_1}$ to $\mathbf{x_2}$. Consider that the controls move the robot about a loop. When at $\mathbf{x_2}$, the **starting pose** $\mathbf{x_0}$ is recognized, and the relative situation between both poses is measured. Robot motions and the sensor measurement are corrupted by noise. We assume no adverse effect due to the robot inertia, false sightings, outliers, etc.

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}\}$.

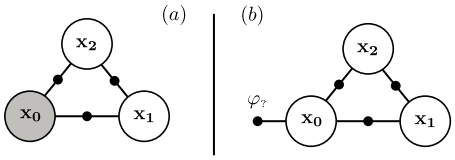


Figure 2.18: A three variable pose-graph problem depicting the gauge freedom dilemma. The scenario is described in box (7). No proper choice between (a) and (b) exists for the characterization of the starting variable. Thus, the probabilistic SLAM formulation cannot solve satisfyingly this seemingly simple three-variable problem (comprised of only 2 decision variables in the scenario).

The problems u1 and u2 are far more concerning than the features f1, f2, f3, f4 explored earlier. For one, if u1 and u2 lead to the production of 'wrong' estimates[7], it is not possible to quantify them at this stage. One would need first to formulate a 'correct' model (devoid of said problems), examine the arguments for and against it, show its value on diverse use cases, etc.

Another concern is the large scope of SLAM scenarios affected by these issues. The gauge freedom dilemma u2 applies to some classical mobile robot scenarios, when the loop is closed on the starting variable, and to most (all?) problems of bundle adjustment. The unwarranted loop-breaking u1 is unfortunately present in all probabilistic SLAM scenarios.

Remark 2.7. *It should be noted, for completeness, that probabilistic SLAM approaches do evade these issues when used in scenarios that contain no loop-closure. But, as noted by Cadena et al. (2016), SLAM without some form of closure reduces to odometry.*

[7] If compelling estimation results are produced, we do not know if it is thanks to the strengths or despite the weaknesses of model.

What are the causes of all these issues? We trace them back to the key assumption 2.2 of probabilistic SLAM, i.e., to the form of the joint pdf. Note that it does not matter which graphical representation is used, nor whether variables are incrementally integrated out like in KF solutions.

The joint pdf defines the frame along which the data are structured. Any inference process on that joint pdf (even if elegant, sophisticated, and/or astutely designed) only transforms, or reduces the data (Fisher, 1922), and only within that probabilistic framework. Consequently, if the pattern of conditional independences defined at the outset is inadequate, it cannot be fixed by better techniques of filtering or optimization.

2.3.3 Complexity/Performance

As SLAM systems may be embedded in devices used for long term operations, the number of factors (ego-motions, observations, etc.) and decision variables (poses, landmark, etc.) inevitably grows. Thus, an issue of scalability arises in the computation of the MAP. This issue can materialize in excessive memory footprint and/or CPU load. This cost essentially derives from the notion of treewidth, which corresponds to the size of the maximum clique[8], both for matrix-based and tree-based graphical solvers (Blair and Peyton, 1993; Paskin and Lawrence, 2003). Note that in the probabilistic graphical model literature, variables are generally assumed to be discrete and the only complexity measure that matters is the treewidth (Chandrasekaran et al., 2008). However, when the product of factors (2.3) is made of Gaussian densities, equivalence can be drawn between the sparsity pattern of a linear system and of the graph representing the joint pdf (factor graphs, BN, etc.). We will give more details in chapter 3 on this equivalence. But even with Gaussian densities, in the simplified SLAM scenarios which are not concerned with outlier rejection or data association, the inference cost is respectively quadratic and cubic to the treewidth, in time and memory footprint (Paskin and Lawrence, 2003). Another measure of complexity is the so-called elimination tree height (Bodlaender et al., 1995) which has to do with the ability to parallelize the elimination process (or the process of decomposition in sparse matrices). Another major source of complexity concerns the recovery of the posterior covariances. In matrix-based solvers, the full posterior covariance is obtained by inversing $A^\top A$, where A is the Jacobian matrix appearing in (2.13) (see also figure 2.7). While $A^\top A$ (a.k.a. the or information matrix) is large and sparse, its inverse is large but dense. Hence, computing the full covariance is often untractable in online robotic applications. In clique tree structures, the situation is better as marginal covariances can be deduced clique by clique once the belief propagation algorithm is done (see the concept of 'ready clique' Koller and Friedman, 2009, def. 10.4). However, the set of queried variables (i.e., the variables for which the posterior covariance is queried) might not in general be a subset of an existing clique, but rather of a span of cliques, leading to additional computations. Unfortunately, knowing the posterior uncertainty of the MAP estimate is important in applications that require data association, active SLAM etc. (Ila et al., 2015). As we can see, both MAP inference and posterior covariance computation cost depends on the sparsity pattern. The sparsity pattern is difficult to anticipate ahead of time, thus, several solutions have been proposed, such as: using approximate algorithms (Ranganathan et al., 2007; Davison and Ortiz, 2022; Ila et al., 2015); 'sparsifying' the graph and/or removing nodes and factors (Vallvé et al., 2018, 2019; Carlevaris-Bianco et al., 2014); eliminating nodes viewed as aux-

[8]The 'maximum clique', or maxclique for short, is a is the largest cluster in a clique tree structure.

iliary (Carlone et al., 2014); proposing pre-integration schemes for IMU data (Forster et al., 2017; Brossard et al., 2022). All in all, fighting against the consequences of adverse sparsity pattern is difficult since many theoretical results have shown that inference is costly in general (Cooper, 1990). Shimony (1994) showed that finding the MAP is NP-hard in a (discrete) belief network. The most important complexity measure, the treewidth, is highly sensitive to the ordering of variable elimination: unfortunately, finding the optimal variable ordering is itself NP-hard (Arnborg et al., 1987). Consequently, for Cooper (1990), "any attempt at a general, exact, efficient solution is unlikely to be successful". Nonetheless, alternative strategies can be used, such as approximate algorithms or specialized algorithms. We put forward, as a hypothesis, that better complexity measures can be achieved by specialization; and that specialization is achievable by incorporating in a better way domain knowledge we have about the SLAM problem.

2.4 Discussion

In this chapter, the existing solutions to represent the adjustment of SLAM were presented. The principles of the main approach, denoted probabilistic SLAM, were presented in §2.1. Probabilistic SLAM are a family of approaches which include the dominant approach (MLE SLAM), but also Kalman filter and particle filters. The main commonality behind this family is to pose the SLAM estimation problem as a joint posterior pdf (key assumption 2.1), and the factorization of the joint into a product of local probabilistic models (key assumption 2.2). A non-exhaustive list of achievements reached through the use of this framework was given (§2.2). In relation to our historical work in chapter 1, we noted that this approach is closer to the one named 'method of variation of coordinates' (see 1.4.4 and table 1.1). In the ensuing section §2.3, we explored the limitations of that framework along 3 axes: in terms of its ability to feature spatial reasoning considerations, in terms of faithfulness representation, and in terms of inherent algorithmic complexity. Deep underlying issues are pointed out which are traceable to the fundamental assumptions mentioned above. These issues have, to our knowledge, not yet been properly identified in whole. There is no deny that current formulations of SLAM have shown (and will continue to show) compelling results in challenging environments (e.g., large scale maps, badly initialized priors, with outliers, etc.). This is substantiated by a large body of literature.

But these problems deserve more attention. Yet, SLAM is sometimes considered solved, in particular the adjustment part. This impression can be found early in highly cited reviews. For instance, Durrant-Whyte and Bailey (2006) writes that "at a theoretical and conceptual level, SLAM can now be considered a solved problem". A decade later, Cadena et al. (2016) are more cautious, but the probabilistic formulation is not directly questioned; it is instead suggested that some additional modules are needed to address a wider array of problems in SLAM. In the book *Probabilistic Graphical Models*, Koller and Friedman (2009, p679-683) address tracking, localization and mapping problems via graphical models (using Bayesian networks) and consider that most instances of the SLAM problem can be tackled, "so that this problem is now essentially considered solved". At minimum, the MLE SLAM method with factor graphs is considered well established (Rosen et al., 2021; Dellaert, 2021).

There are therefore, to our best knowledge, no solutions in sight for the issues depicted in this chapter. But how important is it to solve these issues? From an historical perspective, SLAM inherits many traits from geodetic adjustment. This problem embodied several funda-

mental challenges of the period: the combinations of different observational equations and the specifications of error curves (Stigler, 1986). The processing of geodesic triangulations started from a binary question, related to the shape of earth, in the mid 18th century. Ad-hoc methods were the norm, e.g., not correcting closures or calculating the arithmetic mean. Then, in the late 18th century, new instruments and more demanding requirements appeared for the determination of a unit length for the early metric system. Solving rigorously the geodetic adjustment problem became a scientific basis for a metrological endeavor. In doing so, and in some instances in parallel with the discipline of astronomy, major advances in theory of probabilities and inference techniques were made by a relatively small number of people (see chapter 1). The SLAM problem, or more generally the problem of emulating spatial understanding, should be recognized in continuation as a more complex and rewarding problem, i.e. one which should propel the development of new concepts that go beyond existing estimation techniques (of which some were invented for geodetic adjustment).

So what can be done in our view? We want to pose the SLAM problem in a way that depicts all the *a priori* qualitative judgements, or knowledge, that are available to us. This motivates the search of a more insightful representation. There are no doubts on the ability of graphical models to seamlessly carry and illustrate these assumptions. There are also no doubts that the assumption of Gaussian distributed noises is very practical. The Gaussian noise assumption has tended to lead, since 1809 (Gauss, 1809), to a (nonlinear) least squares optimization problem. In geodetic adjustment, we have seen that there were several ways to state the least squares problem (see 1.4), each one based on different prior qualitative considerations. Our goal is therefore to explore other representations to state the optimization problem in SLAM differently.

The outline of the remainder of the manuscript goes as follows. As the achievements of current SLAM solutions cannot be denied, we propose in chapter 3 to improve on the probabilistic SLAM via a Causal Bayesian Network, which exploits causal relationships in the robot trajectory to gain insights that the solver level can exploit.

Another more ambitious endeavor, but that we believe more promising, is to reformulate SLAM by taking inspiration from the Laplacian conception (§1.5.3), along with relatively recent tools developed by Pearl and colleagues: the framework of structural causal model (SCM) and the causal graph representation. This is done in chapter 4.

Chapter 3

Causal Bayesian Network for SLAM

In this chapter, we propose to re-investigate SLAM on the basis of Causal Bayesian Networks (CBNs), a directed acyclical graph (DAG) structure which has similarities to Bayesian Networks. The key assumptions 2.1 and 2.2 of probabilistic SLAM are mostly kept, but we use additional considerations to insightfully insert causal notions in the ego-motion of the robot. Our motivation is to leverage the fact that new odometry measurements (or control inputs) do not induce propagation of belief to older part the trajectory/map. Consequently, this framework exhibits the fact that sometimes it is unnecessary to compute an adjustment, at least not until a loop-closure event takes place.

In this context, the CBNs (Pearl, 2009, §1.3) are judged as the most suitable models. On the one hand, they feature d-separation properties, and on the other hand, they can capture the fact that each trajectory element is jointly caused by its predecessor and by a relative motion. The directed edges of the CBNs explicitly account for causal relations[1].

Our approach can be summarized as follows. Once the CBN of the SLAM problem is drawn up, the inference proceeds by building an undirected chordal graph, which shows the fill-in (additional connections between nodes) induced by the successive elimination of decision variables. This chordal graph is then turned into a clique tree, whose cliques are fully connected nodes of the chordal graph. Belief propagation is conducted using a message passing algorithm between cliques. It leads both to the Maximum a Posteriori (MAP) and marginal posterior distributions within each clique.

The benefits of our approach over other tree-based approaches materializes during the message passing step. Depending on the location of the SLAM loop closures, constant messages can be *a priori* exhibited in the anti-causal direction, so that efficient incremental inference can be obtained by only updating the subtree subject to changes. Potentialities of the approach include: computation of arbitrary marginal posteriors; application to non-Gaussian (dynamics or measurement) noises; extension to cooperative SLAM. The method is presented on a pose-graph problem and its strengths and limitations are compared with other graphical approaches and matrix-based approaches.

The chapter is organized as follows. Section 3.1 states the problem. Section 3.2 gives a brief reminder on what the involved structures (chordal graph, clique tree) represent, how to convert CBNs into clique trees and message passing for exact inference. Subsequent sections describe a numerical experiment. Our last section 3.7 discusses the merits and limits of the approach. We explain why, ultimately, we foresee a more promising outcome in pursuing a full-fledged causal structure for the SLAM problem, particularly in view of our analysis in the previous chapters 1 & 3. This causal structure will be proposed in chapter 4.

3.1 Problem Statement

3.1.1 Basics

Notations

In the sequel, the set $\mathcal{X}$ gathers the target unknown variables, i.e., trajectory poses and landmarks. The sets $\mathcal{U}$ and $\mathcal{Z}$ are respectively built up with known control inputs and odometric measurements: IMUs, wheel encoders, scan-matchers etc. The set $\mathcal{Y}$ gathers other kind of measurements: absolute localization such as GNSS measurements, relative perception of landmarks, loop-closures on past positions, etc. From now on, any uppercase calligraphic

[1] Recall that this is not the case for BNs, see remark 2.3.

letter refers to a set of variables, each one being labeled in lowercase. When these variables live in a matrix smooth manifold such as $\mathrm{SE}(n)$ or $\mathrm{SO}(n)$, they are typed in small bold font, and so are their corresponding sets, e.g.,

$$\mathcal{X} \triangleq \{\mathbf{x_0}, \mathbf{x_1}, \dots, \mathbf{x_n}\}\,, \qquad \mathcal{Z} \triangleq \{\mathbf{z_0}, \mathbf{z_1}, \dots, \mathbf{z_m}\}\,. \tag{3.1}$$

If variables live in a Euclidean space (e.g., known control inputs), then they and their sets are typed in normal fonts.

The goal is to answer inference questions about the joint posterior pdf $p(\mathcal{X}, \mathcal{L}|\mathcal{Z}, \mathcal{U}, \mathcal{Y})$. The Bayes rule is generally invoked, yet another path is taken. To simplify the notations, but without loss of generality in the reasoning, all bold variables will be assumed to belong to the same manifold (though $\mathcal{Y}$ may be built with heterogeneous elements, but this does not call into question the approach). The forthcoming notations closely follow the tutorial Solà et al. (2021). The operator $\circ$ stands for the matrix Lie group composition. Log denotes the mapping from an element of the manifold to the Euclidean space isomorphic to its Lie algebra, and Exp stands for its reciprocal mapping. Though with a slight notation misuse, elements of the Lie algebra and of its isomorphic Euclidean space will no longer be distinguished. For instance, if $\mathbf{x} \in \mathrm{SE}(2)$, then its image in the Lie algebra (generally written as $\log(\mathbf{x}) \in \mathfrak{se}(2)$) is isomorphic to the 3-dimensional real vector $\mathrm{Log}(\mathbf{x}) \in \mathbb{R}^3$ (also written $(\log(\mathbf{x}))^\vee$), so that $\mathfrak{se}(2)$ and $\mathbb{R}^3$ are not distinguished. Conversely, a vector $\nu \in \mathbb{R}^3$ maps to an element of $\mathfrak{se}(2)$ (generally noted $\nu^\wedge \in \mathfrak{se}(2)$), which itself maps to $\mathrm{Exp}(\nu) = \exp(\nu^\wedge) \in \mathrm{SE}(2)$, see Figure 3.1.

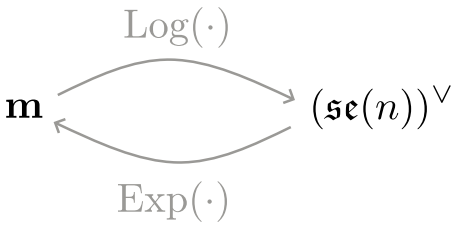


Figure 3.1: Conversion of an element $\mathbf{m}$ in a Lie group to the Euclidean space isomorphic to its Lie algebra, and back, through the Log and Exp mapping functions.

Motivation to work on a pose-graph SLAM

Pose-graph is often considered the backbone of modern SLAM Tian et al. (2021). For instance, many higher-level spatial perception constructs like Kimera Rosinol et al. (2021) are built on top it. It also constitutes a simple enough structure to introduce the approach, and compare it with state-of-the-art back-ends cited above. A toy pose-graph case study is displayed on Figures 3.2 and 3.3 as a factor graph and a CBN, respectively. All its factors are odometric, except a loop-closure connecting $\mathbf{x_2}$ and $\mathbf{x_6}$. A common practice is to set the robot's initial at the origin ϵ of the manifold, i.e., $\mathbf{x}_0 = \epsilon$. Regarding CBN terminology, the initial pose is viewed as an intervention Pearl (2009) rather than a prior (which would open the question of how to fill the covariance) or an observation (for it is not a measurement, though it is the common choice of SLAM practitioners).

The most noteworthy differences between the two models occur on the trajectory part between any two consecutive spatial entities (poses), where additional unknown variables depicting the Log of relative motions are inserted. On the CBN Figure 3.3, their names

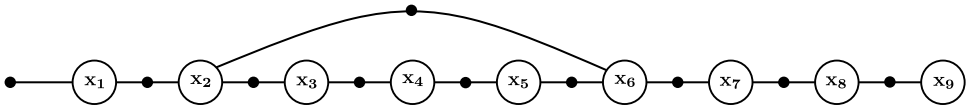


Figure 3.2: Factor graph of a toy pose-graph problem.

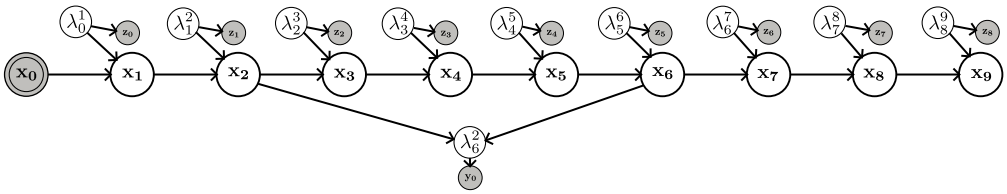


Figure 3.3: CBN of a toy pose-graph problem.

are $\lambda^{\text{to\_variable}}_{\text{from\_variable}}$. On the trajectory part sketched on Figure 3.4, these are named λ. We will denote Λ the set of relative motions λ. Elements of Λ can be caused by a control input and can give rise to one or multiple observations (resp. u and $\mathbf{z}$'s on Figure 3.4).

Remark 3.1. *On the representation of landmarks.*
As per key assumption 2.2, the observation of a landmark is still considered to be of the form $p(\mathbf{z}|\mathbf{x},\mathbf{l})$*, where* $\mathbf{l}$ *is a landmark. Its graphical representation does not change in this chapter: it assumed that the observation* $\mathbf{z}$ *is caused by the robot pose* $\mathbf{x}$ *and the landmark position* $\mathbf{l}$*:* $\mathbf{x} \rightarrow \mathbf{z} \leftarrow \mathbf{l}$*. However, though the approach in this chapter does not exclude scenarios with landmarks, a complete reconceptualization of the processing of landmarks will be given chapter 4.*

3.1.2 Trajectory representation

Consider figure 3.4-(left). The unknown consecutive poses $\mathbf{x}_{(-)}$ and $\mathbf{x}_{(+)}$ belong to the matrix Lie group. The current pose $\mathbf{x}_{(+)}$ is caused by its past $\mathbf{x}_{(-)}$ and the relative motion λ, independently of $\mathbf{x}_{(-)}$'s ancestors, along the definition

$$\mathbf{x}_{(+)} = \mathbf{x}_{(-)} \circ \text{Exp}(\lambda), \tag{3.2}$$

where λ expresses the relative motion in the Lie algebra of the matrix Lie group. λ is observed through the random measurement $\mathbf{z}$.

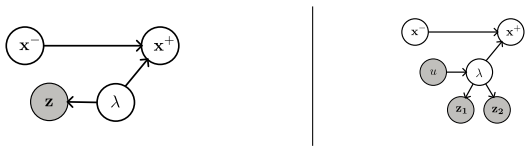


Figure 3.4: Trajectory part between two poses of a pose-graph CBN. $\mathbf{x}_{(-)}$ and $\mathbf{x}_{(+)}$ respectively stand for the past (parent) and the current (child) pose. $\mathbf{x}_{(+)}$ is caused by $\mathbf{x}_{(-)}$ and the relative motion λ. (Left) λ depends on no external control, and is observed through the random measurement $\mathbf{z}$. Note that a prior on λ may exist. (Right) The external control input u causes λ through a conditional probability distribution (CPD), and λ gives rise to multiple measurements (limited to 2, $\mathbf{z_1}, \mathbf{z_2}$, but the approach can be generalized).

Figure 3.4(-right) shows a more general case, where it is assumed that the relative motion λ listens to a stochastic control input u and is then observed by measurements $\mathbf{z_1}$ and $\mathbf{z_2}$. Figure 3.4(-right) is here to show that the model can be adapted to cover the presence of prior controls and/or an arbitrary number of sensors that observe the relative motion λ. However, note that in the case of infrequent and/or unsynchronized sensors, it is assumed that measurements have been preprocessed to be compatible with the associated CBN representation. The joint pdf $p(\mathbf{x}_{(+)}, \lambda, u, \mathbf{z_1}, \mathbf{z_2}|\mathbf{x}_{(-)})$ associated to the trajectory part writes as

$$p(\mathbf{x}_{(+)}, \mathbf{x}_{(-)}, \lambda, u, \mathbf{z_1}, \mathbf{z_2}) = p(\lambda|u)p(\mathbf{z_1}|\lambda)p(\mathbf{z_2}|\lambda)p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)}, \lambda)p(u). \tag{3.3}$$

The mathematical expressions of $p(\lambda|u)$, $p(\mathbf{z_1}|\lambda)$, $p(\mathbf{z_2}|\lambda)$ are known and are parameterized by the input data $u, \mathbf{z_1}, \mathbf{z_2}$. As for $p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)}, \lambda)$, it is derived from domain knowledge (kinematics in view of (3.3)) and boils down to

$$p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)}, \lambda) = \delta\left(\mathrm{Log}\left((\mathbf{x}_{(-)} \circ \mathrm{Exp}(\lambda))^{-1} \circ \mathbf{x}_{(+)}\right)\right) \tag{3.4}$$

$$= \delta\left(\mathbf{x}_{(+)} \ominus (\mathbf{x}_{(-)} \oplus \lambda)\right), \tag{3.5}$$

with $\delta(.)$ the Dirac delta function and $\mathbf{x} \oplus \lambda \triangleq \mathbf{x} \circ \mathrm{Exp}(\lambda)$ and $\mathbf{y} \ominus \mathbf{x} \triangleq \mathrm{Log}(\mathbf{x}^{-1} \circ \mathbf{y})$.

To the best of our knowledge, this model, though key to the approach, has not been used recently. We have found that Paskin (2002, Fig. 6) used a comparable model for the filtering view of probabilistic SLAM and astutely remarks that belief does not propagate backwards in those situations, but it is not exploited further.

3.1.3 Elimination of the relative motion λ

The marginalization of the pdfs entailed in the full problem with respect to the relative motions eases inference by message passing. The procedure is hereafter run through for one trajectory part[2] (Figure 3.4-(right)), by eliminating λ though marginalization of (3.3). Its generalization will be further discussed.

Define

$$\tau(\mathbf{x}_{(+)}, \mathbf{x}_{(-)}) \triangleq \int_\lambda p(\mathbf{x}_{(+)}, \lambda, u, \mathbf{z_1}, \mathbf{z_2}|\mathbf{x}_{(-)})\mathrm{d}\lambda$$

$$= \int_\lambda p(\lambda|u)p(\mathbf{z_1}|\lambda)p(\mathbf{z_2}|\lambda)p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)}, \lambda)p(u)\mathrm{d}\lambda. \tag{3.6}$$

The graph structure of Figure 3.4-(right) implies that, $\mathbf{z_1}$, $\mathbf{z_2}$, u are d-separated from each others given λ. This, together with (3.5), leads to:

$$\tau(\mathbf{x}_{(+)}, \mathbf{x}_{(-)}) = \int_\lambda p(\mathbf{z_1}, \mathbf{z_2}, \lambda|u)p(u)\,\delta\left(\mathbf{x}_{(+)} \ominus (\mathbf{x}_{(-)} \oplus \lambda)\right)\mathrm{d}\lambda,$$

$$= p(\mathbf{z_1}, \mathbf{z_2}, u)\int_\lambda p(\lambda|u, \mathbf{z_1}, \mathbf{z_2})\,\delta\left(\mathbf{x}_{(+)} \ominus (\mathbf{x}_{(-)} \oplus \lambda)\right)\mathrm{d}\lambda. \tag{3.7}$$

In view of the inverse and associativity properties of the composition operator $\circ$, the argument of the delta function comes as

$$\mathrm{Log}\left((\mathbf{x}_{(-)} \circ \mathrm{Exp}(\lambda))^{-1} \circ \mathbf{x}_{(+)}\right) = \mathrm{Log}\left((\mathrm{Exp}(\lambda))^{-1} \circ ({\mathbf{x}_{(-)}}^{-1} \circ \mathbf{x}_{(+)})\right), \tag{3.8}$$

[2]Similar method is used for elimination loop closure relative motion, though structure differs slightly.

and admits the following Baker-Campbell-Hausdorff (BCH) first order expansion (Chirikjian, 2012, §10.2.7):

$$\begin{aligned}\operatorname{Log}\left(\left(\operatorname{Exp}(\lambda)\right)^{-1}\circ\left({\mathbf{x}_{(-)}}^{-1}\circ\mathbf{x}_{(+)}\right)\right) &= -\operatorname{Log}(\operatorname{Exp}\lambda)+\operatorname{Log}({\mathbf{x}_{(-)}}^{-1}\circ\mathbf{x}_{(+)})+\dots\\ &= -\lambda+\operatorname{Log}({\mathbf{x}_{(-)}}^{-1}\circ\mathbf{x}_{(+)})+\dots\end{aligned} \tag{3.9}$$

where $\operatorname{Log}(\operatorname{Exp}\lambda)=\lambda$ is true for small motions, assuming that the frequency of data is sufficient relative to the velocities and dynamics of the robot.

The Dirac mass is concentrated at:

$$\operatorname{Log}\left(\operatorname{Exp}(\lambda)^{-1}\circ\left({\mathbf{x}_{(-)}}^{-1}\circ\mathbf{x}_{(+)}\right)\right)=0, \tag{3.10}$$

i.e., when (3.9) is zero. The Log(.) of (3.10) can thus be unwrapped, and using the inverse property of Lie groups leads to

$$\lambda=\operatorname{Log}({\mathbf{x}_{(-)}}^{-1}\circ\mathbf{x}_{(+)}). \tag{3.11}$$

Incidentally, (3.11) says that the sum of the first two terms of the BCH expansion (3.9) is zero, so that the remainder "$+\dots$" is necessarily zero in view of (3.10). The following equality of Dirac distributions

$$\delta\left(\mathbf{x}_{(+)}\ominus(\mathbf{x}_{(-)}\oplus\lambda)\right)=\delta\left(\lambda-\operatorname{Log}({\mathbf{x}_{(-)}}^{-1}\circ\mathbf{x}_{(+)})\right) \tag{3.12}$$

is henceforth assumed, for their masses are concentrated when (3.10) and (3.11) hold. Injecting (3.12) into (3.7) finally yields

$$\tau(\mathbf{x}_{(+)},\mathbf{x}_{(-)})=p(\mathbf{z_1},\mathbf{z_2},u)p(\operatorname{Log}(\mathbf{x}_{(-)}^{-1}\circ\mathbf{x}_{(+)})|u,\mathbf{z_1},\mathbf{z_2}). \tag{3.13}$$

To understand what is represented by $p(\operatorname{Log}(\mathbf{x}_{(-)}^{-1}\circ\mathbf{x}_{(+)})|u,\mathbf{z_1},\mathbf{z_2})$ in (3.13), consider the analytical elimination of λ from the joint pdf associated with the trajectory part in (3.3):

$$\begin{aligned}\int_{\lambda}p(\mathbf{x}_{(+)},\lambda,u,\mathbf{z_1},\mathbf{z_2}|\mathbf{x}_{(-)})\mathrm{d}\lambda &= p(\mathbf{x}_{(+)},u,\mathbf{z_1},\mathbf{z_2}|\mathbf{x}_{(-)})\\ &= p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)},u,\mathbf{z_1},\mathbf{z_2})p(u,\mathbf{z_1},\mathbf{z_2}|\mathbf{x}_{(-)})\end{aligned} \tag{3.14}$$

And since $(\mathbf{x}_{(-)}\perp\!\!\!\perp u,\mathbf{z_1},\mathbf{z_2})$ can be read from figure 3.4-(right), then $p(u,\mathbf{z_1},\mathbf{z_2}|\mathbf{x}_{(-)})=p(u,\mathbf{z_1},\mathbf{z_2})$.

Thus, $p(\operatorname{Log}(\mathbf{x}_{(-)}^{-1}\circ\mathbf{x}_{(+)})|u,\mathbf{z_1},\mathbf{z_2})$ in (3.13) stands for $p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)},u,\mathbf{z_1},\mathbf{z_2})$ by identification with (3.14).

Remark 3.2. *The independence assumption $(\mathbf{x}_{(-)}\perp\!\!\!\perp u,\mathbf{z_1},\mathbf{z_2})$ holds if no descendant of $\mathbf{x}_{(+)}$ is conditioned on. However, we can see that this is not the case for $\mathbf{x}_{(+)}\equiv\mathbf{x_2}$ in the case study CBN, figure 3.3; as $\mathbf{y_0}$ is an observed descendant of $\mathbf{x_2}$. We give the justification for this case, and all other similar such cases, in §3.2.3.*

Remark 3.3. *In the rest of the chapter, the model for the robot ego-motion will correspond to the simple case, i.e., figure 3.4(-left). This does not mean that the control or other measurements are dropped nor approximated away. Instead, this is due to the limitation of the scenario of the subsequent applications of this chapter (in particular, datasets). Adaptation to more complex scenarios (e.g., figure 3.4(-right)) can still be done thanks to the elimination process described above.*

3.2 Undirected Tree for Exact Inference

The above section was concerned about the CBN representation of SLAM and the analytical elimination of the newly introduced relative motions, denoted by the set Λ.

We now detail how to conduct inference in the CBN by means of belief propagation, i.e., by message passing techniques into a graph which clusters decision variables. This section describes the process of building a so-called clique tree, a special clustered structure that guarantees exact inference after belief propagation. First, a reminder is given on the graphical representation of elimination (a.k.a. marginalization) together with the concepts of chordal graph and clique tree. A choice was made set forth these concepts on a simple example which structure is close to the other CBNs in the present chapter. Then, in the subsequent subsections, the process is explained for the toy pose-graph problem shown on figure 3.3, whose joint pdf writes as

$$p(\mathcal{X},\Lambda,\mathcal{Z},\mathcal{Y}|\mathbf{x_0}) = \prod_{i=0}^{n} p(\mathbf{x_{i+1}}|\mathbf{x_i},\lambda_i^{i+1})p(\lambda_i^{i+1})p(\mathbf{z_i}|\lambda_i^{i+1}) \times \prod_{\text{relevant } k,k'} p(\lambda_k^{k'}|\mathbf{x_k},\mathbf{x_{k'}})p(\mathbf{y_k}|\lambda_k^{k'})p(\lambda_k^{k'}) \quad (3.15)$$

where the set $\{p(\mathbf{y}_k,\lambda_k^{k'}|\mathbf{x_k},\mathbf{x_{k'}})\}_{k,k'}$ includes, among others, the factors related to loop closures; and Λ is the set of all relative motions $\{\lambda_j\}_j$.

3.2.1 Reminder on graphical elimination

This subsection introduces the symbolic manipulations yielding the clique tree, as special data structure which supports exact belief propagation algorithms. We consider, as a motivating example, the set $\Theta \triangleq \{A,B,C,D,E,F\}$, and its joint (continuous, without loss of generality) distribution

$$p(\Theta) = p(A)p(B)p(D)p(C|A,B)p(E|C,D)p(F|A,E), \quad (3.16)$$

which looks like (3.15). Its CBN is portrayed on figure 3.5.

Given $E \subset \Theta$, consider the marginal query:

$$p(E) = \int_{\Theta\setminus E} p(\Theta)d(\Theta \setminus E). \quad (3.17)$$

A practical question that emerges in view of (3.17) is the selection of the elimination ordering to be applied to $p(\Theta)$, and its consequences. Consider two elimination orderings: $\mathbf{O_1} \triangleq [D,B,F,A,C]$ and $\mathbf{O_2} \triangleq [C,A,B,F,D]$. Applying order $\mathbf{O_1}$ develops as:

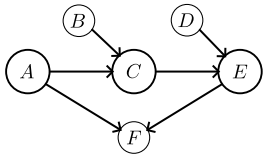


Figure 3.5: (Causal) Bayesian Network G of the distribution $p(\Theta)$.

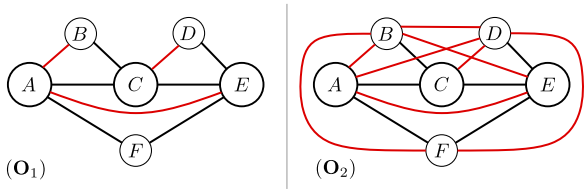


Figure 3.6: Induced (or chordal) graphs reflecting the elimination orderings $\mathbf{O_1}$ (left) and $\mathbf{O_2}$ (right). Fill-in edges, i.e., edges not appearing in the original graph, are in red. Different orderings produced different fill-in schemes.

$$
\begin{aligned}
p(E) &= \int_C \int_A \int_F \int_B \int_D p(\Theta) dD dB dF dA dC \\
&= \int_C \int_A \int_F \int_B p(A)p(B)p(C|A,B)p(F|A,E) \left(\int_D p(D)p(E|C,D)dD \right) dB dF dA dC \\
&\qquad \text{and set} \quad \tau_1(C,E) \triangleq \int_D p(D)p(E|C,D)dD, \\
&= \int_C \int_A \int_F p(A)p(F|A,E)\tau_1(C,E) \left(\int_B p(B)p(C|A,B)dB \right) dF dA dC \\
&\qquad \text{and set} \quad \tau_2(A,C) \triangleq \int_B p(B)p(C|A,B)dB, \\
&= \int_C \int_A p(A)\tau_1(C,E)\tau_2(A,C) \left(\int_F p(F|A,E)dF \right) dA dC \\
&\qquad \text{and set} \quad \tau_3(A,E) \triangleq \int_F p(F|A,E)dF, \\
&= \int_C \tau_1(C,E) \left(\int_A p(A)\tau_2(A,C)\tau_3(A,E)dA \right) dC \\
&\qquad \text{and set} \quad \tau_4(C,E) \triangleq \int_A p(A)\tau_2(A,C)\tau_3(A,E)dA, \\
&= \int_C \tau_1(C,E)\tau_4(C,E)dC. \qquad (3.18)
\end{aligned}
$$

Similarly, ordering along $\mathbf{O_2}$ leads to:

$$
\begin{aligned}
p(E) &= \int_D \int_F \int_B \int_A \int_C p(\Theta) dC dA dB dF dD \\
&= \int_D \int_F \int_B \int_A p(A)p(B)p(F|A,E)p(D) \left(\int_C p(C|A,B)p(E|C,D)dC \right) dA dB dF dD \\
&\qquad \text{and set} \quad \tau_1'(A,B,E,D) \triangleq \int_C p(C|A,B)p(E|C,D)dC, \\
&= \int_D \int_F \int_B p(B)p(D) \left(\int_A p(A)p(F|A,E)\tau_1'(A,B,E,D)dA \right) dB dF dD \\
&\qquad \text{and set} \quad \tau_2'(F,B,E,D) \triangleq \int_A p(A)p(F|A,E)\tau_1'(A,B,E,D)dA, \\
&= \int_D \int_F p(D) \left(\int_B p(B)\tau_2'(F,B,E,D)dB \right) dF dD \\
&\qquad \text{and set} \quad \tau_3'(F,E,D) \triangleq \int_B p(B)\tau_2'(F,B,E,D)dB, \\
&= \int_D p(D) \left(\int_F \tau_3'(F,E,D)dF \right) dD \\
&\qquad \text{and set} \quad \tau_4'(E,D) \triangleq \int_F \tau_3'(F,E,D)dF, \\
&= \int_D p(D)\tau_4'(E,D)dD. \qquad (3.19)
\end{aligned}
$$

The scopes of marginalization, i.e., the variables involved in the blue potentials in (3.18) and (3.19), are a metric for the complexity of the marginalization operation (Koller and Friedman, 2009, chap.9). The larger the size of the scope, the higher the cost, both in space and in time. Figure 3.6 shows undirected graphs associated with the orderings $\mathbf{O_1}$ and $\mathbf{O_2}$. If two variables are both involved in an integration step (i.e., they appear in blue in the same expression in resp. (3.18) and (3.19)), then an edge is drawn between them. Figure 3.6 features the edges non-existent in the original BN (figure 3.5) in red. The red edges are known as "fill-in" edges in the literature[3]. This representation (figure 3.6) is called the "induced graph" (i.e., the graph induced by the original graph from elimination) or the "cover graph" (i.e., the graph that covers the original graph). It is said that the induced graph is "triangulated", or "chordal".

Property 1. *For every cycle of length greater than 4 in a graph $\mathcal{G}$, if $\mathcal{G}$ has an edge connecting two non-consecutive nodes in the cycle, then $\mathcal{G}$ is triangulated, or chordal.*

The process of producing the induced graph from $\mathcal{G}$ can be referred as the graph triangulation. It is said that the induced graph *covers* the original graph[4].

Remark 3.4. *As developing the analytical equations (such as eq 3.18 and 3.19) can be tedious, the induced graph can instead be produced from a more astute graphical process, as follows: at each ordered step of the elimination process, connect with each other the non-eliminated neighbors of the node to be eliminated.*

[3]Depending on the structure of the original graph and the ordering chosen for elimination, no fill-in edge might be produced during elimination. This is typically the case if the original graph is already chordal.

[4]More precisely, the induced graph covers the moralized graph of the original directed graph. The moralized graph replaces the directed edges of the DAG by simple edges and connects parents of a given node between each other.

It can be noticed that ordering $\mathbf{O_2}$ leads to more fill-in edges and larger size of variable scopes than ordering $\mathbf{O_1}$ does. As mentioned in §2.3.3, finding the optimal ordering becomes exponentially difficult as the size of the belief network grows, and heuristics are resorted to. Popular heuristics include approximate minimum ordering (AMD) (Amestoy et al., 1996), nested dissection (Kai Ni and Dellaert, 2010), see also Agarwal and Olson (2012) for ordering strategies for SLAM. In this small example, $\mathbf{O_1}$ is the optimal elimination ordering for the query $p(E)$.

Suppose that the query of $p(C)$ is contemplated, instead of $p(E)$. Consider elimination ordering to be $[D, B, F, A, E]$. It can be quickly noticed that similar operations of integration are carried out as in (3.18), except for the last one. The same remark stands for the query $p(C, E)$. Based on the insight that some of these operations can be cached, the clique tree data structure has been proposed in the literature (Lauritzen and Spiegelhalter, 1988; Shenoy and Shafer, 1990).

Definition 3.1. *Adapted from Darwiche (2009, def 9.13).*
*A **clique tree** $\mathcal{T} = (\mathcal{C}, S)$ of a DAG $\mathcal{G}$ is a tree which connects cliques $\mathcal{C}$ (or clusters) via separator edges S such that:*

- *A clique $\mathcal{C}_i$ is a set of nodes from $\mathcal{G}$.*
- *Each conditional probability distribution (CPD) in the underlying distribution of $\mathcal{G}$ must be embedded in one clique whose scope is compatible (a.k.a. family preservation property).*
- *If a node appears in two cliques $\mathcal{C}_i$ and $\mathcal{C}_j$, then it must also appear in every clique $\mathcal{C}_k$ on the path connecting $\mathcal{C}_i$ and $\mathcal{C}_k$. This is known as the "running intersection property" (RIP).*

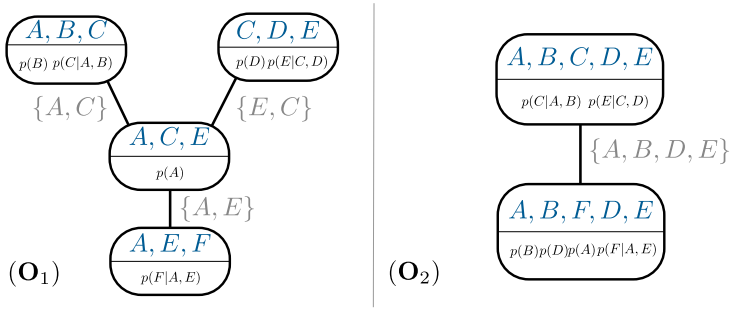


Figure 3.7: Clique trees of the distribution $p(\Theta)$ induced by orderings $\mathbf{O_1}$ and $\mathbf{O_2}$. The treewidth of the clique tree on the left is lowered, which entails better performance.

Every graph is known to have (at least) one associated clique tree (Koller and Friedman, 2009, Theorem 4.12). Clique trees can be built from the induced graphs: the cliques are formed out of the fully connected nodes, and the separators are deduced from the RIP[5]. Then the potentials (make up the joint distribution 3.3) are allocated, each in only one clique to avoid double counting ("family preservation property"). Figure 3.7 shows the clique trees

[5] Alternatively, the edges of the clique tree can be deduced from the maximum spanning tree of the cliques pairwise intersection set (Darwiche, 2009, theorem 9.9).

corresponding to the induced graphs of figure 3.6. The treewidth comes as the size of the scope of the maximum clique. For ordering $\mathbf{O_1}$, the treewidth is 3, while it is 5 for ordering $\mathbf{O_2}$.

The other interest of clique trees, besides to avoid the repeatition of some elimination steps, comes from the fact that new mechanics emerge for inference. These mechanics are difficult to foresee from the analytical perspective. Procedural so-called "message passing" algorithms on such structures have been designed to find the MAP or to compute marginals of arbitrary scope. Conceptually, the structure can even be generalized to other semi-ring operations (Aji and McEliece, 2000).

A clique belief $\Psi_{\mathcal{C}_i}$ (product of potentials affected to a clique $\mathcal{C}_i$) is the posterior marginal over C_i's scope once it has received all messages from neighboring cliques (Koller and Friedman, 2009, §10.2.2 & §10.3.3). Several programming architectures exist to arrange this message passing scheme (Lepar and Shenoy, 2003).

The cost of inference materializes when computing messages from one clique to a neighbor. In the case when the potentials are Gaussian densities, the marginalization operations are matrix decompositions (see, e.g., Dellaert and Kaess (2017, chap.3)). Closely related structures have already been exploited in SLAM (Pinies et al., 2012; Kaess et al., 2010).

3.2.2 Graph triangulation for the CBN and Clique Tree building

Equipped with the necessary tools to conduct inference on a graphical model, we return to the processing of our toy case study presented figure 3.3 whose posterior joint pdf writes as (3.15).

The used triangulation method proceeds as follows: first, an edge is added connecting the parents of each node in the trajectory $\mathbf{x_i} \rightarrow \mathbf{x_{i+1}} \leftarrow \lambda_i$; second, the remaining variables are virtually eliminated following an ordering given by the Approximate Minimum Degree algorithm Amestoy et al. (1996). The elimination is virtual in that no summation is explicitly made: the aim is to just simulate the fact that additional edges appear to link up unconnected and unmarked neighbors of eliminated variables, where a variable is marked once eliminated. The resulting induced graph is depicted on Figure 3.8. Red edges are filled through triangulation.

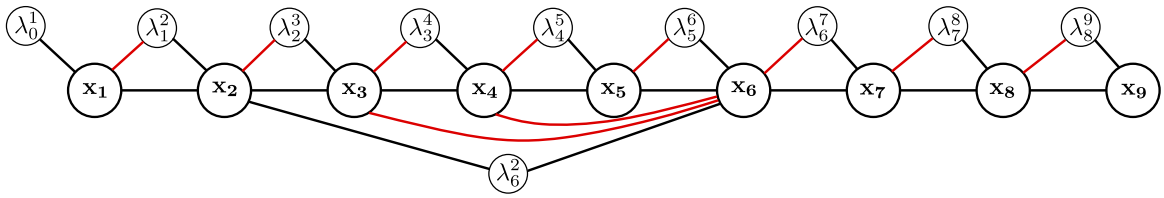


Figure 3.8: Induced graph for the decision variables of the SLAM CBN of Figure 3.3 (fill-in edges in red).

Cliques are retrieved by *maximally* clustering the fully connected nodes of the cover graph. This defines the max-cliques in the clique tree. To produce the clique tree edges (separators) compliant with the RIP, the following interwoven steps are taken: a candidate edge is drawn between any pair of cliques which has a non-empty intersection; the edges are weighted according to the size of the intersection set (or scope); the well-known *maximum spanning tree* algorithm is used to fetch the clique tree (figure 3.9) (Koller and Friedman, 2009, §10.4.2 & Figure 10.10).

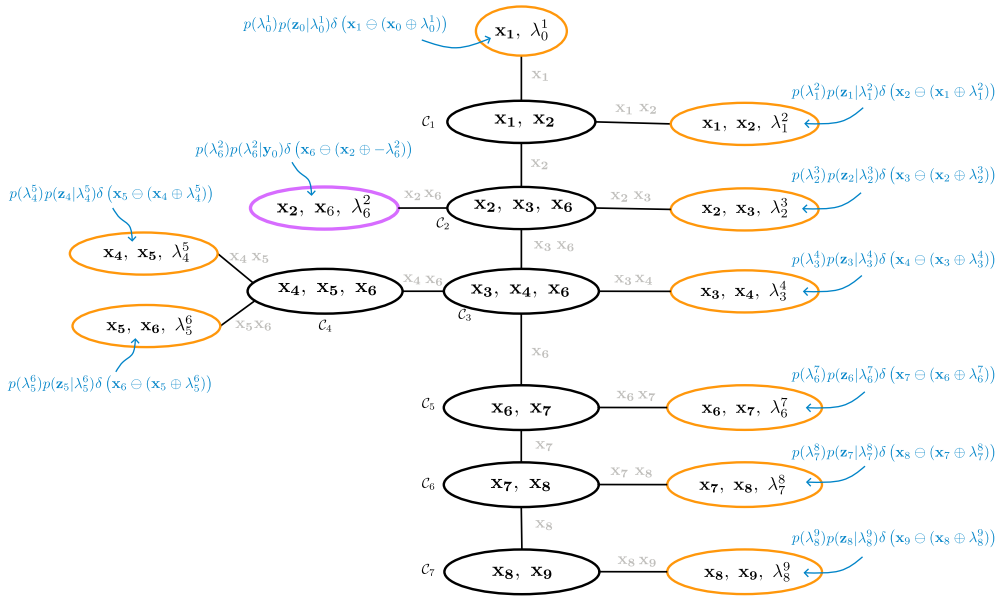


Figure 3.9: Clique tree of the pose-graph Figure 3.3. Cliques involving the relative motions λ are stroked in yellow and in purple in the case of the loop closure. The terms in blue shows in which clique each group of potentials is affected. Cliques in black strokes are initially empty.

Here, the decision variables $\mathbf{x_i}, \mathbf{x_{i+1}}, \lambda_i$ involved in any trajectory part are completely connected in the induced graph ($\mathbf{x_i}$ and $\mathbf{x_{i+1}}$ constitute the Markov blanket of λ_i), so they are guaranteed to belong to a common clique.

The belief $\Psi(\mathcal{C}_i)$ of any clique $\mathcal{C}_i$ associated to a trajectory part (orange and purple in Figure 3.9) is initialized with the potential indicated in blue on figure 3.9, the product $\prod_i \Psi(\mathcal{C}_i)$ being equal to the joint pdf (3.15). Since this toy pose-graph entails no control input u and a single odometric measurement $\mathbf{z}$, the structure of the trajectory corresponds to figure 3.4, so (3.3) simplifies into

$$p(\mathbf{x}_{(+)}, \lambda, \mathbf{z}|\mathbf{x}_{(-)}) = p(\lambda)p(\mathbf{z}|\lambda)p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)}, \lambda). \tag{3.20}$$

The beliefs of the remaining cliques (black in Figure 3.9) are initially empty, i.e., they have a potential of 1 until they receive messages from neighbors. This protocol is described next.

3.2.3 Message Passing Protocol

The next goal is to answer inference questions using the built clique tree by proposing a message passing scheme. Sum-product belief propagation algorithms are selected Darwiche (2009); Koller and Friedman (2009). The messages are passed between cliques by computing the marginal of the clique belief over[6] the separator set. A clique can only send a message to a neighboring clique once it has received all the messages from its other neighbors. This is why message passing can only start from the clique tree "leaves" and process recursively towards a selected clique named "root", then return from the root to the leaves.

[6]To avoid potential ambiguity on the word "over", it means that the variables in the clique belief scope, but not in the separator scope, are to be eliminated.

By virtue of the selected SLAM model, orange and purple cliques are leaves. The remaining black cliques are named "core". The message flowing from clique $\mathcal{C}_i$ to clique $\mathcal{C}_j$ is denoted $\tau_i^j(S_i^j)$, with S_i^j the separator of $\{\mathcal{C}_i, \mathcal{C}_j\}$. Some messages will be shown to be constant. In the other cases, when no analytical result can be leveraged, we revert to the customary message passing computations.

Message passing from leaves to core

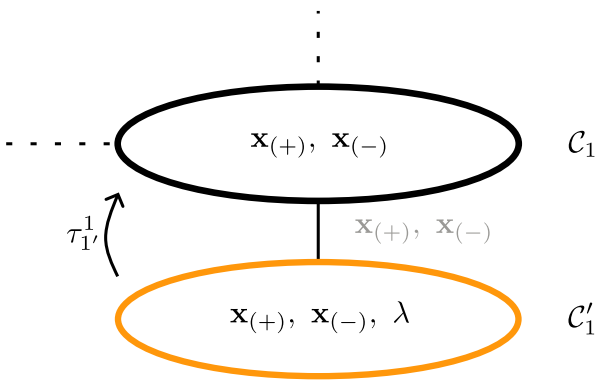


Figure 3.10: Message passing from leaf to core cliques.

Messages starting from leaves must be computed first. Messages flowing from leaves to core are nothing else than the steps described in §3.1 for the elimination of λ. Indeed, from the clique tree excerpt on Figure 3.10, one gets

$$\tau_{1'}^1(S_{1'}^1) \triangleq \int_\lambda \Psi(\mathcal{C}'_1)\mathrm{d}\lambda, \tag{3.21}$$

with $\mathcal{C}_{1'} = \{\mathbf{x}_{(+)}, \mathbf{x}_{(-)}, \lambda\}$, $\mathcal{C}_1 = \{\mathbf{x}_{(+)}, \mathbf{x}_{(-)}\}$, $S_{1'}^1 = \{\mathbf{x}_{(+)}, \mathbf{x}_{(-)}\}$. In view of Subsection 3.2.2, the initial belief of each leaf clique is set to the potentials associated with the data received when the robot is at $\mathbf{x}_{(+)}$(see blue potentials in figure 3.9). Thus, to eliminate λ in (3.21), we finally use a variant of the result (3.13), leading to

$$\tau_{1'}^1(S_{1'}^1) \triangleq p(\mathbf{z})p(\mathrm{Log}({\mathbf{x}_{(-)}}^{-1} \circ \mathbf{x}_{(+)})|\mathbf{z}). \tag{3.22}$$

And, given (3.14), we know that $p(\mathrm{Log}({\mathbf{x}_{(-)}}^{-1} \circ \mathbf{x}_{(+)})|\mathbf{z})$ stands for $p(\mathbf{x}_{(+)}|\mathbf{x}_{(-)}, \mathbf{z})$. Note that this step is purely analytical and requires no computation.

Anti-causal constant message passing for latest odometry measurements

Consider the three lower core cliques $\mathcal{C}_7, \mathcal{C}_6, \mathcal{C}_5$ (upwards) of figure 3.9, reproduced figure 3.11, which is representative of a sequence of odometry measurements. Messages are assumed to have been passed from their respective leaf cliques, according to §3.2.3. For now, the message coming to clique $\mathcal{C}_5$ from core cliques other than $\mathcal{C}_6$ is ignored. Rather, the focus is put on the messages flowing along the direction $\mathcal{C}_7 \to \mathcal{C}_6 \to \mathcal{C}_5 \to \ldots$. From $\mathcal{C}_7$ to $\mathcal{C}_6$, one gets

$$\tau_7^6(\mathbf{x}_8) = \int_{\mathbf{x_9}} p(\mathbf{x}_9|\mathbf{x_8}, \mathbf{z_8})p(\mathbf{z_8}|\mathbf{x_8})\mathrm{d}\mathbf{x_9} = p(\mathbf{z_8}|\mathbf{x_8}) = p(\mathbf{z_8}). \tag{3.23}$$

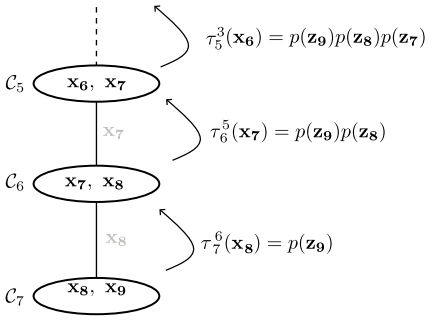


Figure 3.11: Message passing between core cliques representative of odometry.

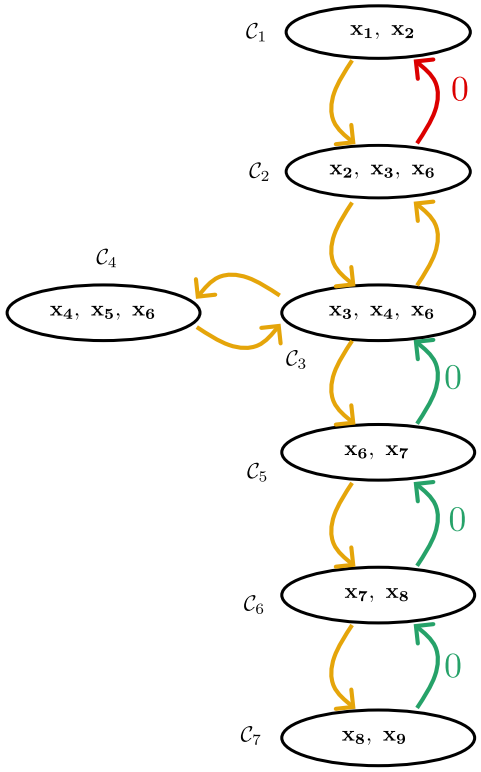


Figure 3.12: Costs of the message passing computations in the clique tree Figure 3.9. The weights of yellow edges is known at runtime. Those associated with green and red edges are structurally zero (see Subsection 3.2.3).

This step is repeated from $\mathcal{C}_6$ to $\mathcal{C}_5$, $\mathcal{C}_5$ to $\mathcal{C}_3$, etc. Consequently, only the evidences $\{p(\mathrm{z_i})\}$ constitute such messages (Figure 3.11). In other words, messages between core cliques in the anti-causal direction are recursively constant. This result is inline with the intuition of many SLAM practitioners that the belief does not propagate backwards in such a structure. In fact, only loop-closures break that structure.

Anti-causal constant message passing for pre-loop trajectory

So far, we have shown that messages in the anti-causal direction are constant, but only for the latest odometry measurements, i.e., the ones occurring after the last loop-closure. However, it can be shown that the computation the message before the loop $\mathbf{x_2} - \mathbf{x_6}$ can also be avoided. Figure 3.12 represents the cliques of the case study problem, except that the undirected edges of the clique have been replaced by directed edges which weights denote the cost of computing messages from the source clique to the destination clique. Yellow edges mean that the cost of

local message passing is non-zero and will be known at runtime. Green edges are associated with a zero-cost in view of the previous paragraph. In what follows it is explained why the red edge is also zero-cost.

Let us consider the variation of the toy problem CBN shown on figure 3.13. The relevant difference with the CBN of figure 3.3, has to do with the parents of the relative motion node λ_2^6, i.e., the loop-closure. Figure 3.3 has considered $\mathbf{x_2}$ and $\mathbf{x_6}$ to be the parents[7] of λ_2^6. By contrast, in figure 3.13, the parents of λ_2^6 are all the intermediate relative motions. Geometrically, both statements are true:

$$\begin{cases} \mathrm{Exp}\left(\lambda_6^2\right) & = \mathbf{x_6}^{-1} \circ \mathbf{x_2}, \\ \mathrm{Exp}\left(\lambda_6^2\right)^{-1} & = \mathrm{Exp}\left(\lambda_2^3\right) \circ \mathrm{Exp}\left(\lambda_3^4\right) \circ \mathrm{Exp}\left(\lambda_4^5\right) \circ \mathrm{Exp}\left(\lambda_5^6\right). \end{cases} \tag{3.24}$$

Figure 3.13 is more appropriate to show that $\tau_2^1(\mathbf{x_2}) = \text{const}$ (zero-cost red edge in figure 3.12), because it becomes graphically evident that: (a) $\mathbf{x_2} \perp\!\!\!\perp \mathbf{y_0}$, and (b), for all $i \geq 2$, $\mathbf{x_2} \perp\!\!\!\perp \mathbf{z_i}$. These d-separations are not depicted by figure 3.3. They graphically prove that no loop-closure observation $\mathbf{y_k}$ influence the estimate of $\mathbf{x_2}$: $(\mathbf{x_2} \perp\!\!\!\perp \mathbf{y_0})$ in figure 3.13. The reasoning can be extended to all poses *before* loop-closures, but not within a loop-closure.

On the other hand, in the context of inference in probabilistic graphical model, the initial representation of figure 3.3 is more convenient for belief propagation algorithms. Indeed, a major flaw of the CBN figure 3.13 is that it leads to a completely different clique tree where sparsity is, essentially, destroyed. As it is known, by construction, that the number of parents per node has a direct effect on treewidth (Darwiche, 2009, §6.8), it can be anticipated that the treewidth will correspond to the value of largest observed loop-closure λ_j^i[8], i.e., $j-i$. This discards the CBN modeled figure 3.13 as a basis for belief propagation. Its single interest is therefore to justify additional zero-cost messages (red arrow in figure 3.12) in clique trees made from the CBN figure 3.3.

Nonetheless, we will return to the merits of figure 3.13 representation in the next chapter (4), where a non-Bayesian view is explored.

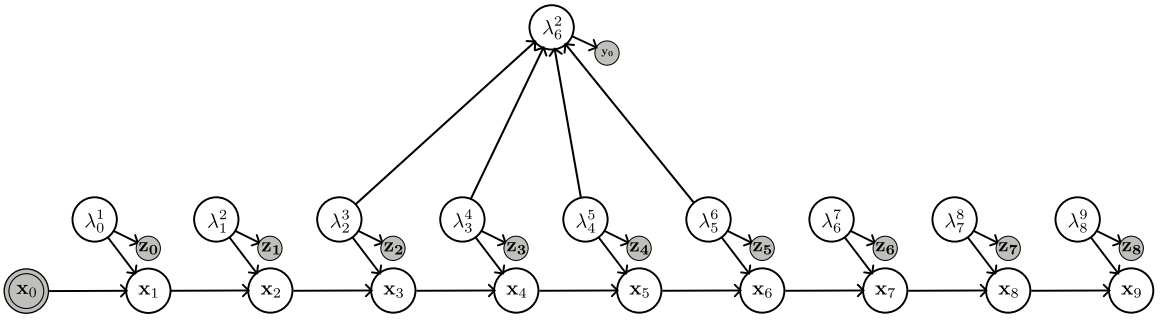


Figure 3.13: Alternative view of the loop-closure were λ_2^6 is made dependent on other λs. On the one hand, this structure allows the exhibition of more independences (e.g., $\mathbf{x_2} \perp\!\!\!\perp \lambda_2^6$). However, this alternative representation has an adverse effect on sparsity as larger cliques will be inevitably produced by the elimination.

[7] This view of the loop-closure is compatible with existing BN for pose-graph, in that a loop-closure creates a statistical link between 2 poses.

[8] For instance, if a pose is re-observed K timesteps after it is created, that alternative view would guarantee a clique which scope size is K.

Root Selection

Belief propagation in the clique tree is achieved through forward/backward message passing to a clique designated as root. Any clique can in fact be the root. But it is clear that choosing a root clique approximately in the "middle" of the clique tree would leverage better parallelism. On the other hand, choosing a leaf clique as the root would create a bottleneck.

Furthermore, the championed CBN view of the SLAM structure exhibits constant message passing in some areas, at zero computational cost. A directed graph can be constructed of all the message passing costs involved in the tree. Its nodes are the cliques, and its directed edges are weighted by the cost of message passing from a source clique to a target clique, once a said source has received all other messages from its other neighbors (Figure 3.12).

The green arrows are 0-weighted because the message passed is constant, as deduced from the structure. The weights of all other arrows can be known analytically only once the mathematical expressions of the clique beliefs are available, independently of the values taken by the variables (e.g., size of matrices that represent the density, cardinality of discrete beliefs). Once the weights of that symbolic graph are evaluated, the goal is to find the center of the graph, in order to best parallelize the belief propagation. An instance of the *betweenness centrality algorithm* was run to find which root is (close-to) equidistant to the leaves, in terms of message computation costs.

3.3 Incremental Inference

The inference is now adapted to incremental cases. The various inference structures described in the previous section and depicted in Figures 3.12-3.9 are studied in view of new measurements. Next Subsections 3.3.1 and 3.3.2 describe in Figure 3.14 the incorporation of new odometry measurements and the handling of a large loop closure $\mathbf{y_1}$.

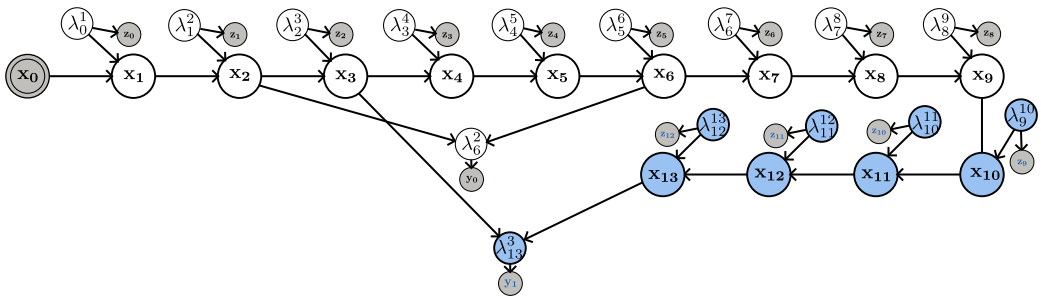


Figure 3.14: Progression of the pose-graph of Figure 3.3. New decisions variables are in blue. New measurements to be incorporated sequentially are bolded in blue.

3.3.1 Additional Odometry Measurements

The incorporation of a sequence of measurements $\mathbf{z_9}, \mathbf{z_{10}}, \mathbf{z}_{11}$ is considered. Three new cliques $(\mathcal{C}_8, \mathcal{C}_9, \mathcal{C}_{10})$ are appended to the clique tree of Figure 3.9. In view of constant message passing in the anti-causal direction, it can be structurally anticipated that a zero-cost message $\tau_8^7(\mathbf{x_9}) = p(\mathbf{z_9})$ is propagated from $\mathcal{C}_8 = \{\mathbf{x}_9, \mathbf{x}_{10}\}$ to $\mathcal{C}_7$. In the causal direction, the new

clique $\mathcal{C}_8$ (and $\mathcal{C}_9$, $\mathcal{C}_9$ thereafter) is passed the message τ_7^8 that does not depend on the new measurements.

Applying a pattern similar to equation (3.23), backwards from clique $\mathcal{C}_{10}$ to $\mathcal{C}_8$, constant messages are exhibited. This proves that the cost of message passing associated with marginal and MAP inference is constant, whenever new (purely) odometric measurements are added. Furthermore, a new clique $\mathcal{C}$ is broken up into $\{\mathcal{C}, \mathcal{C}'\} \leftarrow \mathcal{C}$ whereby $\mathcal{C}'$ is connected to $\mathcal{C}$. The belief of $\mathcal{C}'$ gets initialized with the new measurement density. This is done to maintain the distinction between core and leaves (see Figure 3.9).

3.3.2 Additional Loop closure

A large loop closure is described by measurement $\mathbf{y_1}$. Consequently, a backdoor path between $\mathbf{x_{13}}$ and $\mathbf{x_3}$ is activated. To maintain the clique tree properties (RIP, family preservation), its structure must be partially updated. Algorithm 1 details the process, complemented by Figure 3.15. From the inherited clique tree, the path involving the loop closure decision variables is first identified. For large loop closure, this amounts to a significant part of the clique tree. For measurement $\mathbf{y_1}$, the identified path is the subtree $\mathcal{C}_3$, $\mathcal{C}_5 - \cdots - \mathcal{C}_{10}$. Those cliques are removed and the corresponding joint pdfs form a partial CBN (including pdfs related to $\mathbf{y}$). The fill-in edges issued by the inherited clique tree are maintained (yellow edges in Figure 3.15). The partial CBN is then turned into a new chordal graph along §3.2.2. A clique subtree is assembled and joined to the genuine clique tree. This process shows some similarities with loop closure management in iSAM2.

Algorithm 1: Partial reordering algorithm of a clique tree following a loop closure measurement

Input: Clique Tree T, measurement $\mathbf{y}$
Output: Updated Clique Tree T'

$T_{rm} \leftarrow$ identifySubTree(T, $\mathbf{y}$);
$T' \leftarrow T \setminus T_{rm}$;
$pCBN \leftarrow$ expandDensities(T_{rm}) $\cup\, \mathbf{y}$;
$CG \leftarrow$ triangulation($pCBN$);
$wCG \leftarrow$ weightSetIntersection(CG);
$T' \leftarrow T' \cup$ maxSpanningTree(wCG);

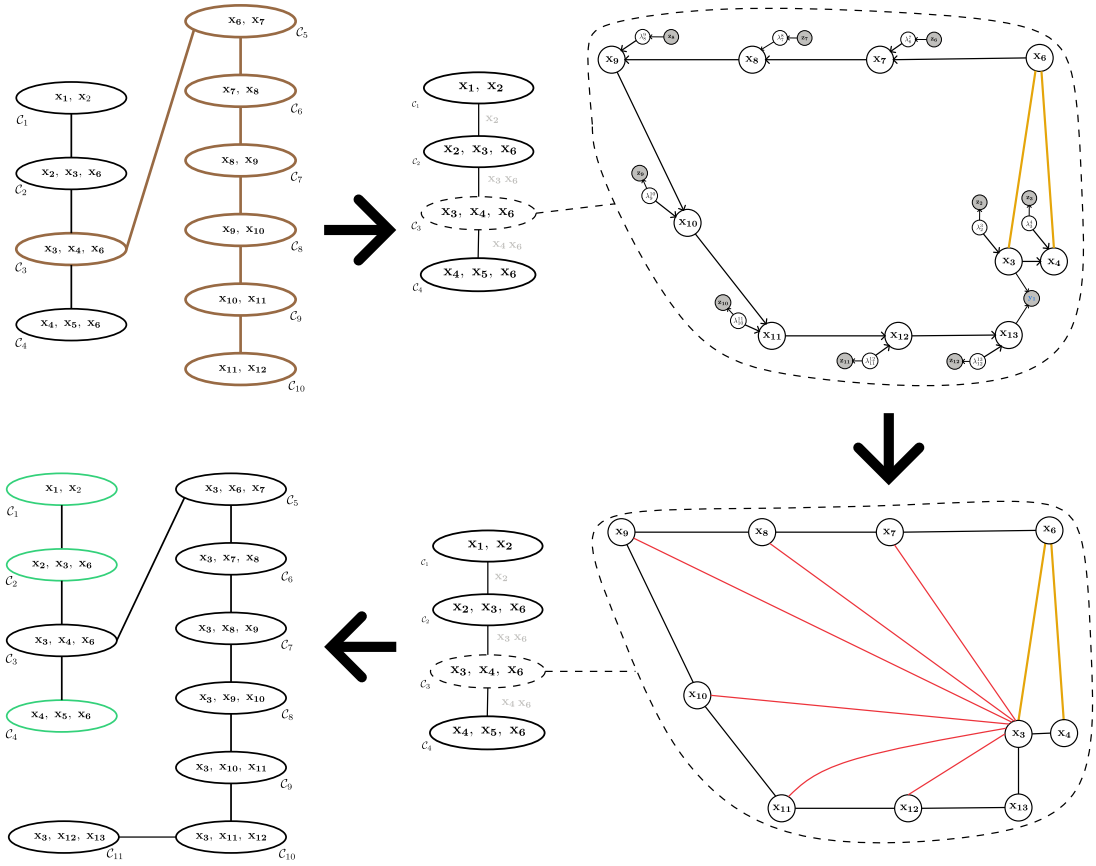


Figure 3.15: The partial reordering algorithm (Algorithm 1) following loop closure. Only core cliques are represented to declutter the depiction. In clockwise order: (a) clique tree before the measurement $\mathbf{y_1}$; (b) the subtree $\mathcal{C}_3,\ \mathcal{C}_5 - \cdots - \mathcal{C}_{10}$ is removed; (c) the joint pdfs in the removed cliques plus the ones associated with new measurements form a partial CBN, fill-in edges (yellow) are inherited from the existing tree structure; (d) the partial CBN is transformed into a chordal graph; (e) the clique tree structure is reformed.

3.4 Numerical application for the case study problem

A numerical experiment in SE(2) is conducted, based on Figure 3.3 assuming additive Gaussian dynamics and measurement noises. The `manif` library (Deray and Solà, 2020) has been used for computations over manifolds. The order of message passing computations described above have been precoded. Figure 3.17 compares the matrix view and Bayes tree of the problem. The Bayes tree structure is similar to the core of the clique tree structure of Figure 3.9. The beliefs in the cliques take the form of dense matrix-vector structure (information matrix and information vector). The messages between cliques (marginalization) are computed using Schur complements. As mentioned above, the elimination of the relative motions λ is analytical, and does not appear explicitly in the implementation. The only explicit information in the implementation concerns which messages are zero-cost, as that a number of message computations can be avoided. We ignore the cost of incorporating the messages into clique beliefs (matrix additions).

Since the densities are only Gaussian locally with respect to the decision variables, the problem is solved iteratively for some δx around a composite linearization point $\tilde{\mathbf{x}}$, until convergence. The message passing is visually depicted figure 3.16.

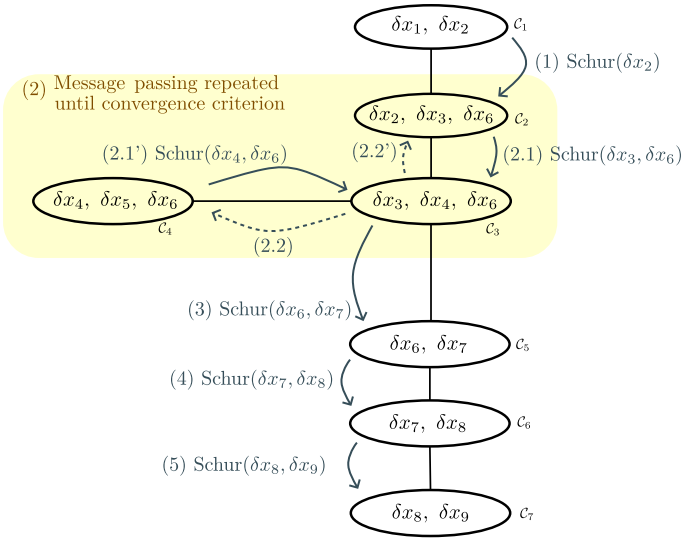


Figure 3.16: Depiction of the message passing between the cliques. First, the message $\mathcal{C}_1 \to \mathcal{C}_2$ is sent (only once). Clique $\mathcal{C}_3$ is selected as the root node. In the K iterative steps required during the run, the messages only need to flow from $\mathcal{C}_2 \to \mathcal{C}_3$, and from $\mathcal{C}_3$ back to $\mathcal{C}_2$ (dashed arrow). The backward message $\mathcal{C}_3 \to \mathcal{C}_2$ only occurs when $\mathcal{C}_4$ is received, and without sending back the message $\mathcal{C}_2 \to \mathcal{C}_3$ to avoid double counting. Concurrently, the message $\mathcal{C}_4 \to \mathcal{C}_3$ flows and then back $\mathcal{C}_3 \to \mathcal{C}_4$ (dashed arrow) once $\mathcal{C}_3$ has received the message from $\mathcal{C}_2$. Then, the potentials in the clique in the yellow area re-linearized, and this process is repeated, until convergence condition is met. A number of re-linearizations and message computations can thus be avoided, compared to a standard matrix based implementation of MLE SLAM. The exact number depends on the structure of the graph and the number of iterative steps K.

We fix the measurement noise as $\Sigma \triangleq \begin{bmatrix} 90 & 200 & -50 \\ 200 & 470 & -120 \\ -50 & -120 & 100 \end{bmatrix}$ (including for the loop closure) and we make it so that the randomly generated trajectory stays within a 10×10 square meters area. Applying the message passing protocol, step (2) in figure 3.16 requires between 3 and 4 iterations in our tests. For each generated experiment, our MAP results from this protocol is similar to the solution of the matrix-based view of MLE SLAM (figure 3.17-(left)), which was expected. Compared to the matrix view of the problem the proposed approach enables less re-linearizations thanks to zero-cost message passing (green are red arrows in figure 3.12). This has been tested over several runs with different noise characteristics.

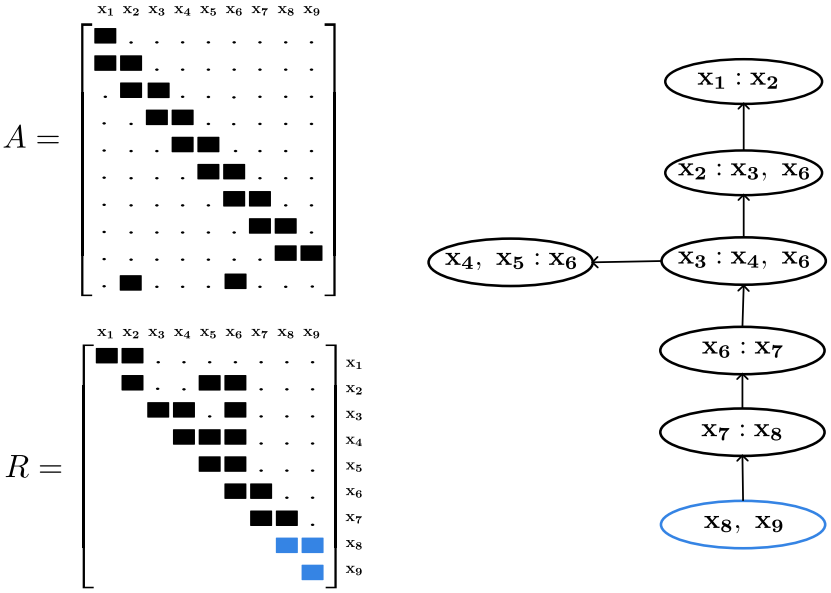


Figure 3.17: Matrix view Kaess et al. (2008); Ila et al. (2017); Wang et al. (2018) *vs* Bayes tree Kaess et al. (2010) of the problem of Figure 3.3. For both approaches, the root is highlighted in blue. Notice the similarity of the core part (i.e. leaves excluded) of the (undirected) clique tree Figure 3.9 with the Bayes tree.

3.5 Experiment on a larger graph

An experiment is conducted on the dataset 'Manhattan M3500' (Olson et al., 2006; Carlone and Censi, 2014). It features a robot travelling within a city block urban environment, which measures its own displacement thanks to a fairly poor odometry and is able to detect loop closures. Its view from above is portrayed figure 3.18. Our motivation for this dataset, in the context of this chapter, is the causal link between the right side (older) and left side (more recent), as highlighted in figure 3.18. The node $\mathbf{x}_{2947}$ is chosen as a separator. This loosely-coupled situation that can be representative of many real world SLAM problems. Constant message passing is exhibited from the loosely connected left side towards the right side. This means, in accordance with the presented message passing scheme, that inference can be conducted separately on both sides. In some sense, this is a causal dissection.

The MAP inference conducted for the left side expresses as:

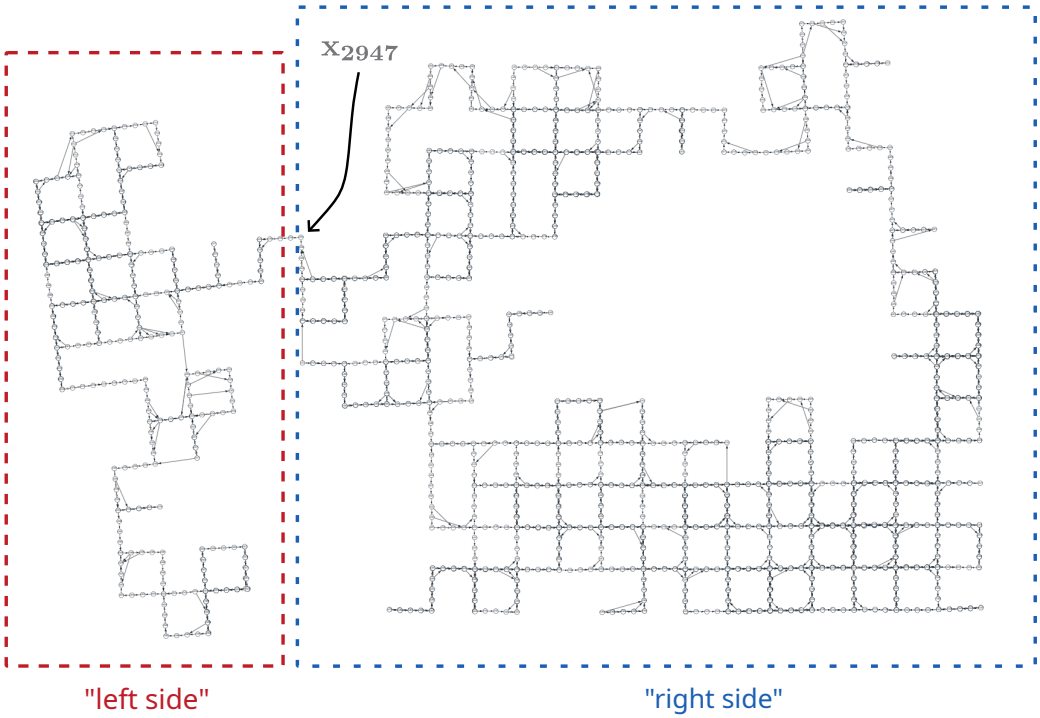


Figure 3.18: Depiction of a *solved* CBN of the 'Manhattan 3500' dataset. Left and right side are distinguished and solved separately in this work.

$$\arg\max_{\mathcal{X}_{\text{left}}} \; p(\mathcal{X}_{\text{left}}|\mathcal{Z}, \text{do}(\mathbf{x_{2947}} = \varepsilon)),$$

where $\mathcal{X}_{\text{left}} = \{\mathbf{x_i}, i \in (2947, ..., 3500)\}$, and $\mathbf{x}_{2947}$ is set at the origin (via Pearl's do-operator notation). The clique tree can be visualized on a web support, described hereafter.

3.6 Visualization of the data structure

An animated visualization of the inference process (from initial guess to MAP) is provided on a web support (figure 3.19 and 3.20). It consists in the visualization of the CBN (which nodes are positions in the trajectory) and the associated clique tree (as a force graph). We provide below links to visualize several trajectories (from a small, to medium, to large graph)[9]:

- The visualization of a small graph `https://jtari.net/sub_cbn/small`.
- The left side of the M3500 dataset `https://jtari.net/sub_cbn/left_side`.
- The right side of the M3500 dataset `https://jtari.net/sub_cbn/right_side`.

In the CBN/trajectory screen, pressing the keys 'backspace', 's' and 'space' respectively increase, decrease, and reset the dimensions of the element to declutter the view. Pan and zoom are available via the mouse. To replay the animation, simply refresh the page.

[9]NOTE FOR REVIEWERS: the links proposed here are temporary for the draft version. For the final version, the content will eventually be served on a more permanent institutional website.

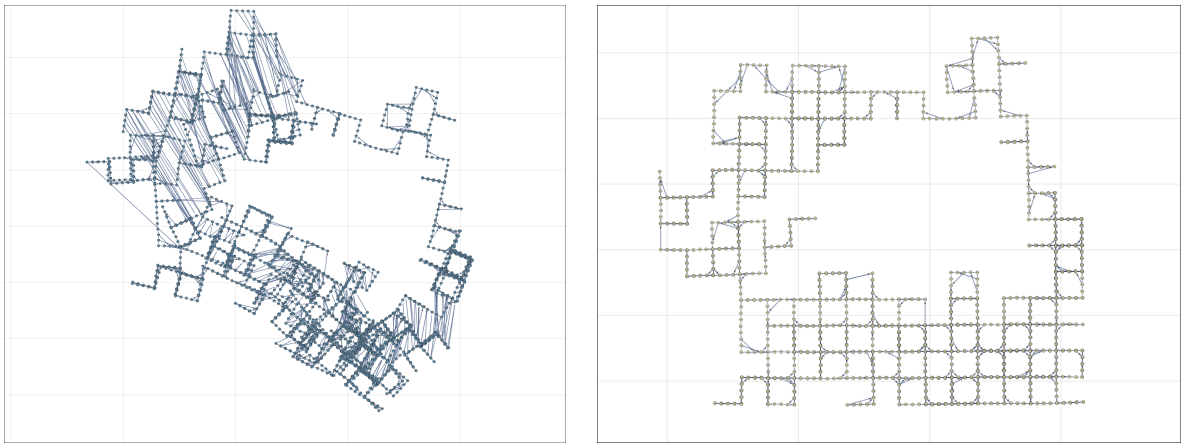


Figure 3.19: Screenshot of the animation step of adjustment of the graph.

The corresponding *clique tree* can be accessed through the drop-down menu (burger shaped button) located in the top left of the screen. The *clique tree* visualization is implemented as a force graph. Hovering above a given variable makes the running intersection property apparent (black stroke in figure 3.20). However, due to the large number of nodes, which implies a large number of 'html' elements, the clique tree of the right side of the graph (the largest one among examples provided) is *currently* very slow.

The current implementation is available here: https://github.com/JoelTari/CarlitViz and is based on the "D3JS" Javascript library. Under the hood, our solution (C++ and Julia scripts) produces a "json" file containing the results. Future work will provide a user interface as well as additional features for our next contribution (CARLIT) presented in the next chapter.

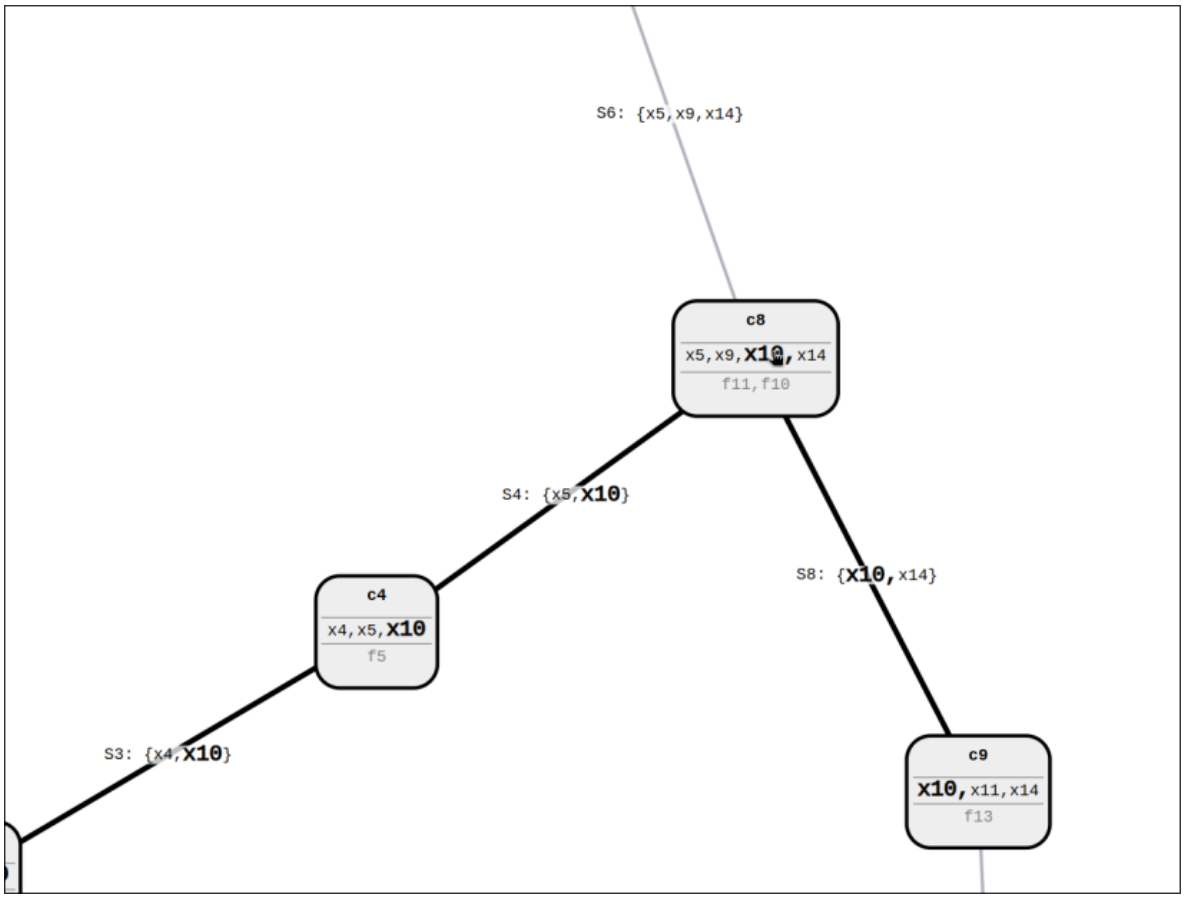


Figure 3.20: Clique tree visualization. Hovering over a given variable makes the running intersection property apparent. This screenshot shows the RIP of $\mathbf{x}_{10}$, i.e., all cliques and separator containing this variable. This clique tree visualization has proved useful observing how loop closures that overlap each other produce a greater treewidth.

3.7 Discussion

In this chapter, we presented a new conceptualization of the SLAM problem and belief propagation based inference. Contrarily to other strategies, causal intuition in the PGMs is enforced, that enables to benefit from zero-cost backward propagation along the robot trajectory (in some sense, the intuition that supports preintegration schemes Forster et al. (2017) is made explicit). A symbolic graph can be set up, whose edges are message passing costs, that helps to decide which root is best suited to each inference question. No assumption is made on the nature of the involved pdfs as long as analytical calculations and computational cost evaluations can be conducted (they could be nonparametric Fourie et al. (2016), etc.). Compared to state-of-the-art, additional unknowns (λ's) are needed, but they can be discarded analytically at no cost. The method shows theoretical advantages, and besides, the proposed "rootless" clique tree may be an asset for collaborative SLAM.

On the one hand, more can be done on this basis, and a thorough timing analysis could be done to flourish the approach. However, the structure only addresses fully one aspect of the issues mentioned in 2.3, which is $\mathfrak{f}_2$ (no backward influence of odometry). Consequently,

we rather spend more focus on the important scientific question, which concerns the research of a new structure for the SLAM problem able to reproduce more basic aspects of spatial understanding.

One of the motivations behind this CBN SLAM contribution was to maintain compatibility with the modern probabilistic SLAM formulation, which has many successes (see §2.2). Unfortunately, most of the other issues identified in 2.3 (that is $\mathfrak{f}1, \mathfrak{f}3, \mathfrak{f}4, \mathfrak{u}1, \mathfrak{u}2$) are also transferred to this CBN approach of SLAM.

The visualization system we presented highlights another concerning fact: the size of the cliques gets considerably bigger as the size of SLAM size grows. In the biggest cliques of medium-sized problems (e.g., Manhattan 3500), grouped variables may be separated by an appreciably big distance so that there may be no reason, *a priori*, to think of them as potentially gathered. The complexity issue comes from the known theoretical result that the size of the largest clique is the prime indicator of the inference cost (see §2.3.3 and §3.2.1).

A profound ramification of this last point comes from the fact that this issue is not specific to any clique based representation. We have seen that the sparsity pattern between the decomposed sparse matrix structure and the clique tree structure are similar (see figure 3.17). The problem is traced back to the key assumptions 2.1 and 2.2 of probabilistic SLAM. These assumptions lead us to solve a system heavier that initially thought, whichever support is used (decomposed sparse matrix, clique tree, etc...).

To overcome these issues, our next solution is to come up again with a different structure. Consider again the interesting structure on figure 3.13. Its single interest was to make a point about an independence relation, yet it was discarded because it leads to a solver with unfavorable metric (high treewidth). In the next chapter, inspired by a variant of this figure, we propose a novel view of the SLAM problem which finally articulates the Laplacian conception (1.5.3), the theory of errors, and Pearl's paradigm on causality.

Chapter 4

Causal SLAM for Internalized Topography

This chapter attempts to synthesize our initial historical analysis, our observations on modern SLAM, and our understanding of causality and graphical models. This results in an extra probabilistic view of the SLAM problem.

CARLIT (Causal Approach to Represent Localization and Internalized Topography) is the name given to our approach. It is based on the Structural Causal Model (SCM) framework of Pearl & colleagues (Pearl, 2009). Its privileged mode of graphical representation is the causal graph.

Figure 4.1 shows different modes of representation of the small toy pose-graph scenario studied in chapter 3. Each representation depicts a different approach to the SLAM problem (see caption). In this chapter, we will progressively introduce formally the graphs (c) and (d) which concern the CARLIT approach. But first, notice that the graph (c) is reminiscent of the graph described in figure 3.13. The main difference versus figure 3.13 is that additional ζ-variables have been incorporated, which represent explicitly the external errors affecting the measurements.

The overall goal of our approach is to address the faithfulness and feature issues identified in section 2.3 and, then, only by opportunity, to discuss the improvements brought in terms of complexity.

Since this approach is appreciably different from existing SLAM formulations, we present the process step by step. Section 4.1 will formally introduce the SCM framework and discuss its merits. Section 4.2 studies the causal representation of pose-graph scenarios. This will give us the opportunity to restate the fundamental concept of loop closure. In section 4.3, we detail our constrained least squares solution to a CARLIT pose-graph. Nonetheless, pose-graph problems are only a limited subset of SLAM problems. This is why in section 4.4, we address the question of mapping with landmarks. The notion of internalized topography ('IT' in 'CARLIT') is presented. Conceptually, it is an important section. To improve the sparsity pattern of the solver, the composition of summarized motion is presented in section 4.5. In section 4.6, we showcase on a small example how covariance queries can be answered in such quasi-deterministic framework. Finally, concluding remarks and perspectives are given in section 4.7.

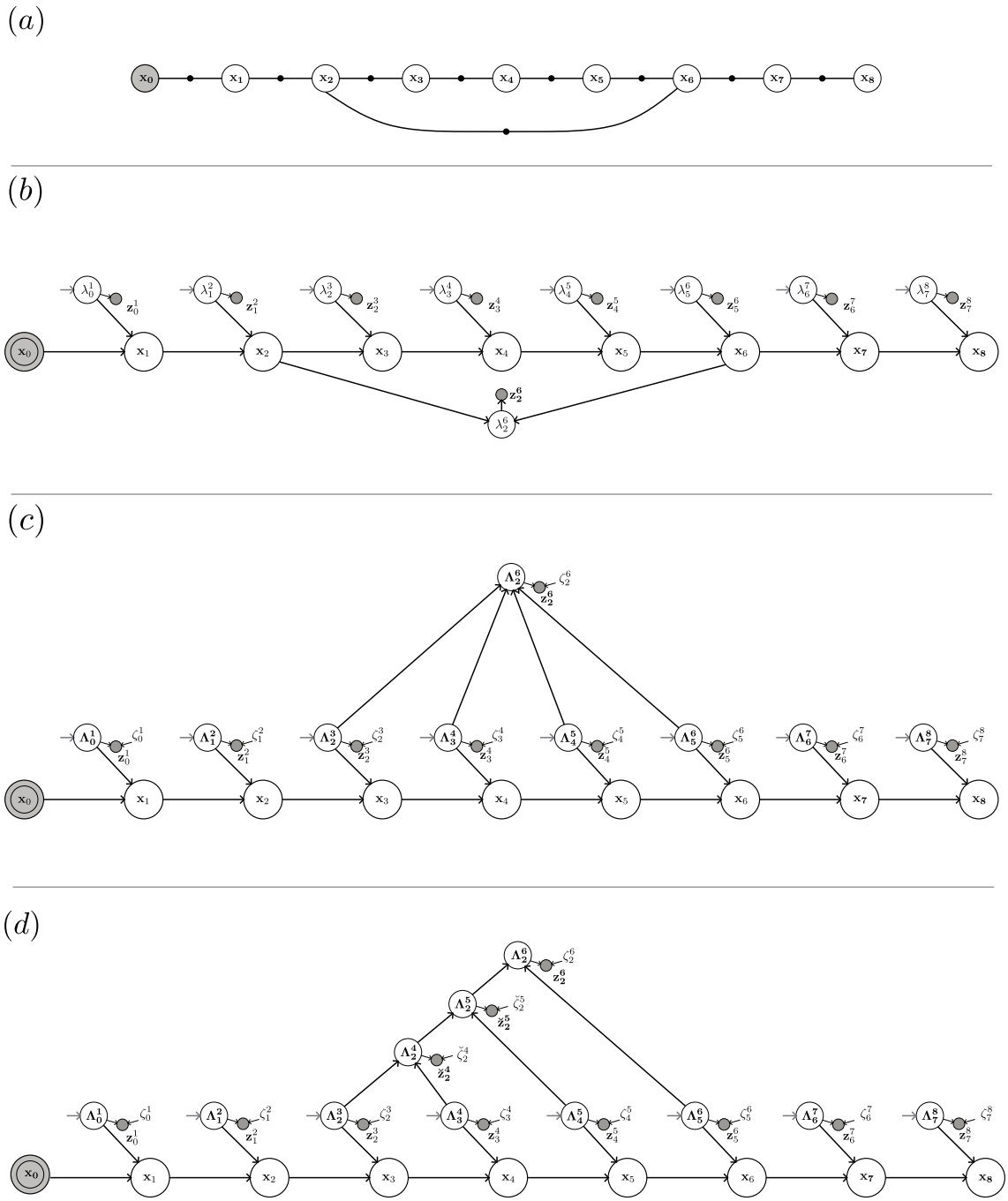


Figure 4.1: Various representation of the same toy problem. (a) is the factor graph representation of probabilistic SLAM. (b) is the CBN representation we introduced in chapter 3. (c) is the causal graph of the CARLIT approach. (d) is the wrapped causal graph for which compositions and virtual measurements are introduced to safeguard sparsity; this will be described in section 4.3.

4.1 The Structural Causal Model framework

4.1.1 On Structural Causal Models and their implied conceptual hierarchies

As we assume the reader to be less familiar with causal framework, we present in this section the Structural Causal Model (SCM) and its ramifications when approaching the SLAM problem.

The definition given is based on Pearl (2009); Pearl and Dechter (2013); Pearl et al. (2021).

Definition 4.1 (Structural Causal Model (SCM))**.** *An SCM is a* $4-$*tuple,* $\text{SCM} =< V, \mathcal{E}, p(\mathcal{E}), \{f_i\} >$ *where:*

- V *is the set of endogenous variables.*
- $\mathcal{E}$ *is the set exogenous variables. They are generally the disturbances or the error terms. They will commonly be referred to as 'errors' in this work.*
- $p(\mathcal{E})$ *is the joint pdf over the errors.*
- $\mathcal{F} = \{i \in (1, \dots, |V|) : f_i\}$ *is a set of deterministic functions such that:*

$$\forall v_i \in V,\ v_i := f_i(\text{pa}_i, \mathcal{E}_i).$$

The set $\text{pa}_i \subset V$ *are the parents of* v_i*.* $\mathcal{E}_i \subseteq \mathcal{E}$ *is the set of errors affecting* v_i*. The symbol* $:=$ *denotes the assignment operator, meaning that* v_i *is determined by the right-hand side of the assignment operator.*

In a strictly causal deterministic view of the world, the errors $\mathcal{E}$ do not result of inherent randomness but rather of ignorance of some underlying processes. As noted by Laplace (and before him Condorcet, Maupertuis, d'Hollbach and Boscovich, see §1.5.3), since knowledge of all laws in Nature is out of reach, and/or for convenience reasons[1], we do not explicitly attempt to detail the causes of the errors. The formalism of probability theory captures the behavior of these exogenous variables. As was remarked in chapter 1, this is in fact precisely how probability theory was introduced in the first paragraphs of *Essay* (Laplace, 1814a).

Definition 4.2 (Causal graph)**.** *from Pearl (2009, §7.1).*
Every SCM (M) *can be associated with a directed graph,* $\mathcal{G}_M$*, in which each node corresponds to a variable in the set* $\mathcal{E} \cup V$*. The directed edges point, for each* $v_i \in V$*, from members of* pa_i *and* $\mathcal{E}_i$ *toward* v_i*.*

Remark 4.1. *From definitions 4.1 and 4.2, error nodes have no parent in the graph. They are root nodes.*

The graphical criterion of d-separation applies for $\mathcal{G}_M$ in the same way as it does for Bayesian networks (BN). Causal graphs could conceivably be cyclical, but they are all acyclical in this work (DAGs). A causal graph is a graphical abstraction of the cause-and-effect relations

[1] One might be willing to ignore phenomena he/she knows about for convenience reasons. Such phenomena often ignored can be countless and are left captured by the exogenous variables. They include, e.g., the known effect of frictions, known play in the robot joints, effect of shadows on perception systems, etc.

that the SCM embodies, and thus, of the causes-and-effects *assumed* by the modeler of a given SCM. There comes its attractiveness, from the fact that it is able to convey transparently assumptions through the absence of arrows, the direction of arrows and d-separation. In relation to d-separation and the fact that it is a directed graph, it is worth stressing the conceptual difference between a BN and a causal graph, especially given the fact that SLAM practitioners are relatively familiar with the former. The BN, as noted in §2.1.2 (remark 2.3), is only a representation of the patterns of independences in a corresponding joint pdf. The direction of the arrows in a BN do not imply causal assumptions, since the joint distribution can be reformulated in the form of various products, by Bayes rules. Thus, there are as many possible BNs for a joint distribution as there are applications of the Bayes rules in this distribution. Hence, BN cannot be the only basis for problems that necessitate the manipulation of causal considerations.

One such potential manipulation is an intervention. It consists in setting a variable v_j in V to an arbitrary value to study the effect. By "setting the arbitrary value of v_j", it is meant that the value of variable v_j is no longer assigned by its parents. Instead, v_j is controlled by something external, intervening from outside the scope of the SCM. The analytical operator $\text{do}(v_j)$, and the double circled nodes in the graph, are both used to denote an intervention (e.g., see $\mathbf{x_0}$ figure 4.1-(b), -(c) and -(d)). A DAG can interpret interventions only if the arrows represent cause-and-effect relations and not just Bayesian conditionals. The Causal Bayesian Network (CBN - used in the previous chapter), for its part and as its name suggests, remains a representation of a joint pdf enforcing causal constraints. It opens the opportunity to model interventions. The causal graph (definition 4.2) goes beyond the BN representation: it abstracts the structural equations of the SCM (definition 4.1).

Definition 4.3 (Set of observations and set of interventions)**.** *Let $\mathcal{Z}$ and $\mathcal{A}$ respectively denote the set of observations and interventions applied on a subset of the endogenous variables V of an SCM.*

- *The set $\mathcal{Z}$ correspond to the variables being observed, i.e., conditioned on in the inference process. Its corresponding nodes in the causal graph $\mathcal{G}$ are depicted in gray.*
- *The intervention set $\mathcal{A}$ corresponds to the variables appearing in the do-operator,* $\text{do}(\mathcal{A})$*. Its corresponding nodes in the causal graph $\mathcal{G}$ are depicted in gray with a double-circle. Unlike other endogenous variables, nodes in $\mathcal{A}$ do not have parents.*

In this work, the inference objectives are achieved by first computing the MAP estimate of the errors, the peak of the posterior pdf $p(\mathcal{E}|\mathcal{Z}, \text{do}(\mathcal{A}))$. Then, the robot poses $\mathcal{X}$ and the landmarks $\mathcal{L}$ are deduced. Under precautions to be detailed, these are also the MAP estimate of $p(\mathcal{X}, \mathcal{L}|\mathcal{Z}, \text{do}(\mathcal{A}))$.

Remark 4.2 (Inference with interventions)**.** *In the causal inference literature, especially in comparative studies, inferring the effect of interventions can be difficult because implementing an intervention may be costly, impossible, or even unethical (e.g., typically, one cannot force a cohort to smoke in order to study the effect of smoking). In those situations, a set of methods called* do-calculus *(Pearl, 2012) is leveraged. While* do-calculus *is a powerful tool to keep in mind for future work, we will see that setting the interventions* $\text{do}(\mathcal{A})$ *poses no difficulties in this work. For instance, the implementation intervention* $do(\mathbf{x_0})$ *is a matter of tweaking a configuration file. In fact, one can ask why not just cast $\mathbf{x_0}$ and other elements of $\mathcal{A}$ as observations. The reason for this is twofold: first, in all rigor, only interventions have no*

parents, while observations do (because observations have cause(s) in the model), and $\mathbf{x_0}$ *has no parents; second there are conceptual ramifications concerning the causal interpretation in the model, to be discussed for each entity in* $\mathcal{A}$*. In this instance, setting of* $\mathbf{x_0}$ *is an arbitrary choice, external to the model.*

The SCM approach proposes a coherent framework to answer inferential queries on (1) an associational level (e.g., estimating the marginal over the last pose given observations), (2) an interventional level (e.g., what happens if we send a control signal u?), and (3) a counterfactual level (e.g., what would have happened had a control signal u' been sent, given the fact that we did send control signal u?). These three layers (or rungs) constitute the *ladder of causation*, or the *Pearl Causal Hierarchy* (PCH). Bareinboim et al. (2022) emphasized the need for causal models to answer estimation queries on all 3 layers of the PCH.

4.1.2 On the relevance of SCMs for SLAM

In the robotics context, one can naturally expect great benefits in designing a causal SLAM system if his/her objective includes answering interventional and counterfactual queries about the environment[2]. But independently of that fact, one can argue that a SLAM system formulated on the first layer of the PCH, i.e., via a joint pdf represented by a BN or by a factor graph, is sufficient to solve questions related to SLAM, such as joint state estimation. However, in this chapter, we intend to show the merits of a causal model for SLAM, i.e., of a system based on the whole PCH, *even* when the SLAM problem is restricted to the first rung of the PCH, i.e. to state estimation conditioned on local observations.

The idea according to which the SLAM endeavor can be formulated as an extra probabilistic problem is actually far from novel. This can be identified already in one of seminal paper on robotic localization and mapping (not yet termed as SLAM) by Chatila and Laumond (1985). Table 4.1 shows a lexicometric analysis of this paper. For each keyword of interest, we give the occurrence and corresponding layer in the PCH. This corresponding layer means that the considered keyword advocates for a problem formulation via a framework of layer i (in the PCH) or above.

This paper describes general ideas to correct the errors in order to model and understand the environment. The notion of (spatial) understanding is presented as being the ultimate goal. The word "model", which occurs most often, could refer to several constructs (e.g., a probabilistic model, a geometric model); however, in conjunction with other words (such as "perception", "error", "structure"), it then tends to suggest a richer conception. It can be interpreted that the potential solution system, i.e. the "model" the authors were after, should be capable of answering the broadest types of inference queries, not limited to state estimation. Overall, we can see that the vocabulary employed by Chatila and Laumond (1985), if put in an SCM context as a thought experiment, suggests the need of a third layer, e.g. a structural model. One can also notice that the vocabulary significantly differs from SLAM contributions, from the 'classical age' (as coined by Cadena et al. (2016)) to today, which words are generally focused on the first layer, i.e. probability theory considerations and state estimation techniques. Conversely, what makes Chatila and Laumond (1985) paper intriguing is precisely its non-commitment to a purely probabilistic view. It should be recalled here that Bayesian networks were just invented at the time (Pearl, 1985) and that further research on the SCM framework would not take shape until around the late 90s. The formalization of

[2]This will be addressed in future work.

Keyword	Count	PCH layer
model	100+	-
percep* + perceive	58	3
frame	46	2
uncertain*	33	-
error	25	3
(in)accuracy	18	1
system	15	-
structure	15	3
consistent*	10	1
prediction	10	1
knowledge	8	3
understand	6	3
data	4	1
prior	1	1
probabil*	1	1
likel*	0	1
learn*	0	1
estim*	0	1
correlation	0	1
causal	0	3

Table 4.1: Lexicometric analysis of Chatila and Laumond (1985) and would-be correspondence with the PCH layers in the third column.

causal SLAM through SCM can now be explored, in view of: the issues mentioned in chapter 2, the compatibility in spirit with the historical approaches of geodetic adjustment in chapter 1.

Before proceeding, an important disclaimer must be raised concerning the nature of assumptions made in the next sections. It should be understood beforehand that causal assumptions are stronger than statistical assumptions, in that the former imply the latter. Usually, the implied statistical relations are testable[3]. However, 'purely' causal parts often rely on arguments and our qualitative judgements, and no attempts are made to prove them. For instance, consider the consecutive robot poses $(\mathbf{x_i},\ \mathbf{x_{i+1}})$, and the relative motion noted $\mathbf{\Lambda_i^{i+1}}$ between them. It is algebraically equivalent to say that $\mathbf{x_{i+1}} = \mathbf{x_i} \circ \mathbf{\Lambda_i^{i+1}}$ and that $\mathbf{x_i} = \mathbf{x_{i+1}} \circ (\mathbf{\Lambda_i^{i+1}})^{-1}$. However, only the assignment $\mathbf{x_{i+1}} := \mathbf{x_i} \circ \mathbf{\Lambda_i^{i+1}}$ is assumed to be causal here. This assumption, already used in chapter 3, relies solely on the intuitive and largely shared notion of temporal precedence as an indicator of causal directionality.

Causal assumptions are introduced progressively in the rest of the chapter. In the next section, we explore a causal representation of pose graph SLAM. In section 4.3, we address a

[3]Note that this is usually true for sparsely connected causal networks, but may not hold for small & completely connected networks. For instance, this causal graph: $\{X \to Y; Z \to X; Z \to Y\}$ inspired from (Pearl, 2019, p3) is made by definition of causal assumptions, but does not contain any testable conditional independence statement. And hence this model cannot be refuted by studying the statistics of the experimental data.

new method for adjustment for this representation. The so-called CARLIT approach is subsequently extended to landmarks in section 4.4. In section 4.5, we explore how to compose together intermediary motions in an effort to lower complexity in the presence of overlapping loop-closure. In section 4.6, we show how to process covariance queries in such quasi-deterministic system. Our conclusions and perspectives for causal SLAM, and causal robotics, are given in section 4.7.

4.2 Representation of pose graph problem

Assumption 4.1. *In this work, all errors of the set $\mathcal{E}$ are mutually independent*[4]*, see also (Pearl, 2009, def 2.2.2):*

$$p(\mathcal{E}) = \prod_i p(\mathcal{E}_i) \ .$$

Of course, the factorization of $p(\mathcal{E}|\mathcal{Z})$ is DAG dependent. Furthermore, every $\mathcal{E}_i$ only affects one node in the graph (which is not always the case in causal inference literature), and, unless stated otherwise, the errors are always zero-mean Gaussian.

4.2.1 Simple robot trajectory (no loop-closure)

Box 8: Simple trajectory scenario for a controlled robot.

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot is fed sequentially 4 independent control inputs $\mathcal{U} \triangleq \{u_0, u_1, u_2, u_3\}$. We assume absence of inertia in the robot motion, no false sightings, no outliers, etc.

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}\}$.

Consider first a sequence of N robot poses $\{\mathbf{x_0}, \dots, \mathbf{x_N}\}$, as described per the toy scenario (Box 8), for which no form of loop-closure observation is available. The SCM approach is introduced via the interpretation of the following statement:

Causal Statement 4.1 (Robot Trajectory)**.** *The trajectory of a robot is the ordered set of poses $\mathcal{X} \triangleq \{\mathbf{x_0}, \dots, \mathbf{x_N}\}$. It is determined by a sequence of N relative motions $\{\mathbf{\Lambda}^i_{i-1}\}_{i\in 1\dots N}$ cumulated on a pose $\mathbf{x_0}$.*

On the basis of this toy scenario (Box 8) and causal statement 4.1, we build the SCM (M_1) by stating the set of functionals:

$$(M_1): \begin{cases} \forall i \in (1, \dots, N), & \mathbf{\Lambda}^i_{i-1} := \mathrm{Exp}\left(u_{i-1} + v_{i-1}\right), \\ \forall i \in (1, \dots, N), & \mathbf{x_i} := \mathbf{x_{i-1}} \circ \mathbf{\Lambda}^i_{i-1}, \end{cases} \tag{4.1}$$

where it is assumed that the elements $\mathbf{x_i}$ and $\mathbf{\Lambda}^i_{i-1}$ are Lie group elements (typically, of the Special Euclidean groups $SE(2)$ or $SE(3)$); and the control inputs u_i and corresponding errors

[4]The notation $\mathcal{E}_i$ is kept for meta purposes, it abstracts the different types of errors v, ζ, ϵ, ν that exist in this chapter, owing to the variety of robot sensors and control signals.

υ_i are expressed in a Euclidean space isomorphic to a Lie algebra (typically $\mathfrak{se}(2)^\wedge$ or $\mathfrak{se}(3)^\wedge$). Notations are similar as in the previous chapter (Solà et al., 2021), where $\mathrm{Exp}(\cdot)$ and $\mathrm{Log}(\cdot)$ transform elements to/from the manifold from/to the Euclidean space isomorphic to their Lie algebra (see figure 3.1). Note that half of the structural equations of (4.1) are deterministic.

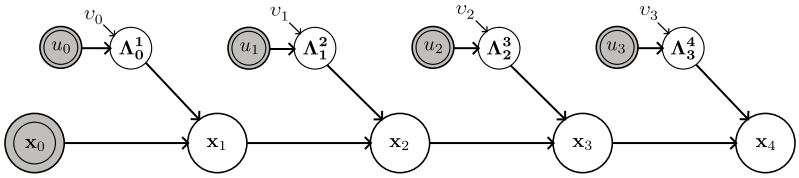


Figure 4.2: Representation of the toy scenario (Box 8) for $N = 4$. The d-separation criterion shows that all errors υ_i are independent from each other.

The corresponding causal graph $\mathcal{G}_{M_1}$ on figure 4.2 abstracts (M_1). We can learn a few important points from it:

- First, the errors on the control inputs are, under this model, independent of each other: for any $i \neq j$, $(\upsilon_i \perp\!\!\!\perp \upsilon_j)_{\mathcal{G}_{M_1}}$ is verified in figure 4.2. Consequently, no correction is possible through adjustment, which is useful information in itself. Remark that this statement is derived from modelling assumptions, and not claimed from start, which is partly the purpose of our efforts on representation.

  In a noise-free problem, the poses can be obtained by applying the functionals (4.1) recursively, and considering the zero mean of each error:

$$\forall i \in (1 \dots N), \quad \mathbf{x_i} = \mathbf{x_0} \circ \prod_{j=1}^{i} \mathbf{\Lambda}_{j-1}^{j} = \mathbf{x_0} \circ \prod_{j=1}^{i} \mathrm{Exp}\,(u_{j-1}) \ ,$$

  where the product $\prod$ stands for the composition $\circ$ developed on the right, as it is non-commutative on the usual Lie groups encountered in SLAM.

- Second, the drift, i.e., the growing uncertainty on the localization of the robot, is easily visualized on the graph $\mathcal{G}_{(M_1)}$. As each pose $\mathbf{x_i}$ is only influenced by independent error terms that are its ancestors, i.e., the set $\{\upsilon_j\}_{j \in 1 \dots i-1}$, then the bigger the value i, the bigger the set of error terms influencing $\mathbf{x_i}$.

- The third point concerns the modeling of control signals. They are modeled as interventions, which amounts to consider that their values are arbitrary, and that their causes are out of scope, as per remark 4.2. However, in practical real world scenarios, there is always an intent underlying the control signals sent. This could be: a feedback control law or some higher level planning, or the desire of an operator via a remote device. Any of these causes could, with some efforts, be causally incorporated in the SCM, so that some solutions could create additional paths of influence between relative motions. It is a matter of what the modeler is willing to incorporate.

- Lastly, the strongest assumption exhibited by figure 4.2 is the absence of arrow between relative motions. It assumed that the relative motions from one pose to the next are only determined by the control signal, and that no effect of inertia remains from past motions (as stated in Box 8). For dynamical systems and/or under-actuated systems

scenarios, this assumption would not be realistic, so that either the value of the previous relative motion should be included $\mathbf{\Lambda}_{i-1}^{i}$ in the functional of $\mathbf{\Lambda}_{i}^{i+1}$ (i.e., an arrow should connect one relative motion to the next), or more involved family of Lie groups should be required. In the remainder of this work, we will carry on with the current assumption (no connection), as it is often the case in existing SLAM problems. The main point is that, in contrast to other representations, the causal graph makes explicit some of the judgements that were implicit before, which is viewed positively.

The previous scenario describes a simple trajectory described by a robot with known controls. However, sometimes, SLAM scenarios are only 'passive', i.e., only sensor measurements are known, and control inputs are ignored (or unknown). This variant is presented in Box 9.

Box 9: Simple trajectory scenario for a robot (odometry variant).

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot carries out 4 relative motions $\{\mathbf{\Lambda_0^1}, \mathbf{\Lambda_1^2}, \mathbf{\Lambda_2^3}, \mathbf{\Lambda_3^4}\}$. These motions are measured by an odometer: $\mathcal{Z} \triangleq \{\mathbf{z_0^1}, \mathbf{z_1^2}, \mathbf{z_2^3}, \mathbf{z_3^4}\}$, corrupted by noise. We assume absence of inertia in the robot motion, no false sightings, no outliers, etc.

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}\}$.

Let (M_2) denote the SCM of the toy scenario Box 9:

$$(M_2): \begin{cases} \forall i \in (1, \dots, N), & \mathbf{z}_{i-1}^{i} := \mathbf{\Lambda}_{i-1}^{i} \circ \mathrm{Exp}\left(\zeta_{i-1}^{i}\right), \\ \forall i \in (1, \dots, N), & \mathbf{x_i} := \mathbf{x_{i-1}} \circ \mathbf{\Lambda}_{i-1}^{i}, \end{cases} \tag{4.2}$$

where the error terms of the odometry ζ_{i-1}^{i} are expressed in the Euclidean space isomorphic to the Lie algebra of the manifold.

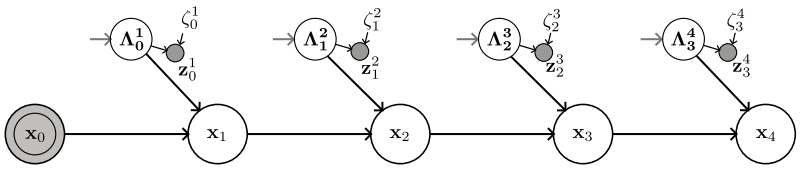


Figure 4.3: Representation of the toy scenario (Box 9) for when $N = 4$. The d-separation criterion shows that all measurement errors ζ_{i-1}^{i} are independent from one another.

This toy scenario (Box 9), in contrast to previous scenario (Box 8), completely ignores the causes of relative motions $\mathbf{\Lambda}$. Consequently the $\mathbf{\Lambda}$ nodes have no parents in the DAG $\mathcal{G}_{M_2}$ figure 4.3. In this scenario (Box 9), this is an unknown. In all rigour, all endogenous variables should have parents (according to Pearl et al. (2021, §1.5.1)), except those which are intervened upon like $\mathbf{x_0}$. Aside from that, 3 points out of the 4 exhibited for the previous scenario remain valid: the errors are independent, the drift is graphically exhibited and, explicit assumptions are made concerning the null effect of inertia.

Similarly, the deduced noise-free values for the pose $\mathbf{x}_i$ can be established by setting the error terms $\zeta^i_{i-1} = 0$:

$$\forall\, i \in 1 \dots N, \quad \mathbf{x_i} = \mathbf{x_0} \circ \prod_{j=1}^{i} \mathbf{\Lambda}^j_{j-1} = \mathbf{x_0} \circ \prod_{j=1}^{i} \mathbf{z}^j_{j-1},$$

which holds because $\mathbf{\Lambda}^i_{i-1} = \mathbf{z}^i_{i-1} \circ \mathrm{Exp}\left(-\zeta^i_{i-1}\right) = \mathbf{z}^i_{i-1}$ as $\mathrm{Exp}\left(\zeta^i_{i-1}\right)$ is identity.

Remark 4.3 (Data as intervention versus data as observation)**.** *We can see by comparing the two previous toy scenarios that there are two ways of specifying errors in relation to the data. By data, we gather sensor readings and control signals. Firstly, one can associate the error and the data as a common cause of the endogenous variable. This is the case for data corresponding to noisy control signals (scenario Box 8): $u \rightarrow \mathbf{\Lambda} \leftarrow v$. In this scenario, the error and the data are considered to be jointly responsible for the endogenous variable. Secondly, one can explain the data as the joint effect of the endogenous variable and the error. This is the case for odometry measurements (scenario Box 9): $\mathbf{\Lambda} \rightarrow \mathbf{z} \leftarrow \zeta$.*

At this point, it might be tempting to systematically categorize data as follows: actuator data in the first way and measurement data in the second way. However, we will see in the section about landmarks (§4.4) that this is not true.

4.2.2 Incorporation of loop-closures through summarized motions

The formulation of pose-graph loop-closures in the CARLIT framework is inspired by what was done in the CBN (see figure 3.13 & corresponding comments). However, the structure was not fully exploited as it was deemed impractical from the solver perspective. Importantly, relative motions in the used Lie groups can be composed, or summarized, as follows:

$$\forall j > i+1, \quad \mathbf{\Lambda}^j_i = \mathbf{\Lambda}^{i+1}_i \circ \mathbf{\Lambda}^{i+2}_{i+1} \circ \cdots \circ \mathbf{\Lambda}^j_{j-1}. \tag{4.3}$$

In causal terms, the summarized motion value $\mathbf{\Lambda}^j_i$ is assigned by all the relative motions between i and j, which are therefore the parents of $\mathbf{\Lambda}^j_i$. To include (4.3) in the SCM, the following functional defined by the assignment operator $:=$ is used (see definition 4.1):

$$\forall j > i+1, \quad \mathbf{\Lambda}^j_i := \prod_{k=i}^{j-1} \mathbf{\Lambda}^{k+1}_k. \tag{4.4}$$

Figure 4.4 shows such summarized motion in the node $\mathbf{\Lambda^4_1}$. Note that the summarized motion $\mathbf{\Lambda^4_1}$ is an open collider, thus no path of influence is added between its parents $\mathbf{\Lambda^2_1}, \mathbf{\Lambda^3_2}, \mathbf{\Lambda^4_3}$. The error terms ζ remain d-separated so that the previous analysis applies: no adjustment is possible. Conceivably, all combinations of summarized motions could be added as collider nodes in the graph. But as these nodes are purposeless in terms of belief propagation and to avoid cluttering the visual representation, they are not explicit drawn.

Summarized motions are interesting when they generate the loop-closure measurements data, i.e., when they are the cause of loop-closure measurement.

Causal Statement 4.2 (Loop-closure in pose-graphs)**.** *A loop-closure event occurs when the observation of a transform between two non-consecutive poses is available. The value of this observation is generated by (a) the summarized motion between the two poses, and (b) a measurement error.*

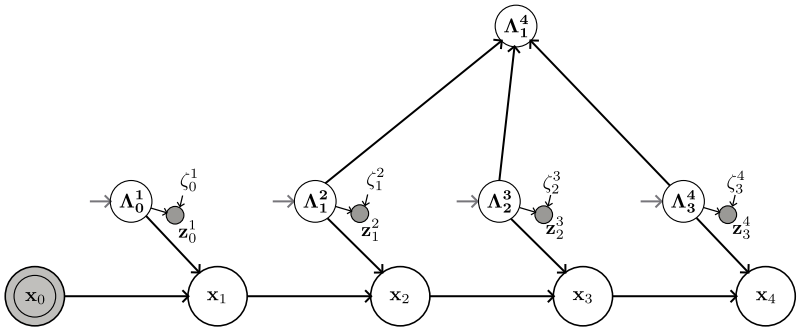


Figure 4.4: Representation of the toy scenario (Box 9) for $N = 4$, with the explicit representation of the summarized motion $\mathbf{\Lambda_1^4}$.

Expanding the previous scenario (Box 9), we consider the incorporation of a loop-closure between the poses $\mathbf{x_4}$ and $\mathbf{x_1}$, i.e, a relative transform measurement $\mathbf{z_1^4}$ between the two poses is available. The SCM (M_2'), which builds on (M_2) in (4.2), writes as:

$$(M_2') : (M_2) \bigcup \begin{cases} \mathbf{\Lambda_1^4} & := \mathbf{\Lambda_1^2} \circ \mathbf{\Lambda_2^3} \circ \mathbf{\Lambda_3^4}, \\ \mathbf{z_1^4} & := \mathbf{\Lambda_1^4} \circ \mathrm{Exp}\left(\zeta_1^4\right) \ . \end{cases} \tag{4.5}$$

Figure 4.5 shows the loop-closure measurement $\mathbf{z_1^4}$ as caused by the summarized motion and an error term.

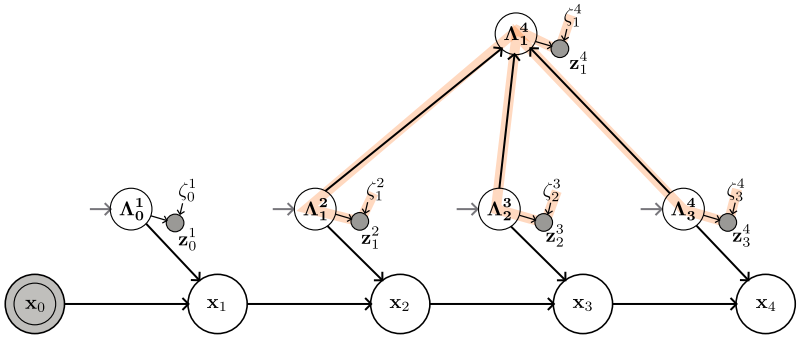


Figure 4.5: Extension of toy scenario (Box 9) for $N = 4$, see SCM (M_2') in (4.5). Compared to figure 4.3, this SCM contains the loop-closure measurement $\mathbf{x_4}$—$\mathbf{x_1}$ and error term ζ_1^4. The overlaid orange rubans highlight new paths of influence between the errors that are created thanks to the observed node $\mathbf{z}_1^4$ (a closed collider).

At this point, we should discuss the consequences of our causal assumptions, and how they differ from the core assumptions in probabilistic SLAM (see key assumption 2.2). We can see from the DAG on figure 4.5 that the loop-closure measurement is not a descendant of the involved poses $\mathbf{x_1}$ & $\mathbf{x_4}$. By contrast, for models which adhere to the probabilistic SLAM key assumption 2.2, whenever a geometric transform is observed between two poses, it entails a statistical relation between them. Here, a statistical relation between $\mathbf{x_1}$ and $\mathbf{x_4}$ would be introduced, bypassing over $\mathbf{x_2}$ and $\mathbf{x_3}$[5]. Figure 4.6 shows these probabilistic graphical model

[5]This is the case for BN, factor graphs, as well as the CBN model of chapter 3, through the mediation of

representations (BN is not shown, which is similar to CBN for this discussion).

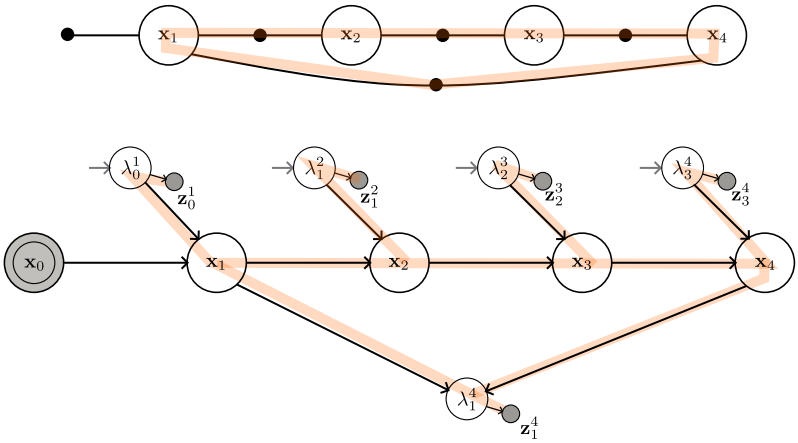


Figure 4.6: Factor graphs (top) and CBN (bottom), had the key assumptions of probabilistic been followed for the extension of toy scenario (Box 9) for when $N = 4$, with loop-closure and using probabilistic models.

From a probabilistic viewpoint, the following conditional independency can be read from graphs figure 4.6:

$$\mathbf{x_2} \perp\!\!\!\perp \mathbf{x_4} | \mathbf{x_1}, \mathbf{x_3},$$

which is not the case in the causal graph. Conversely, probabilistic graphical models do not depict well the following property seen on figure 4.5, i.e.,

$$\mathbf{x_1} \perp\!\!\!\perp \mathbf{z_1^4} | \mathcal{Z} ,$$

which is the fact that the loop-closure does not influence the value of $\mathbf{x_1}$ (and, carrying on, does not influence the first error ζ_0^1). Note that our modelling assumptions lead to the fact that we are compliant with the reasoning feature ƒ4 (see §2.3), which is the guarantee that the first variable in the loop is not influenced by the closure (without, so far, any additional assumption on the distribution of errors). Moreover, conditioning on $\mathbf{x_1}$ or intervening on $\mathbf{x_1}$ does not interfere with the inner statistical relations introduced by the loop-closure. In other words, it does not "break the loop" as it is the case for probabilistic representations. This is the issue $\mathfrak{u}_1$, in §2.3, noted in the probabilistic framework. In graphical terms, $\mathbf{x_1}$ is not in the orange path drawn on figure 4.5. Issue $\mathfrak{u}_1$, it is henceforth considered solved now.

On the important distinction between equality and assignment, remark that both following statements are true:

$$\mathbf{z}_1^4 \circ \mathrm{Exp}\left(-\zeta_1^4\right) = \mathbf{x_1}^{-1} \circ \mathbf{x_4} \quad \text{and} \quad \mathbf{z}_1^4 \circ \mathrm{Exp}\left(-\zeta_1^4\right) = \mathbf{\Lambda}_1^4 \, . \tag{4.6}$$

However, causal modelling require committing to a structural relation (with the := operator), which is chosen to be $\mathbf{z}_1^4 := \mathbf{\Lambda}_1^4 \circ \mathrm{Exp}\left(\zeta_1^4\right)$.

The rationale behind this conceptualization of the loop-closure as being explained by a sequence of relative motions, and not descendant of the poses, relies purely on an opinionated

an intermediary transform.

adherence to causal considerations. We consider that the poses composing the trajectory are artificial entities which build on an arbitrary starting point. In other words they do not exist per se, the experiment, they are only the (Cartesian) byproduct of the relative motions. Therefore, in our view they cannot be causes of a measurement owing to their fictional nature. On the other hand, relative motions $\{\mathbf{\Lambda}_{i-1}^{i}\}$ are considered as practical entities of the experiment and can explain, by composition of a number of them, the value of the loop-closure measurement.

Remark 4.4. *In practice, the loop-closure measurement is governed not by* $\mathbf{\Lambda_1^4}$*, but from its inverse. Indeed, typically the data is obtained from the robot current pose (there,* $\mathbf{x_4}$*) to the 'revisited' pose (in this case,* $\mathbf{x_1}$*) by front algorithms such as ICP. The real available data is hence denoted* $\mathbf{z}_4^1$ *with the measurement error captured as* $\zeta_4^{1\prime}$ *(of which we assume to know the noise characterics). Thus, the structural equation should be, in rigor:* $\mathbf{z_4^1} := (\mathbf{\Lambda}_1^4)^{-1} \circ Exp\left(\zeta_4^{1\prime}\right)$*. Nonetheless, we rely on the inversion property of Lie groups to state the structural equation as* $\mathbf{z_1^4} := \mathbf{\Lambda}_1^4 \circ Exp\left(\zeta_1^4\right)$*, without adverse effect on the causal interpretation developed so far. A pre-process step is implied to translate the pair* $(\mathbf{z}_4^1, \zeta_4^{1\prime})$ *into* $(\mathbf{z}_1^4, \zeta_1^4)$*, whereby* $\mathbf{z}_1^4 \triangleq (\mathbf{z}_4^1)^{-1}$*. For the relation between error terms, we use the adjoint relation (Solà et al., 2021, eq 30) to retrieve:* $\zeta_1^4 \triangleq -\mathbf{Ad}_{\mathbf{z}_4^1}\zeta_4^{1\prime}$*.*

We do this to conveniently standardize the notations and facilitate the derivation of various operations in the remainder of the chapter.

It is now established that the loop-closure introduces new paths of statistical influence between the errors $\{\zeta_1^2, \zeta_2^3, \zeta_3^4, \zeta_1^4\}$. By default, if one considers these errors to be zero a priori (assumption 4.1), a mismatch inevitably appears in that:

$$\mathbf{z_1^2} \circ \mathbf{z_2^3} \circ \mathbf{z_3^4} \neq \mathbf{z_1^4}. \tag{4.7}$$

The mismatch motivates the procedure of adjustment of errors, described next.

Note that several adjustment problems can be treated concurrently. Consider the extension of the previous scenario (Box 9) with another loop-closure, as represented in figure 4.7. In this case, as the loop-closures do not overlap with each other, the adjustments denoted in the blue boxes are independent of one another. For a similar type of problem, it was shown in the CBN chapter that the computation of message passing could be avoided in the anti-causal direction for the clique tree structure. However, conceptually, the CARLIT representation is arguably more appropriate for it translates the benefits of this scenario at the representational level (where independences can be exploited), instead of at the solver level (i.e., the clique tree structure).

Finally, we can come back on the issue u2 (the gauge freedom dilemma) uncovered in chapter 2, §2.3. It was noted that whenever we close the loop on the first variable $\mathbf{x_0}$, there is a dilemma whether $\mathbf{x_0}$ should be left as a fixed variable (at the risk of changing the statistics inside the loop) or whether a prior should be prepended on $\mathbf{x_0}$ (thus making it a decision variable, and drifting from the letter of the scenario). For convenience, we reproduce in figure 4.8 the factor graph of the scenario Box 7, with the corresponding CARLIT new representation.

We can see from figure 4.8-(c) that closing the loop on $\mathbf{x}_0$ requires no different considerations than closing the loop on another variable (e.g., the 2nd loop-closure in figure 4.7), so that the gauge freedom dilemma u2 is avoided in CARLIT.

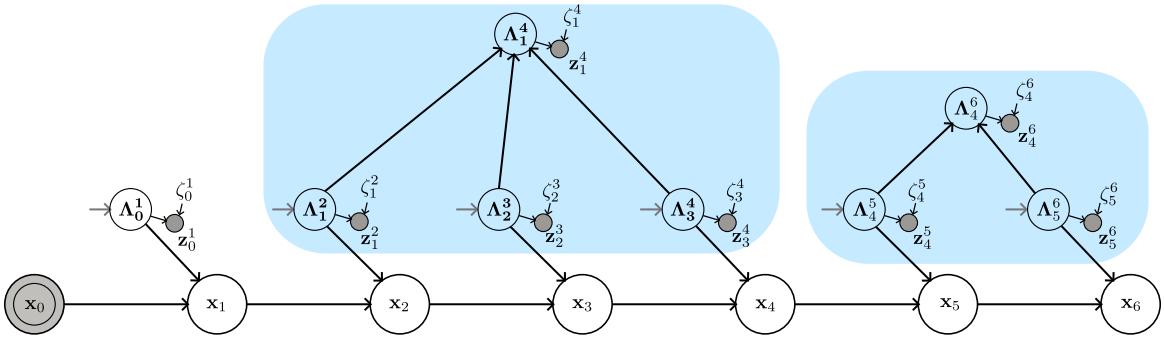


Figure 4.7: Further extension of toy scenario (Box 9) with another loop-closure (not overlapping with the first one). The two blue boxes depict the fact that two *independent* adjustment problems exist. As for the error ζ_0^1 (leftmost error), there is no way to correct it.

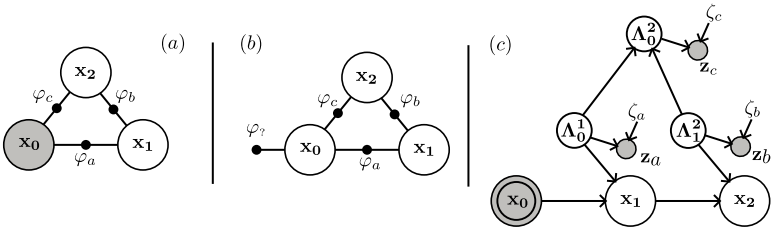


Figure 4.8: Reproduction of figure 2.18 with the addition in (c) of our causal approach.

4.3 Adjustment of errors

In the previous section, a causal representation of pose-graph SLAM was proposed. Important aspects of representation are yet to be introduced for landmark-based SLAM. This is deferred to the next section. In this section instead, we set forth to eliminate the geometric mismatch which emerges following a loop-closure. "Harmony" is restored (as coined by Delambre, see §1.4.1) by bringing correction to the error terms. Throughout the manuscript, this has been referred to as the "adjustment process".

Algorithm 2 describes the steps leading to the posterior estimate of the trajectory $\mathcal{X}$. The inference actually proceeds by temporarily removing the trajectory in order to conduct concurrent adjustment on K independent blocks. For instance, if the toy scenario figure 4.7 is considered, the function **SplitIndependentMotionBlocks** lays $K = 2$ adjustments: the two d-separated blue blocks. The most probable corrections to the errors are sought for the two subset of $\mathcal{E}$, namely, $\hat{\mathcal{E}}^{(1)} \triangleq \left\{\hat{\zeta}_1^2, \hat{\zeta}_2^3, \hat{\zeta}_3^4, \hat{\zeta}_1^4\right\}$ and $\hat{\mathcal{E}}^{(2)} \triangleq \left\{\hat{\zeta}_4^5, \hat{\zeta}_5^6, \hat{\zeta}_4^6\right\}$. These correspond to the function **AdjustMotionBlock** calls of algorithm 2. As for the error ζ_0^1 (i.e., the leftmost error in figure 4.7), the d-separation criterion guarantees that no adjustment can be made. Thus, $\hat{\zeta}_0^1 = 0$, i.e., the mode (and mean) of the prior pdf of ζ_0^1. The elements of the trajectory are deduced in the last step of algorithm 2 by using the deterministic functionals of the SCM given in (4.2). This leads to the relation:

Algorithm 2: CARLIT: Inference procedure in pose-graph problems.

Input: SCM & Causal graph $\mathcal{G}$ containing evidence $\mathcal{Z}$, intervention $\mathbf{x_0}$
Output: Adjusted errors $\hat{\mathcal{E}}$, trajectory $\hat{\mathcal{X}}$

$\mathcal{G}_{\setminus\mathcal{X}} \leftarrow$ `RemoveTrajectory(`$\mathcal{G}$`)` ;
$\left\{\mathcal{G}_{\mathbf{\Lambda}}^{(k)}\right\}_{k\in 1\dots K} \leftarrow$ `SplitIndependentMotionBlocks(`$G_{\setminus\mathcal{X}}$`)` ; // Split
for $k \in 1\dots K$; // Concurrently
do
  /* Adjust errors $\hat{\mathcal{E}}^{(k)}$ in k-th block */
  $\hat{\mathcal{E}}^{(k)} \leftarrow$ `AdjustMotionBlock(`$\mathcal{G}_{\mathbf{\Lambda}}^{(k)}$`)`;
end
$\hat{\mathcal{E}} \leftarrow \bigcup_{k\in 1\dots K}\left\{\hat{\mathcal{E}}^{(k)}\right\}$; // Join adjusted errors
$\hat{\mathcal{X}} \leftarrow$ `DeduceTrajectory(`$\mathcal{G}, \hat{\mathcal{E}}$`)` ; // Deduce trajectory
return $\left(\hat{\mathcal{E}}, \hat{\mathcal{X}}\right)$;

$$\forall i \in 1\dots N, \quad \mathbf{x}_i = \mathbf{x_0} \circ \prod_{j=1}^{i} \mathbf{z}_{j-1}^{j} \circ \mathrm{Exp}\left(-\zeta_{j-1}^{j}\right) \ . \tag{4.8}$$

In subsection 4.3.1, the adjustment method is presented for one 'block' and it is henceforth experimented.

Remark 4.5 (From MAP of the errors to MAP of the trajectory)**.**
The ***AdjustMotionBlock*** *calls of algorithm 2 aim at finding the maximum a posteriori (MAP) of the errors, i.e.,*

$$\hat{\mathcal{E}} = \underset{\mathcal{E}}{\mathrm{argmax}}\, p(\mathcal{E}|\mathcal{Z}) \ . \tag{4.9}$$

The details are described below, in §4.3.1. When it comes to deducing the MAP of the trajectory from $\hat{\mathcal{E}}$, the deterministic function (4.8) is nonlinear. This is slippery, because for a nonlinear function f linking a pose $\mathbf{x}_i$ to the errors, $\mathbf{x_i} = f(\mathcal{E})$, it cannot be said (without precautions) that the MAP of $\mathbf{x}_i$ is given by $\hat{\mathbf{x}}_i = f(\hat{\mathcal{E}})$.

Instead, we will limit ourselves in stating that it holds true that $\hat{\mathbf{x}}_i = f(\hat{\mathcal{E}})$ around a first order development of (4.8) around the point $\hat{\mathcal{E}}$.

4.3.1 Adjustment principles

The following SCM represents the relations between n relative motions and the summarized motion:

$$\text{SCM}: \begin{cases} \forall\ i \in (1,\dots,n), \quad \mathbf{z}_{i-1}^{i} := \mathbf{\Lambda}_{i-1}^{i} \circ \mathrm{Exp}\left(\zeta_{i-1}^{i}\right), \\ \mathbf{\Lambda}_0^n := \prod_{i=1}^{n} \mathbf{\Lambda}_{i-1}^{i}, \\ \mathbf{z}_0^n := \mathbf{\Lambda}_0^n \circ \mathrm{Exp}\left(\zeta_0^n\right). \end{cases} \tag{4.10}$$

We assume, as before, Gaussianity the error terms $\mathcal{E}$: $\forall\ i \in (1,\dots,n)$, $\zeta_{i-1}^{i} \sim \mathcal{N}(0\ ,\ \Sigma_{\mathrm{rm}})$ and $\zeta_0^n \sim \mathcal{N}(0\ ,\ \Sigma_{\mathrm{sm}})$, where subscripts rm and sm correspond to relative and summarized motions, respectively. Figure 4.9 shows the corresponding causal graph of SCM (4.10).

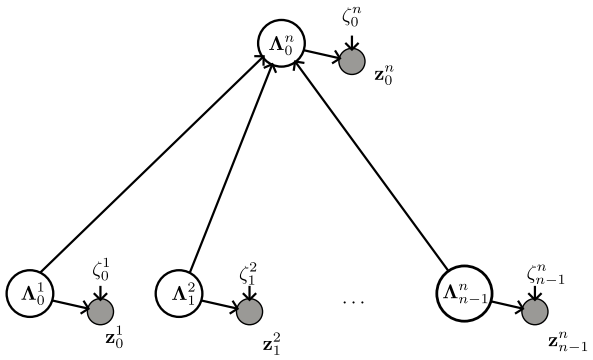


Figure 4.9: Loop-closure over $n+1$ relative motions.

A geometric constraint can be deduced from SCM (4.10):

$$\mathbf{z}_0^n \circ \mathrm{Exp}\left(-\zeta_0^n\right) = \prod_{i=1}^{n} \mathbf{z}_{i-1}^i \circ \mathrm{Exp}\left(-\zeta_{i-1}^i\right). \tag{4.11}$$

The above constraint is expressed under the form $C_s(\mathcal{E}) = d_s$, where:

$$\begin{cases} C_s(\mathcal{E}) \triangleq \mathrm{Log}\left(\left(\prod_{k=0}^{n-1} \mathrm{Exp}\left(\mathbf{Ad}^{-1}_{\prod_{j=n-k}^{j<n} \mathbf{z}_j^{j+1}} \zeta_{n-1-k}^{n-k}\right)\right)^{-1} \circ \mathrm{Exp}\left(\zeta_0^n\right)\right), \\ d_s \triangleq \mathrm{Log}\left(\left(\prod_{i=1}^{n} \mathbf{z}_{i-1}^i\right)^{-1} \circ \mathbf{z}_0^n\right). \end{cases} \tag{4.12}$$

This is achieved by grouping together the $\mathbf{z}$ terms and the error terms $\mathcal{E}$ in eq (4.11), so that the adjoint terms **Ad** show up. Then $\mathrm{Log}(\cdot)$ is applied on both sides[6].

Example 4.3.1. *For $n = 2$, we have: $\mathcal{E} = \{\zeta_0^1, \zeta_1^2, \zeta_0^2\}$, and $C_s(\mathcal{E}) = d_s$ with*

$$\begin{cases} C_s(\mathcal{E}) = Log\left(\left(Exp\left(\zeta_1^2\right) \circ Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2} \zeta_0^1\right)\right)^{-1} \circ Exp\left(\zeta_0^2\right)\right), \\ d_s = Log\left(\left(\mathbf{z}_0^1 \circ \mathbf{z}_1^2\right)^{-1} \circ \mathbf{z}_0^2\right). \end{cases} \tag{4.13}$$

Example 4.3.2. *For $n = 3$, we have: $\mathcal{E} = \{\zeta_0^1, \zeta_1^2, \zeta_2^3, \zeta_0^3\}$, and $C_s(\mathcal{E}) = d_s$ with*

$$\begin{cases} C_s(\mathcal{E}) = Log\left(\left(Exp\left(\zeta_2^3\right) \circ Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_2^3} \zeta_1^2\right) \circ Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2 \mathbf{z}_2^3} \zeta_0^1\right)\right)^{-1} \circ Exp\left(\zeta_0^3\right)\right), \\ d_s = Log\left(\left(\mathbf{z}_0^1 \circ \mathbf{z}_1^2 \circ \mathbf{z}_2^3\right)^{-1} \circ \mathbf{z}_0^3\right). \end{cases} \tag{4.14}$$

[6] This assumes that the mismatch is not large as, in general, it is not automatically true that $\mathrm{Log}(\mathrm{Exp}(\mathbf{x})) = \mathbf{x}$ for $\mathbf{x} \in \mathrm{SE}(\cdot)$. Additional consideration regarding this is given in §4.5.

Example 4.3.3. *For problems expressed in a vector space $\mathbb{R}^d$ (i.e., both $\mathbf{z}$ and ζ are in $\mathbb{R}^d$), the constraint (4.12) simplifies to the linear function:*

$$\begin{cases} C_s(\mathcal{E}) = \zeta_0^n - \sum_{i=1}^{n} \zeta_{i-1}^i \ , \\ d_s = \mathbf{z}_0^n - \sum_{i=1}^{n} \mathbf{z}_{i-1}^i \ . \end{cases} \tag{4.15}$$

Note that the problem in $\mathbb{R}^d$ is also linear in the probabilistic SLAM formulation, thus, it should not be surprising that the causal approach preserves this property.

The adjustment process consists in searching the most probable corrections to be brought to the errors $\mathcal{E}$ given the observations $\mathcal{Z}$, i.e., the MAP of $\mathcal{E}$, denoted $\hat{\mathcal{E}}$.

Lemma 4.3.1. *The following constrained least squares minimization problem enables to find out the MAP estimate of the errors $\hat{\mathcal{E}}$ given the observations $\mathcal{Z}$*

$$\hat{\mathcal{E}} \triangleq \underset{\mathcal{E}}{\operatorname{argmin}} \frac{1}{2} \left\| \rho^\top \mathcal{E} \right\|_2^2 \quad s.t. \quad C_s(\mathcal{E}) = d, \tag{4.16}$$

with

$$\rho\rho^\top \triangleq \begin{bmatrix} \Sigma_{rm} & & & \\ & \ddots & & \\ & & \Sigma_{rm} & \\ & & & \Sigma_{sm} \end{bmatrix}^{-1} ,$$

obtained by Cholesky decomposition of the information matrix of the errors.

Proof. The MAP of the errors writes as:

$$\hat{\mathcal{E}} \triangleq \underset{\mathcal{E}}{\operatorname{argmax}}\, p(\mathcal{E}|\mathcal{Z}) = \underset{\mathcal{E}}{\operatorname{argmax}}\, p(\mathcal{Z}|\mathcal{E}) p(\mathcal{E}) \tag{4.17}$$

The prior term $p(\mathcal{E})$ is known to be, according to assumption 4.1,

$$p(\mathcal{E}) = p(\zeta_0^n) \prod_{i=1}^{n} p(\zeta_{i-1}^i) = \mathcal{N}\left(\zeta_0^n; 0, \Sigma_{\text{sm}}\right) \prod_{i=1}^{n} \mathcal{N}\left(\zeta_{i-1}^i;\ 0, \Sigma_{\text{rm}}\right) \ . \tag{4.18}$$

The main insight of the proof consists in identifying that the likelihood function $p(\mathcal{Z}|\mathcal{E})$ is a Dirac formulated from the geometrical constraint (4.10):

$$p(\mathcal{Z}|\mathcal{E}) = \delta\left(\operatorname{Log}\left(\mathbf{\Lambda}_0^n\right) - \operatorname{Log}\left(\mathbf{\Lambda}_0^1 \circ \cdots \circ \mathbf{\Lambda}_{n-1}^n\right)\right) \ , \tag{4.19}$$

where each term $\mathbf{\Lambda}$ can be replaced by the form $\mathbf{z} \circ \operatorname{Exp}(-\zeta)$, as in (4.10).

Thus, the likelihood function can be expressed in the form $\delta\left(C_s(\mathcal{E}) - d_s\right)$. Consequently, the 'argmax' in (4.17) is only achieved under the explicit constraint $C_s(\hat{\mathcal{E}}) = d_s$:

$$\hat{\mathcal{E}} = \underset{\mathcal{E}}{\operatorname{argmax}}\, p(\mathcal{E}) \quad \text{s.t.} \quad C_s(\mathcal{E}) = d_s \ . \tag{4.20}$$

Furthermore, using (4.18), and developing the negative log of $p(\mathcal{E})$, yields:

$$\hat{\mathcal{E}} = \underset{\mathcal{E}}{\operatorname{argmin}} \frac{1}{2} \left({\zeta_0^n}^\top \Sigma_{\mathrm{sm}}^{-1} \zeta_0^n + \sum_{i=1}^{n} {\zeta_{i-1}^i}^\top \Sigma_{\mathrm{rm}}^{-1} \zeta_{i-1}^i \right) \quad \text{s.t.} \quad C_s(\mathcal{E}) = d_s \ . \tag{4.21}$$

And, lastly, the term $\frac{1}{2}(\dots)$ of (4.21) can be formatted in a matrix, leading to (4.16). □

Unfortunately, due to the expression of the constraint (4.11), the system cannot be solved directly, except for the simplest Lie groups such as vector spaces in (4.15). Therefore, (4.16) is solved iteratively in the local tangent space close to a changing linearization point $\tilde{\mathcal{E}}$. The choice of the initial linearization point is naturally $\tilde{\mathcal{E}} = 0$. Notice that, by contrast, the choice of an initialization point can be delicate in probabilistic SLAM. The constraint equation around $\tilde{\mathcal{E}}$ rewrites as:

$$C_s(\tilde{\mathcal{E}} + \delta\mathcal{E}) \approx C_s(\tilde{\mathcal{E}}) + C_{s,\tilde{\mathcal{E}}} \, \delta\mathcal{E} \ , \quad \text{with } C_{s,\tilde{\mathcal{E}}} \triangleq \left. \frac{\partial C_s}{\partial \mathcal{E}} \right|_{\tilde{\mathcal{E}}} \ . \tag{4.22}$$

Then, the linearized constrained least squares system is solved for $\delta\mathcal{E}$:

$$\widehat{\delta\mathcal{E}} \triangleq \underset{\delta\mathcal{E}}{\operatorname{argmin}} \frac{1}{2} \left\| A\delta\mathcal{E} - \tilde{b} \right\|_2^2 \quad \text{s.t.} \quad C_{s,\tilde{\mathcal{E}}} \, \delta\mathcal{E} = \tilde{d} \ , \tag{4.23}$$

for which,

$$A \triangleq \rho^\top, \qquad \tilde{b} \triangleq -\rho^\top \tilde{\mathcal{E}} \qquad \text{and} \quad \tilde{d} \triangleq d - C_s(\tilde{\mathcal{E}}) \ . \tag{4.24}$$

The Lagrangian $\mathcal{L}$ of the optimization problem (4.16) is used to reformulate it as a matrix system. Introducing λ as the vector of Lagrange coefficients[7] leads to

$$\mathcal{L}(\delta\mathcal{E}, \lambda) = \frac{1}{2} \left\| A\delta\mathcal{E} - \tilde{b} \right\|_2^2 + \lambda^\top \left(C_{s,\tilde{\mathcal{E}}} \, \delta\mathcal{E} - \tilde{d} \right) . \tag{4.25}$$

The saddle point $\left\{ \widehat{\delta\mathcal{E}}, \hat{\lambda} \right\}$ of (4.25) satisfies:

$$M \begin{bmatrix} \widehat{\delta\xi} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} A^\top \tilde{b} \\ \tilde{d} \end{bmatrix} \quad \text{with} \quad M \triangleq \begin{bmatrix} A^\top A & C_{s,\tilde{\mathcal{E}}}^\top \\ C_{s,\tilde{\mathcal{E}}} & 0 \end{bmatrix} \ . \tag{4.26}$$

The whole step of the adjustment is summarized in algorithm 3.

The cost for the adjustment is dominated by the inversion of the matrix (4.26). Since this matrix is sparse, symmetric (but not generally definite positive), an implementation of the Cholesky decomposition variant LDL$^\top$ might be most suited, as discussed later. The formulae for the computations of the Jacobians of $C_{s,\tilde{\mathcal{E}}}$ from (4.22), are given hereafter. We have used the derivative rules from Solà et al. (2021, eq 65,66,75,79,82,83):

$$\begin{cases} \left. \dfrac{\partial C_{s,\tilde{\mathcal{E}}}}{\partial \zeta_0^n} \right|_{\tilde{\mathcal{E}}} = \mathbf{J}_r^{-1} \left(C_{s,\tilde{\mathcal{E}}}(\tilde{\mathcal{E}}) \right) \mathbf{J}_r \left(\tilde{\zeta}_0^n \right) , \\ \forall \ i \in (1 \dots n), \\ \left. \dfrac{\partial C_{s,\tilde{\mathcal{E}}}}{\partial \zeta_{i-1}^i} \right|_{\tilde{\mathcal{E}}} = -\mathbf{J}_l^{-1} \left(C_{s,\tilde{\mathcal{E}}}(\tilde{\mathcal{E}}) \right) \mathbf{Ad}^{-1}_{\prod_{k=0}^{k<i-1} \mathrm{Exp}\left(\mathbf{Ad}^{-1}_{\prod_{j=i-k}^{j<n-1} \mathbf{z}_j^{j+1}} \tilde{\zeta}_{i-1-k}^{i-k} \right)} \\ \qquad\qquad \times \mathbf{J}_r \left(\mathbf{Ad}^{-1}_{\prod_{j=i}^{j<n-1} \mathbf{z}_j^{j+1}} \tilde{\zeta}_{i-1}^i \right) \mathbf{Ad}^{-1}_{\prod_{j=i}^{j<n-1} \mathbf{z}_j^{j+1}} , \end{cases} \tag{4.27}$$

[7] Not to be confused with the λ_i^j in the previous chapter.

Algorithm 3: Iterative Adjustment of a Motion closure.

Input: Normed matrix A, vector d & function of constraints $C_s(\mathcal{E})$
Output: Most probable adjustment for the errors $\hat{\mathcal{E}}$

```
/* Initialize the linearization point to null errors                    */
```
$\tilde{\mathcal{E}} \leftarrow 0$;
while `!StopCondition()` **do**
  `// Determine the necessary quantities from eq (4.24), (4.27)`
  $\tilde{d} \leftarrow d - C_s(\tilde{\mathcal{E}})$;
  $C_{s,\tilde{\mathcal{E}}} \leftarrow \frac{\partial C_s}{\partial \mathcal{E}}\Big|_{\tilde{\mathcal{E}}}$;
  $\tilde{b} \leftarrow -\rho^\top \tilde{\mathcal{E}}$;
  `// Apply eq (4.26)`
  $\left\{\widehat{\delta\mathcal{E}}, \hat{\lambda}\right\} \leftarrow$ `SaddlePoint`$(A, C_{s,\tilde{\mathcal{E}}}, \tilde{b}, \tilde{d})$;
  `// Update the linearization point of the errors`
  $\tilde{\mathcal{E}} \leftarrow \tilde{\mathcal{E}} + \widehat{\delta\mathcal{E}}$;
end
$\hat{\mathcal{E}} \leftarrow \tilde{\mathcal{E}}$;
return $\hat{\mathcal{E}}$

where notations follow Solà et al. (2021, eq 67,71,75), i.e., $\mathbf{J}_r(\zeta) \triangleq \frac{\partial \mathrm{Exp}(\zeta)}{\partial \zeta}\Big|_{\tilde{\zeta}}$ is the Jacobian "on-the-right" and $\mathbf{J}_l$ is the Jacobian"on-the-left". Both relate via $\mathbf{J}_r(\zeta) = \mathbf{J}_l(\zeta) \circ \mathbf{Ad}^{-1}_{\mathrm{Exp}(\zeta)}$.

Examples of implementation of the Jacobians blocks are given for $n = 2$ in (4.28) and $n = 3$ in (4.29).

Example 4.3.4. *For $n = 2$, we have:*

$$\begin{cases} \dfrac{\partial C_s}{\partial \zeta_0^2}\Big|_{\tilde{\mathcal{E}}} = \mathbf{J}_r^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{J}_r(\tilde{\zeta}_0^2), \\ \dfrac{\partial C_s}{\partial \zeta_0^1}\Big|_{\tilde{\mathcal{E}}} = -\mathbf{J}_l^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{J}_r\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2}\tilde{\zeta}_0^1\right)\mathbf{Ad}^{-1}_{\mathbf{z}_1^2}, \\ \dfrac{\partial C_s}{\partial \zeta_1^2}\Big|_{\tilde{\mathcal{E}}} = -\mathbf{J}_l^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{Ad}^{-1}_{Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2}\tilde{\zeta}_0^1\right)}\mathbf{J}_r\left(\tilde{\zeta}_1^2\right). \end{cases} \tag{4.28}$$

Example 4.3.5. *For $n = 3$, we have:*

$$\begin{cases} \left.\dfrac{\partial C_s}{\partial \zeta_0^3}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}_r^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{J}_r(\tilde{\zeta}_0^3), \\ \left.\dfrac{\partial C_s}{\partial \zeta_0^1}\right|_{\tilde{\mathcal{E}}} = -\mathbf{J}_l^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{J}_r\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2\mathbf{z}_2^3}\tilde{\zeta}_0^1\right)\mathbf{Ad}^{-1}_{\mathbf{z}_1^2\mathbf{z}_2^3}, \\ \left.\dfrac{\partial C_s}{\partial \zeta_1^2}\right|_{\tilde{\mathcal{E}}} = -\mathbf{J}_l^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{Ad}^{-1}_{Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2\mathbf{z}_2^3}\tilde{\zeta}_0^1\right)}\mathbf{J}_r\left(\mathbf{Ad}^{-1}_{\mathbf{z}_2^3}\tilde{\zeta}_2^3\right)\mathbf{Ad}^{-1}_{\mathbf{z}_2^3} \\ \left.\dfrac{\partial C_s}{\partial \zeta_2^3}\right|_{\tilde{\mathcal{E}}} = -\mathbf{J}_l^{-1}\left(C_s(\tilde{\mathcal{E}})\right)\mathbf{Ad}^{-1}_{Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_2^3}\tilde{\zeta}_1^2\right)Exp\left(\mathbf{Ad}^{-1}_{\mathbf{z}_1^2\mathbf{z}_2^3}\tilde{\zeta}_0^1\right)}\mathbf{J}_r\left(\tilde{\zeta}_2^3\right). \end{cases} \tag{4.29}$$

Since the pose-graph loop closures do not necessarily start at index 0, a more general formula is given for summarized motion $\mathbf{\Lambda}_s^{s+n}$ that starts at index s:

$$\begin{cases} \left.\dfrac{\partial C_{s,\tilde{\mathcal{E}}}}{\partial \zeta_s^{s+n}}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}_r^{-1}\left(C_{s,\tilde{\mathcal{E}}}(\tilde{\mathcal{E}})\right)\mathbf{J}_r\left(\tilde{\zeta}_s^{s+n}\right), \\ \forall\, i \in (1\ldots n),\ \left.\dfrac{\partial C_{s,\tilde{\mathcal{E}}}}{\partial \zeta_{s+i-1}^{s+i}}\right|_{\tilde{\mathcal{E}}} = \\ \quad -\mathbf{J}_l^{-1}\left(C_{s,\tilde{\mathcal{E}}}(\tilde{\mathcal{E}})\right)\mathbf{Ad}^{-1}_{\prod_{k=0}^{k<i-1}\mathrm{Exp}\left(\mathbf{Ad}^{-1}_{\prod_{j=i-k}^{j<n-1}\mathbf{z}_{s+j}^{s+j+1}}\tilde{\zeta}_{s+i-1-k}^{s+i-k}\right)} \\ \qquad \times\mathbf{J}_r\left(\mathbf{Ad}^{-1}_{\prod_{j=i}^{j<n-1}\mathbf{z}_{s+j}^{s+j+1}}\tilde{\zeta}_{i-1}^{i}\right)\mathbf{Ad}^{-1}_{\prod_{j=i}^{j<n-1}\mathbf{z}_{s+j}^{s+j+1}}, \end{cases} \tag{4.30}$$

Remark 4.6 (Implementation tricks)**.** *In the first iteration, since the initialization point is $\tilde{\mathcal{E}} = 0$, only the rightmost adjoint term in (4.30) remains, as all other terms simplify to the matrix identity. Therefore, the computations of most terms can be avoided if the relevant measurement values $\mathbf{z}$ are updated between iterations instead, $\mathbf{z}' \leftarrow \mathbf{z} \circ Exp\left(-\hat{\zeta}\right)$. Using that trick, every iteration can proceed at $\tilde{\mathcal{E}} = 0$. However, if such strategy is followed, then the expression of d_s must be updated with the new values $\mathbf{z}'$.*

Moreover, for $n = 2$, as instantiated in example (4.28), only one of the three Jacobians block has to be updated under that strategy, i.e., $\mathbf{Ad}^{-1}_{\mathbf{z}_1^2}$. It is useful to keep that in mind because adjustments having $n > 2$ can be, by compositions, transformed into connected adjustments each sized $n = 2$, as will be discussed in section 4.5.

4.4 Landmarks and Internalized Topography

This section addresses questions related to landmarks: how to represent their discovery, their re-observation (loop-closure), and when and how the adjustment process is triggered.

To support our approach to landmark incorporation, the toy scenario box 10 is presented, which is closely similar to the toy scenario box 3 introduced in chapter 2. Figure 4.10 provides a sketch as well as the factor graph representation for comparison with causal graph. The landmarks are denoted by $l_{1,\mathrm{I}}$. The Arabic subscript $1, 2, \ldots$ of a landmark stands for the time (or robot pose index) when the landmark is discovered. The roman letter $_{\mathrm{I,II,III,IV,\ldots}}$ terms the

landmark index at this pose (conceivably, several landmarks $l_{1,\mathrm{I}}, l_{1,\mathrm{II}}, \dots$ can be discovered from the same robot pose). The relative position from the robot pose to the landmark $l_{1,I}$ is noted $y^k_{1,\mathrm{I}}$, with k terms the time elapsed since discovery, i.e., $y^k_{1,\mathrm{I}}$ is the relative position from $\mathbf{x}_{1+k}$ to $l_{1,\mathrm{I}}$. A measurement is noted with a 'bar', e.g. $\bar{y}^k_{1,\mathrm{I}}$, and the noise (error term) is symbolized $\nu^k_{1,\mathrm{I}}$. It is considered that relative positions y are expressed in the action group of the Lie group of $\mathbf{x}$ and $\mathbf{\Lambda}$. Landmarks belonging in the Lie group (oriented landmarks) are also considered in what follows.

Box 10: SLAM Scenario 1 (variant of box 3)

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot operates 5 relative motions. A first landmark, denoted $l_{1,\mathrm{I}}$ is discovered (no prior knowledge) by a robot exteroceptive sensor at time t_1 and re-observed at t_4. A second landmark $l_{2,\mathrm{I}}$ is discovered at t_2 and re-observed at t_3. Robot motion measurements and landmark relative measurements $\mathcal{Z} \triangleq \left\{\mathbf{z}^1_0, \mathbf{z}^2_1, \mathbf{z}^3_2, \mathbf{z}^4_3, \mathbf{z}^5_4, \bar{y}^0_{1,\mathrm{I}}, \bar{y}^3_{1,\mathrm{I}}, \bar{y}^0_{2,\mathrm{I}}, \bar{y}^1_{2,\mathrm{I}}\right\}$ are affected by noise, denoted $\mathcal{E} = \left\{\zeta^1_0, \zeta^2_1, \zeta^3_2, \zeta^4_3, \zeta^5_4, \nu^0_{1,\mathrm{I}}, \nu^3_{1,\mathrm{I}}, \nu^0_{2,\mathrm{I}}, \nu^1_{2,\mathrm{I}}\right\}$. We assume no adverse effect due to the robot inertia, false sightings, outliers, etc.

Objective: The inference problem consists in establishing the joint estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}\}$ and the landmarks positions $\mathcal{L} \triangleq \{l_{1,\mathrm{I}}, l_{2,\mathrm{I}}\}$.

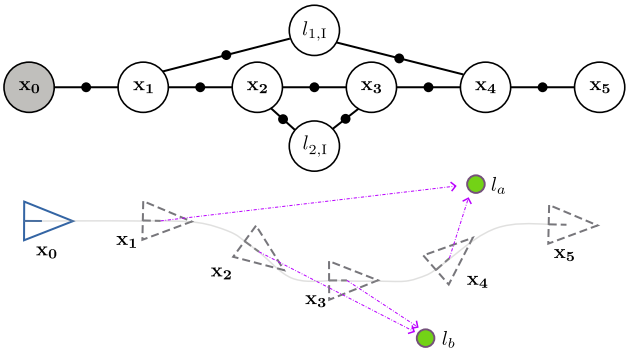


Figure 4.10: Factor graph (top) and sketch (bottom) of a toy scenario as per box 10 description.

It was noted, under the reasoning feature f3 in §2.3, that it would be relevant that the representation upholds the fact that the discovery of new landmarks does not induce adjustment (which is featured neither by BN nor by factor graphs). To express the reasoning feature f3 in terms of desirable (in)dependences between errors, we want the error associated with the landmark first observation to be independent of all other errors. However, when the landmark is re-observed, i.e., when a loop is closed, then the errors should become dependent conditioned on the first and last observations. From those two restrictions comes the difficulty of coming with a representation.

Hereafter, the landmark discovery process is enunciated first (§4.4.1), and then the closure associated with the re-observation is explored (§4.4.2).

4.4.1 Landmark discovery

Let us consider the scenario at the time when the robot pose is $\mathbf{x}_2$. The landmark $l_{1,\mathrm{I}}$ and $l_{2,\mathrm{I}}$ have been discovered and, by scenario box 10, no prior knowledge of their existence was available. Once more, we introduce the representation via a causal statement.

Causal Statement 4.3 (Landmark discovery)**.** *At discovery time, The relative position of a landmark, is dictated by the value read from the measurement (e.g., range-bearing), and affected by the error that comes with that measurement. The landmark absolute position is determined by (a) the robot pose and (b) the relative position of the landmark, both at discovery time.*

The causal statement 4.3 is implemented by the last 4 lines of this SCM:

$$(\mathrm{SCM}): \begin{cases} \mathbf{x_1} := \mathbf{x_0} \circ \mathbf{\Lambda}_0^1 \\ \mathbf{x_2} := \mathbf{x_1} \circ \mathbf{\Lambda}_1^2 \\ \mathbf{z}_0^1 := \mathbf{\Lambda}_0^1 \circ \mathrm{Exp}\left(\zeta_0^1\right) \\ \mathbf{z}_1^2 := \mathbf{\Lambda}_1^2 \circ \mathrm{Exp}\left(\zeta_1^2\right) \\ y_{1,\mathrm{I}}^0 := \bar{y}_{1,\mathrm{I}}^0 + \nu_{1,\mathrm{I}}^0 \\ y_{2,\mathrm{I}}^0 := \bar{y}_{2,\mathrm{I}}^0 + \nu_{2,\mathrm{I}}^0 \\ l_{1,\mathrm{I}} := \mathbf{x_1} \cdot y_{1,\mathrm{I}}^0 \\ l_{2,\mathrm{I}} := \mathbf{x_2} \cdot y_{2,\mathrm{I}}^0 \end{cases}, \tag{4.31}$$

where the operator $\cdot$, as in $\mathbf{x} \cdot y$, denotes the action of $\mathbf{x}$ on y. Typically, if $\mathbf{x} \in \mathrm{SE}(3)$, then $y \in \mathbb{R}^3$.

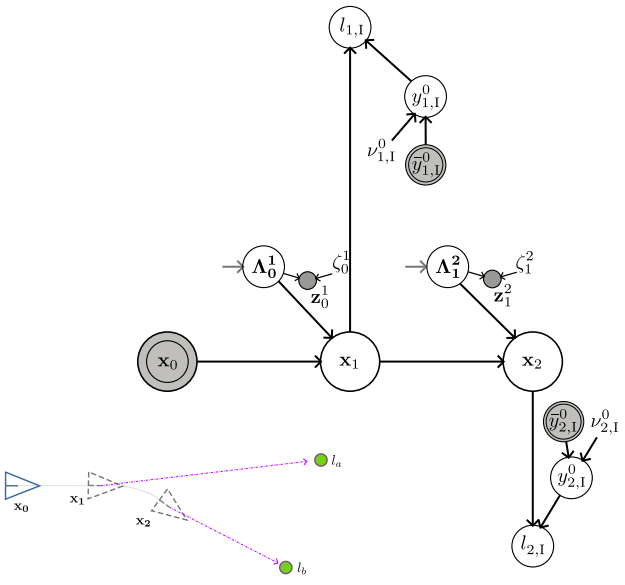


Figure 4.11: Causal graph of SCM (4.31). The error terms ν are d-separated from the remainder errors.

The causal graph on figure 4.11 depicts the d-separation between each landmark observation error term ($\nu_{1,\mathrm{I}}^0$ and $\nu_{2,\mathrm{I}}^0$) and all other errors, thanks to the open colliders $l_{1,\mathrm{I}}$ and $l_{2,\mathrm{I}}$.

Thus, this scheme conforms with reasoning feature ƒ3: there is at this stage no way to correct those errors as they are mutually independent.

One seemingly curious aspect of the representation figure 4.11 is the fact that the observations $\bar{y}^0_{1,\mathrm{I}}$ and $\bar{y}^0_{2,\mathrm{I}}$ are root nodes, i.e., interventions (hence the double-circled nodes). Furthermore, the measurement value is not *caused* by another variable through some physical process: it is instead modeled an arbitrary causal intervention, by a force/stimulus external to the scope of the model. This marks a conceptual difference with the approaches of BN and factor graphs that process landmark measurements as likelihood functions, caused by the robot pose and the landmark position. The causality implied by those schemes is {pose → measurement ← landmark}, e.g., see Stachniss et al. (2016, fig 46.1) and figure 2.4.

This aspect will be justified in further remarks in the next subsection.

4.4.2 Landmark re-observation

The full steps of the toy scenario box 10, when both landmarks $l_{1,\mathrm{I}}$ and $l_{2,\mathrm{I}}$ are re-observed, is addressed through the following causal statement.

Causal Statement 4.4 (Landmark re-observation)**.** *The relative position of a landmark at re-observation time is determined by (a) the relative position of the landmark at discovery time, and (b) the robot interval motion between discovery and re-observation.*
The value read from the re-observation measurement is dictated by the relative position of the landmark at re-observation time. It is affected by an error from said measurement.

The causal statement above is implemented by supplementing new functionals to the SCM (4.31):

$$(\mathrm{SCM}) : (4.31) \bigcup \begin{cases} y^1_{2,\mathrm{I}} := (\mathbf{\Lambda}^3_2)^{-1} \cdot y^0_{2,\mathrm{I}}, \\ \mathbf{\Lambda}^4_1 := \mathbf{\Lambda}^2_1 \circ \mathbf{\Lambda}^3_2 \circ \mathbf{\Lambda}^4_3, \\ y^3_{1,\mathrm{I}} := (\mathbf{\Lambda}^4_1)^{-1} \cdot y^0_{1,\mathrm{I}}, \\ \bar{y}^1_{2,\mathrm{I}} := y^1_{2,\mathrm{I}} + \nu^1_{2,\mathrm{I}}, \\ \bar{y}^3_{1,\mathrm{I}} := y^3_{1,\mathrm{I}} + \nu^3_{1,\mathrm{I}}, \end{cases} \tag{4.32}$$

Figure 4.12 shows the causal graph corresponding to SCM (4.32). It also portrays the path of influence between errors terms $\mathcal{E}^{(1)} \triangleq \left\{\zeta^2_1, \zeta^3_2, \zeta^4_3, \nu^0_{1,\mathrm{I}}, \nu^0_{2,\mathrm{I}}, \nu^3_{1,\mathrm{I}}, \nu^1_{2,\mathrm{I}}\right\}$ (orange ruban), following the incorporation of the observations $\bar{y}^1_{2,\mathrm{I}}$ and $\bar{y}^3_{1,\mathrm{I}}$ (closed colliders).

The reasoning is similar to the one in pose graphs: mismatches come into play between the deduced values of $y^3_{1,\mathrm{I}}$, $y^1_{2,\mathrm{I}}$ and their measured values $\bar{y}^3_{1,\mathrm{I}}$, $\bar{y}^1_{2,\mathrm{I}}$. This compels an adjustment over the set $\mathcal{E}^{(1)}$ to restore harmony.

Remark 4.7. *The landmarks nodes $l_{1,\mathrm{I}}$ and $l_{2,\mathrm{I}}$ are open colliders in the graph. Although the inference objective is to estimate their values, they play no role in the adjustment process. In that regard, a similar point is made for the set of landmarks $\mathcal{L}$ as was made for set of poses $\mathcal{X}$ in the previous section: the landmarks are Cartesian constructs, and we deduce their estimate after adjustments occurring in the error space.*

This representation preserves the intuitive notion that the position of a landmark gets corrected after every re-observation of it. Indeed, the adjustment process will bring correction to the discovery error $\nu^0_{1,\mathrm{I}}$ (resp. $\nu^0_{2,\mathrm{I}}$), which will then affect $l_{1,\mathrm{I}}$ (resp. $l_{2,\mathrm{I}}$) downstream. Moreover, echoing our discussion for representation of loop-closures in pose-graph SLAM

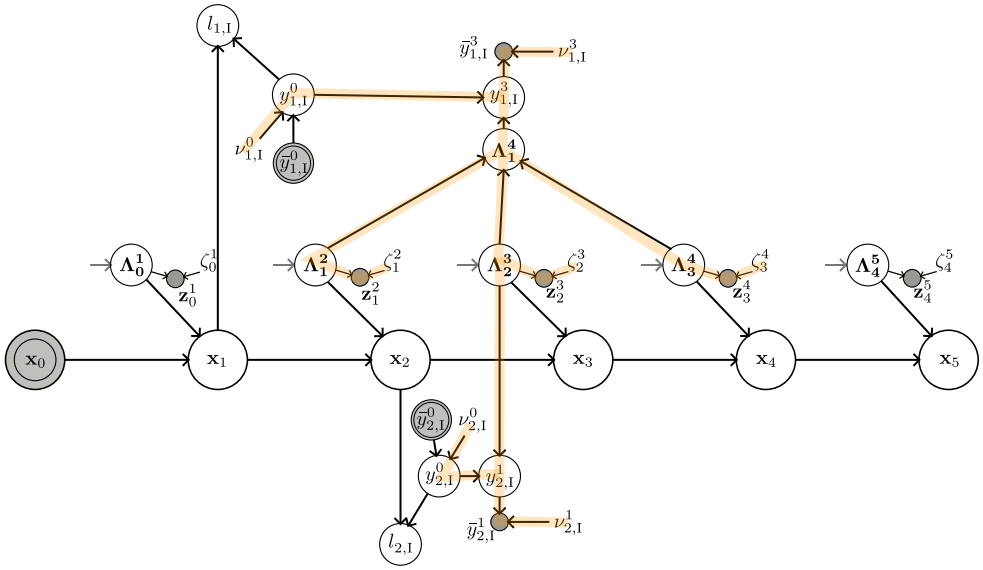


Figure 4.12: Causal graph of SCM (4.32) which represents, in comparison to figure 4.11, the re-observation of errors. Orange path depicts the path of influence between the exogenous variables.

§4.2.2, we can see that $\mathbf{x_1}$ does not 'break the loop' in the DAG figure 4.12 like it does in the corresponding factor graphs representation figure 4.10. Indeed, it graphically verified that, in figure 4.12, $(\mathbf{x_1} \perp\!\!\!\perp \nu^3_{1,\mathrm{I}} | \mathcal{Z}, \mathrm{do}(\mathcal{A}))$[8]. Therefore, in this causal representation, reasoning feature ʃ4 is maintained (i.e., no influence on the first variable in a loop), and issue ʉ1 (loop-breaking) is avoided.

World model or Internalized Topography? At this point, a fundamental question must be raised about *what* is being modelled. In the 'real world', there is no doubt: the physical process that generates the sensor measurement is similar whether the landmark is discovered or re-observed. Thus, a candidate causal model, which aims at describing a world model (i.e., a model of how things physically work), would not be able to differentiate between the discovery and the re-observation, and would then be unable to exhibit a satisfying pattern of conditional independences when building the map. However, in CARLIT, the SCM (4.32) depicts the sensor measurement either as a set root cause for the hidden landmark relative position (i.e., $\mathrm{do}(\bar{y}^0) \rightarrow y^0 \leftarrow \nu^0$ when discovering), or as an observed effect of the hidden landmark relative position (i.e., $k > 0,\ y^k \rightarrow \bar{y}^k \leftarrow \nu^k$ when re-observing).

Although it has often been assumed that mobile robots must build world model through mapping (Chatila and Laumond, 1985), the causal approach in this chapter clearly takes a different approach by being centered around the robot/device viewpoint. What is being modelled here is the internal process of topography: the landmark, as a *thing in itself*, does not exist as an entity generating data. In this scheme, any explicit requirement to build a world model or a rendering structure (such as the set of landmarks and poses) would rather be achieved by appending open colliders, which do not influence the adjustment process.

Remark 4.8 (Magnet analogy)**.** *To offer an analogy for the process of internalized topography, one can reproduce the cartographic process with magnets on a board. The board is*

[8] Or, more bluntly put, the orange rubans in figure 4.12 cannot be extended to $\mathbf{x_1}$.

initially blank. When a new entity is discovered, an action is taken to place the magnet on the board, this is the intervention. And whenever an entity is re-observed, the magnet position on the board is re-adjusted thanks to the new observation.

Adjustment of errors (with landmarks)

The constraint established on by the re-observation of $l_{2,\mathrm{I}}$ formulates as:

$$\bar{y}^1_{2,\mathrm{I}} = \left[\mathrm{Exp}\left(\zeta^3_2\right) \circ (\mathbf{z}^3_2)^{-1}\right] \cdot \left[\bar{y}^0_{2,\mathrm{I}} + \nu^0_{2,\mathrm{I}}\right] + \nu^1_{2,\mathrm{I}} \ , \tag{4.33}$$

from which the forms C_{ro} and d_{ro} are defined:

$$\begin{cases} C_{ro}(\mathcal{E}) \triangleq \left[\mathrm{Exp}\left(\zeta^3_2\right) \circ (\mathbf{z}^3_2)^{-1}\right] \cdot \left[\bar{y}^0_{2,\mathrm{I}} + \nu^0_{2,\mathrm{I}}\right] + \nu^1_{2,\mathrm{I}}, \\ d_{ro} = \bar{y}^1_{2,\mathrm{I}} \ . \end{cases} \tag{4.34}$$

To leverage a variant of algorithm 3, the expression of the Jacobian blocks of C_{ro} relative to the errors must be determined, at linearization point $\tilde{\mathcal{E}}$:

$$\begin{cases} \left.\dfrac{\partial C_{ro}}{\partial \nu^1_{2,\mathrm{I}}}\right|_{\tilde{\mathcal{E}}} = I \ , \\ \left.\dfrac{\partial C_{ro}}{\partial \nu^0_{2,\mathrm{I}}}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}^{\mathbf{M}\cdot\mathrm{p}}_{\mathrm{p}}\left(\tilde{\mathcal{E}}\right) \ , \\ \left.\dfrac{\partial C_{ro}}{\partial \zeta^3_2}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}^{\mathbf{M}\cdot\mathrm{p}}_{\mathbf{M}}\left(\tilde{\mathcal{E}}\right) \mathbf{Ad}_{\mathbf{z}^3_2} \mathbf{J}_r\left(\tilde{\zeta}^3_2\right) \ , \end{cases} \tag{4.35}$$

with $\mathbf{J}^{\mathbf{M}\cdot\mathrm{p}}_{\mathrm{p}}()$ and $\mathbf{J}^{\mathbf{M}\cdot\mathrm{p}}_{\mathbf{M}}()$ being defined as the Jacobian blocks of the group action (operator $\cdot$), which expression depends on the manifold considered in the scenario. For the commonly used SE(2) and SE(3) manifolds, see Solà et al. (2021, eq 166,167,182,183).

The more general formulae are given below, for a landmark $l_{s,\mathrm{I}}$, i.e. discovered at time s, and re-observed after n steps. First, the constraint is formulated as:

$$\begin{cases} C_{ro}(\mathcal{E}) \triangleq \left(\displaystyle\prod_{i=1}^{n} \mathbf{z}^{s+i}_{s+i-1} \circ \mathrm{Exp}\left(-\zeta^{s+i}_{s+i-1}\right)\right)^{-1} \cdot \left[\bar{y}^0_{s,\mathrm{I}} + \nu^0_{s,\mathrm{I}}\right] + \nu^n_{s,\mathrm{I}} \ , \\ d_{ro} \triangleq \bar{y}^n_{s,\mathrm{I}} \ . \end{cases} \tag{4.36}$$

Then, the Jacobian blocks write as:

$$\begin{cases} \left.\dfrac{\partial C_{ro}}{\partial \nu^n_{s,\mathrm{I}}}\right|_{\tilde{\mathcal{E}}} = I \ , \\ \left.\dfrac{\partial C_{ro}}{\partial \nu^0_{2,\mathrm{I}}}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}^{\mathbf{M}\cdot\mathrm{p}}_{\mathrm{p}}\left(\tilde{\mathcal{E}}\right) \ , \\ \forall\ i \in (1,\ldots,n), \ \left.\dfrac{\partial C_{ro}}{\partial \zeta^i_{i-1}}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}^{\mathbf{M}\cdot\mathrm{p}}_{\mathbf{M}}\left(\tilde{\mathcal{E}}\right) \mathbf{Ad}_{\prod_{j=1}^{n} \mathbf{z}^{s+j}_{s+j-1}} \left.\dfrac{\partial h}{\partial \zeta^i_{i-1}}\right|_{\tilde{\mathcal{E}}} \ , \end{cases} \tag{4.37}$$

where the quantity $\left.\frac{\partial h}{\partial \zeta^i_{i-1}}\right|_{\tilde{\mathcal{E}}}$, is defined hereafter, for space reasons:

$$\begin{cases} h(\tilde{\mathcal{E}}) \triangleq \prod_{i=1}^{n} \mathrm{Exp}\left(-\mathbf{Ad}^{-1}_{\prod_{j=i+1}^{n} \mathbf{z}^{s+j}_{s+j-1}} \tilde{\zeta}^{i}_{i-1}\right) , \\ \left.\dfrac{\partial h}{\partial \zeta^{i}_{i-1}}\right|_{\tilde{\mathcal{E}}} = \mathbf{Ad}_{\prod_{j=1}^{i-1} \mathrm{Exp}\left(-\mathbf{Ad}^{-1}_{\prod_{k=j+1}^{n} \mathbf{z}^{s+k}_{s+k-1}} \tilde{\zeta}^{s+j}_{s+j-1}\right)} \mathbf{J}_r\left(\mathbf{Ad}^{-1}_{\prod_{k=i+1}^{n} \mathbf{z}^{s+k}_{s+k-1}} \tilde{\zeta}^{s+i}_{s+i-1}\right) \mathbf{Ad}^{-1}_{\prod_{k=i+1}^{n} \mathbf{z}^{s+k}_{s+k-1}} , \end{cases} \tag{4.38}$$

Example 4.4.1. *For problems expressed in a vector space $\mathbb{R}^d$, (4.36) simplifies to the linear function:*

$$C_{ro}(\tilde{\mathcal{E}}) = \bar{y}^{0}_{s,\mathrm{I}} + \nu^{0}_{s,\mathrm{I}} + \nu^{n}_{s,\mathrm{I}} - \left(\sum_{i=1}^{n} \mathbf{z}^{s+i}_{s+i-1} - \zeta^{s+i}_{s+i-1}\right) . \tag{4.39}$$

Figure 4.13 depicts the pattern of the sparse matrix M involved in (4.26), for the adjustment of SCM (4.32). The lower-left part of M are the Jacobian blocks given by (4.37): the first and second rows correspond respectively to the re-observation of landmark $l_{1,\mathrm{I}}$ and $l_{2,\mathrm{I}}$. The upper-right part of M is the transpose of the lower-left.

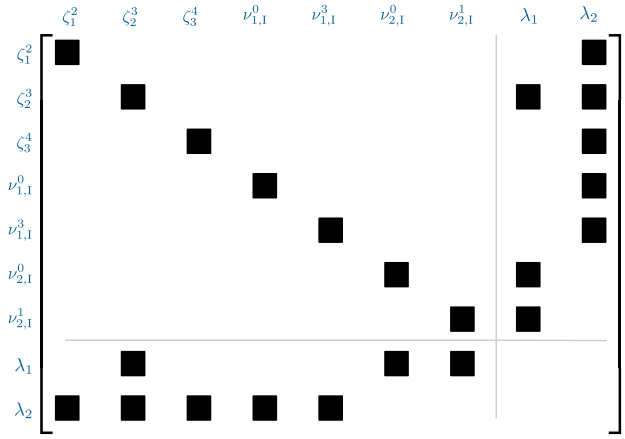


Figure 4.13: Pattern of the square matrix in the Langragian expression (4.26), with the last 2 lines and last 2 columns contain the Jacobian blocks of the constraints related to the re-observation of landmarks $l_{2,\mathrm{I}}$ and $l_{1,\mathrm{I}}$ (resp. first and second line in the constraint block).

When landmarks are specified in the manifold

For landmarks which are expressed in the manifold instead of the action group, we denote by $\mathbf{y}$ the relative position (noted before as boldless y). The constraint and Jacobian expressions must be adapted for this variant. The constraint following re-observation, noted $C_{ro,m}$ is:

$$\begin{cases} C_{ro,m}(\mathcal{E}) \triangleq \left(\prod_{i=1}^{n} \mathbf{z}^{s+i}_{s+i-1} \circ \mathrm{Exp}\left(-\zeta^{s+i}_{s+i-1}\right)\right)^{-1} \circ \bar{\mathbf{y}}^{0}_{s,\mathrm{I}} \circ \mathrm{Exp}\left(\nu^{0}_{s,\mathrm{I}}\right) \circ \mathrm{Exp}\left(\nu^{n}_{s,\mathrm{I}}\right) , \\ d_{ro,m} \triangleq \bar{\mathbf{y}}^{n}_{s,\mathrm{I}} . \end{cases} \tag{4.40}$$

And the Jacobian blocks are consequently:

$$\begin{cases} \left.\dfrac{\partial C_{ro,m}}{\partial \nu^n_{s,\mathrm{I}}}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}_r^{-1}\left(C_{ro,m}(\tilde{\mathcal{E}})\right)\mathbf{J}_r(\nu^n_{s,\mathrm{I}}) \ , \\ \left.\dfrac{\partial C_{ro,m}}{\partial \nu^0_{s,\mathrm{I}}}\right|_{\tilde{\mathcal{E}}} = \mathbf{J}_r^{-1}\left(C_{ro,m}(\tilde{\mathcal{E}})\right)\mathbf{Ad}^{-1}_{\mathrm{Exp}\left(\tilde{\nu}^n_{s,\mathrm{I}}\right)}\mathbf{J}_r(\nu^0_{s,\mathrm{I}}) \ , \\ \forall\ i \in (1,\dots,n),\ \left.\dfrac{\partial C_{ro,m}}{\partial \zeta^i_{i-1}}\right|_{\tilde{\mathcal{E}}} = \mathbf{Ad}^{-1}_{\bar{\mathbf{y}}^0_{s,\mathrm{I}}\circ\mathrm{Exp}\left(\tilde{\nu}^0_{s,\mathrm{I}}\right)\circ\mathrm{Exp}\left(\tilde{\nu}^n_{s,\mathrm{I}}\right)}\mathbf{Ad}_{\prod_{j=1}^n \mathbf{z}^{s+j}_{s+j-1}}\left.\dfrac{\partial h}{\partial \zeta^i_{i-1}}\right|_{\tilde{\mathcal{E}}} \ , \end{cases} \tag{4.41}$$

with h holding the same definition as in eq (4.38).

4.4.3 Experiment

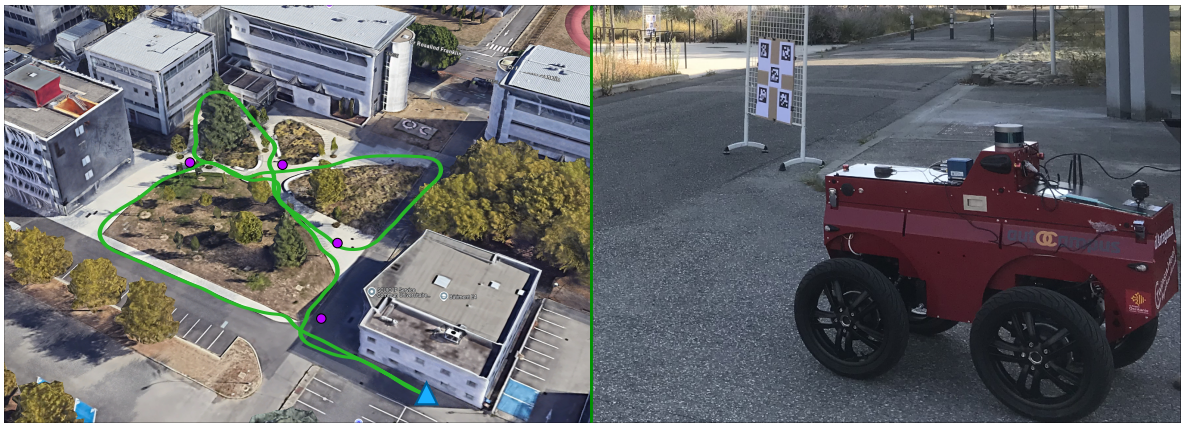


Figure 4.14: SLAM experiment conducted at Université de Toulouse, in the court and parking of building U4 (right photo - Google Maps). The "D'Artagnan" mobile robot (right photo) from Twinswheel starts from the parking spot (blue triangle in the lower left of the right photo), makes loops and goes back to the original spot. GNSS trace is in green (precision $< 2\ cm$). Four landmarks, each being a bundle of five "April Tags" (Wang and Olson, 2016), are scattered in the scene (purple points), to be detected by the robot frontal camera (Intel RealSense D435).

The CARLIT approach has been tested on a large scale experiment conducted in the Toulouse III Paul Sabatier Campus. All the facilities were provided by the autOCampus plaftorm (https://www.irit.fr/autocampus/en/accueil-eng/ funded by Occitanie Region (laregion.fr). The "d'Artagnan" mobile robot, provided by TwinswHeel company (https://www.twinswheel.fr/) was navigated in the Campus. Its RTK-GNSS track was computed by GUIDE GNSS Geolocation Test Lab (https://guide-gnss.com/). We gratefully acknowledge the autOCampus steering staff, TwinswHeel and GUIDE-GNSS, as well as Anushree Shrivastava and Augusto Capone, intern students at LAAS-CNRS.

Figure 4.14 portrays the experimental setting. The robot, running the Robot Operating System (ROS), is controlled manually up to a max-speed of $2m/s$. An embedded frontal camera registers the April Tag detections and an internal node provides a measure of their relative position $\bar{y}$ using the known camera parameters and the tags configuration.

The robot path is made of a total of $N \triangleq 6797$ poses, and 203 landmarks relative position measurements (including 4 discoveries). The problem is approximated to be planar: the set

of poses $\mathbf{x_0}, \ldots, \mathbf{x}_N$ is in SE(2), the motion errors are in the Euclidean space isomorphic to the Lie algebra of SE(2) (i.e., $\forall i$, $\zeta_{i-1}^i \in \mathfrak{se}(2)^\vee$), and the relative position measurement errors are in $\mathbb{R}^2$.

The first step of this SLAM algorithm consists in identifying, from the causal representation, the set of independent blocks that are to be adjusted (as per the **SplitIndependentMotionBlocks** function in algorithm 2). In this problem, 1743 of such independent blocks are identified, and all but one are made of a lone, independent error. A single block of mutually dependent errors regroups all the landmark re-observations: it will be denoted as 'main' block in what follows. The remaining 1742 independent systems are: (a) 979 lone motion errors gathered at the beginning of the experiment, before the first landmark re-observation, and (b) 763 lone motion errors after the last landmark re-observation. Only the 'main' block can be adjusted. In the experiment on figure 4.14-(right), the independent errors are associated with the poses close to the starting and final point (blue triangle in the parking).

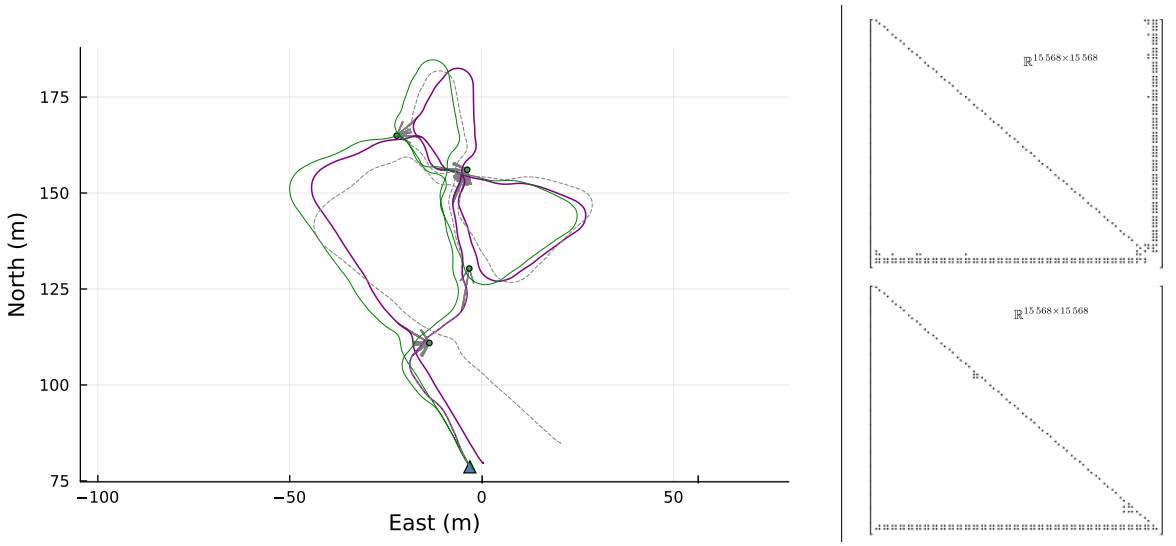


Figure 4.15: Left-Side: Path of the robot according to: fusion of IMU and wheel odometry (dash gray), GNSS trace (green), and MAP estimate resulting from the application of the CARLIT adjustment (purple). Right-side: the pattern of the sparse matrix M appearing in (4.26) and the pattern of its Cholesky decomposition $LDL^\top$. The matrix block of the constraints is visible at the bottom part of M and its transpose at the left part.

The approach is implemented in a Julia module. The 'main' adjustment block is composed of 5054 motion errors and 199 re-observations of landmarks. All these errors, noted $\mathcal{E}^{(1)}$, are mutually dependent a posteriori. In view of the above, 199 constraints of the form (4.36) are exhibited, and the MAP estimate is sought for according to the variant of the formulation (4.16):

$$\hat{\mathcal{E}}^{(1)} \triangleq \underset{\mathcal{E}}{\operatorname{argmin}} \frac{1}{2} \left\| \rho^\top \mathcal{E}^{(1)} \right\|_2^2 \quad \text{s.t.} \quad C_{ro}(\mathcal{E}^{(1)}) = d_{ro} \ , \tag{4.42}$$

with

$$\rho\rho^\top \triangleq \begin{bmatrix} \Sigma_{\text{rm}} & & & & & \\ & \ddots & & & & \\ & & \Sigma_{\text{rm}} & & & \\ & & & \Sigma_{\text{l}} & & \\ & & & & \ddots & \\ & & & & & \Sigma_{\text{l}} \end{bmatrix}^{-1} . \tag{4.43}$$

The covariance of the motion errors and of the landmark relative position measurements are set to be, respectively, $\Sigma_{\text{rm}} \triangleq \begin{bmatrix} 10^{-7} & & \\ & 10^{-9} & \\ & & 10^{-6} \end{bmatrix}$, and $\Sigma_l \triangleq \begin{bmatrix} 1/44 & \\ & 1/44 \end{bmatrix}$.

Then, from an initialization point of $\mathcal{E}^{(1)}$, a linearized constrained least squares system is solved, as in (4.26), and reiterated 7 times, until convergence is achieved. The pattern of the involved sparse matrices are shown figure 4.15-(right). The robot poses and landmarks are then deduced.

The purple path on figure 4.15-(left) depicts the MAP estimate, and it can be seen that the estimate is appreciably closer to the ground truth (GNSS path in green) than the filtered odometry (in dash gray).

More thorough analysis of this experiment is deferred for later, as additional developments are needed to represent (causally) the structural relationship between the robot relative motions and the GNSS data. A difficulty arises from the fact that the GNSS data and the robot motions are not at the same frequency, and that the IMU-odometry fusion (out of reach) is imperfect. This is an interesting problem that the causal approach can address in the future.

4.5 Composition of motions to decrease treewidth

The adjustment processes in the previous sections 4.3 and 4.4 notably include the solutions of a matrix system, as per (4.26). Specifically, the pattern of the symmetric square matrix in such a system is:

$$M \triangleq \begin{bmatrix} A^\top A & C^\top \\ C & 0 \end{bmatrix} , \tag{4.44}$$

where $A^\top A$ is block diagonal, and the sparsity of C is dependent on the problem. The complexity of the adjustment process can be studied through the sparsity pattern of M, since, to solve (4.26), the decomposition step of M is a computational bottleneck for such large matrix.

Consider the classical variant of the Cholesky decomposition such that $LDL^\top = M$, with L a lower triangular matrix and D a diagonal matrix. It is known, in scientific computing, that the graph $\mathcal{G}_L$ associated with the sparsity pattern of L is the induced/cover graph of M (Kepner and Gilbert, 2011).

With the material presented so far in this chapter, the complexity of the pose-graph Manhattan 3500 dataset (Olson et al., 2006) is analyzed and compared with the complexity of a matrix based probabilistic SLAM solution. Firstly, on a positive note, the causal representation can divide this adjustment problem into several independent problems, following the approach described in section 4.2. The split step (line 2 in algorithm 2) shows that this

particular dataset is divisible in 15 independent adjustment blocks, including 3 blocks which span, respectively, 2945, 417 and 96 motion steps, 2 other small adjustment blocks, and the remainder 10 blocks being lone errors (and thus non adjustable).

Approach	Matrix size	non-zeros number	treewidth	timing of $LDL^\top$ decomposition
graph SLAM	3 499	21 992	44	$2.2ms \pm 276\mu s$
CARLIT M3500_2945	6 226	1 025 286	561	$218ms \pm 32ms$
CARLIT M3500_417	914	37 456	178	$3.867ms \pm 301\mu s$
CARLIT M3500_96	38	2 863	38	$210\mu s \pm 301\mu s$

Table 4.2: Metrics of the sparsity patterns of the matrices shown figure 4.16. Timing data are produced in the Julia language with an i7 10th gen CPU.

In figure 4.16, the semantic pattern[9] of the involved sparse matrices for both a probabilistic view of SLAM and the CARLIT approach are depicted. In table 4.2, we analyze performance metrics in the treewidth, the number non-zeros count and timings associated with the Cholesky decomposition.

The results show that, although the causal approach can leverage the independence property to decompose its blocks in parallel, each of the 2 largest blocks have each worse performance than the matrix based approach of probabilistic SLAM. The fact that the sparsity is worse for the causal case can in fact be judged directly on figure 4.16, especially for the lower triangular matrices. More concerning for large scale applications, the largest block coined M3500_2945 (which gathers the right side of the Manhattan map) is analyzed to have a treewidth of 561 (i.e., the size of the maximum clique), which is dramatically greater than the treewidth of 44 for probabilistic SLAM.

In view of our discussion §2.3.3, we consider the treewidth to be a significant metric for judging the ability of the approach to perform well at scale. Therefore, this motivates to research improvements to decrease the treewidth. The main insight is to observe that motions can be heuristically composed in order to gather overlapping summarized motions in another relative motion. This is described thereafter.

4.5.1 Wrapping Procedure

In what follows, the term 'unwrapped' graph denotes the causal graph built according to the approach described in the previous sections 4.2, 4.3, 4.4. The goal here is to 'wrap' the causal graph (or made it 'compact'), by introducing new intermediary summarized motions, such that each summarized motion has only two parents and as few children as possible. For those purposes, virtual measurements $\breve{\mathbf{z}}$ and virtual errors $\breve{\zeta}$ are introduced, distinguishable from 'real' measurements and 'real' errors by the breve "$\breve{}$" overlay. Note that this is an exact step, as no approximation is introduced.

Consider the causal graph figures 4.1-(c) and 4.1-(d) in the introduction of this chapter, which main motion block is reproduced figure 4.17 for convenience. The (partial) SCM, which

[9] The 'blocks' of the matrix are either i or -1, as if the problem was projected in $\mathbb{R}$. This is done in order to reason at the semantic level.

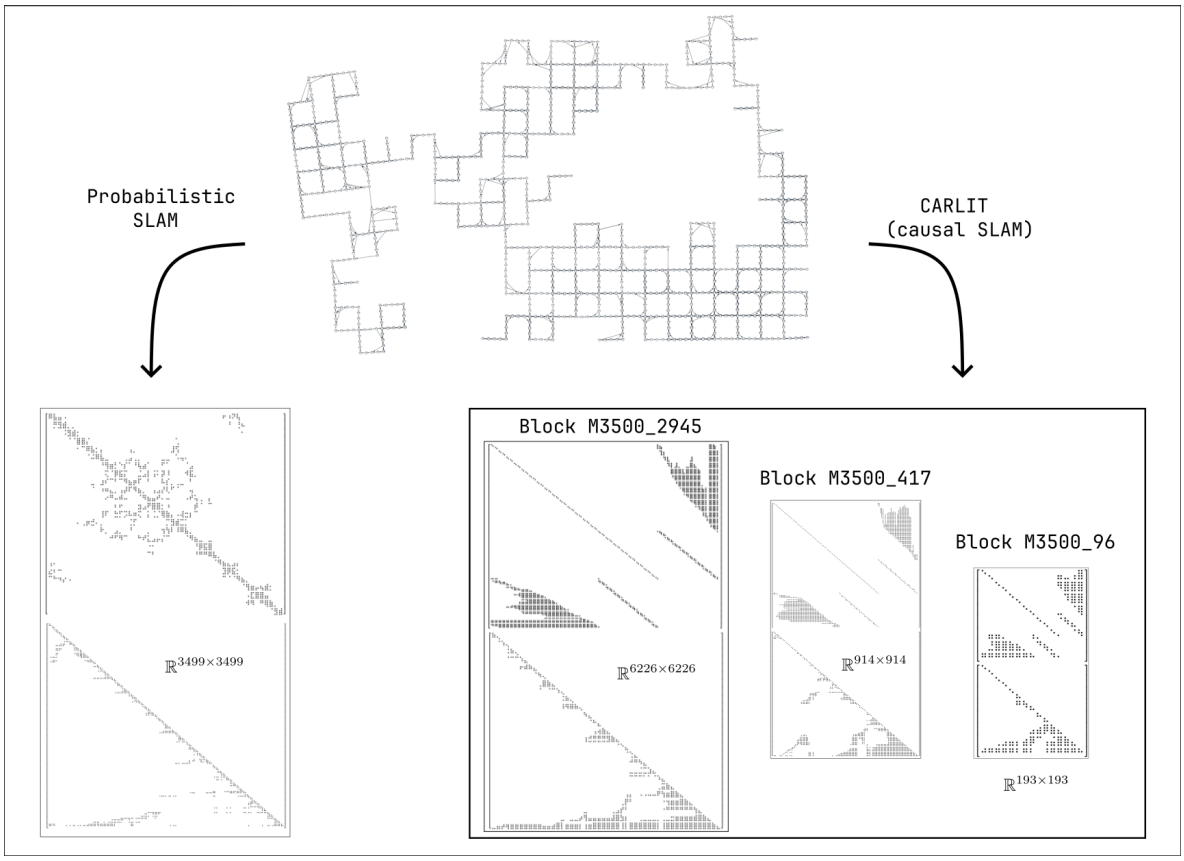


Figure 4.16: Comparison of the sparsity patterns of the matrix based probabilistic approach (left) and the CARLIT approach (right), on the Manhattan 3500 dataset (drawn at the top). The matrix based probabilistic approach portrays the information matrix and its $LDL^\top$ decomposition below it. The CARLIT approach only depicts the 3 largest independent blocks (there are 15 in total) with, for each, the matrix M from (4.44) and its corresponding $LDL^\top$ decomposition below it. Sizes of the matrices are shown on heterogeneous scale.

the motion block of figure 4.17-(b) is based on, writes as:

$$(SCM): \begin{cases} \mathbf{\Lambda}_2^6 & := \mathbf{\Lambda}_2^3 \circ \mathbf{\Lambda}_3^4 \circ \mathbf{\Lambda}_4^5 \circ \mathbf{\Lambda}_5^6 \\ \mathbf{z}_2^3 & := \mathbf{\Lambda}_2^3 \circ \mathrm{Exp}\left(\zeta_2^3\right) \\ \mathbf{z}_3^4 & := \mathbf{\Lambda}_3^4 \circ \mathrm{Exp}\left(\zeta_3^4\right) \\ \mathbf{z}_4^5 & := \mathbf{\Lambda}_4^5 \circ \mathrm{Exp}\left(\zeta_4^5\right) \\ \mathbf{z}_5^6 & := \mathbf{\Lambda}_5^6 \circ \mathrm{Exp}\left(\zeta_5^6\right) \\ \mathbf{z}_2^6 & := \mathbf{\Lambda}_2^6 \circ \mathrm{Exp}\left(\zeta_2^6\right) \end{cases} . \tag{4.45}$$

The structural equation $\mathbf{\Lambda}_2^6 := \mathbf{\Lambda}_2^3 \circ \mathbf{\Lambda}_3^4 \circ \mathbf{\Lambda}_4^5 \circ \mathbf{\Lambda}_5^6$ can be decomposed into three structural equations, with no repercussion, on for the causal interpretation: $\mathbf{\Lambda}_2^4 := \mathbf{\Lambda}_2^3 \circ \mathbf{\Lambda}_3^4$, $\mathbf{\Lambda}_2^5 := \mathbf{\Lambda}_2^4 \circ \mathbf{\Lambda}_4^5$ and $\mathbf{\Lambda}_2^6 := \mathbf{\Lambda}_2^5 \circ \mathbf{\Lambda}_5^6$.

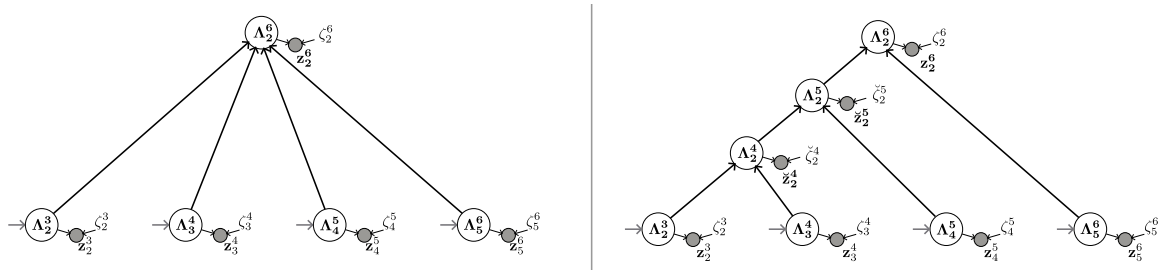


Figure 4.17: Partial graphs reproduced from figure 4.1-(c) and 4.1-(d). These causal graphs only depict of the central motion block (loop-closure between $\mathbf{x_2}$—$\mathbf{x_6}$).

This leads to the wrapped SCM:

$$(SCM): \begin{cases} \mathbf{\Lambda}_2^4 := \mathbf{\Lambda}_2^3 \circ \mathbf{\Lambda}_3^4, \\ \mathbf{\Lambda}_2^5 := \mathbf{\Lambda}_2^4 \circ \mathbf{\Lambda}_4^5, \\ \mathbf{\Lambda}_2^6 := \mathbf{\Lambda}_2^5 \circ \mathbf{\Lambda}_5^6, \\ \breve{\mathbf{z}}_2^4 := \mathbf{\Lambda}_2^4 \circ \mathrm{Exp}\left(\breve{\zeta}_2^4\right), \\ \breve{\mathbf{z}}_2^5 := \mathbf{\Lambda}_2^5 \circ \mathrm{Exp}\left(\breve{\zeta}_2^5\right), \\ \mathbf{z}_2^3 := \mathbf{\Lambda}_2^3 \circ \mathrm{Exp}\left(\zeta_2^3\right), \\ \mathbf{z}_3^4 := \mathbf{\Lambda}_3^4 \circ \mathrm{Exp}\left(\zeta_3^4\right), \\ \mathbf{z}_4^5 := \mathbf{\Lambda}_4^5 \circ \mathrm{Exp}\left(\zeta_4^5\right), \\ \mathbf{z}_5^6 := \mathbf{\Lambda}_5^6 \circ \mathrm{Exp}\left(\zeta_5^6\right), \\ \mathbf{z}_2^6 := \mathbf{\Lambda}_2^6 \circ \mathrm{Exp}\left(\zeta_2^6\right), \end{cases} \tag{4.46}$$

where we introduce virtual measurements $\{\breve{\mathbf{z}}_2^4, \breve{\mathbf{z}}_2^5\}$ and virtual errors $\left\{\breve{\zeta}_2^4, \breve{\zeta}_2^5\right\}$. Virtual errors are given flat priors. Virtual measurement values are set arbitrarily, which should be a middle-of-the-road guess between the real measurements in the ancestor and descendant motion set. Different types of averaging methods can be imagined depending on the Lie group. This step can be viewed as replacing the selection of the initial point which bootstraps the optimization process in MLE SLAM. A 'good' guess would ensure that no overflows occurs for large mismatches between real measurements, as per the assumption made in footnote 6, i.e. that $\mathrm{Log}\left(\mathrm{Exp}\left(\mathbf{x}\right)\right) = \mathbf{x}$ for $\mathbf{x} \in \mathrm{SE}(\cdot)$. Thus, the setting of $\breve{\mathbf{z}}$ values deserves its own study, and is not the focus in this section (decreasing the treewidth).

Remark 4.9 (No change in existing pattern of conditional independences)**.** *The introduction of the virtual measurements produce closed collider nodes in the wrapped causal graph, figure 4.17-(b). As the error nodes within the motion block are already mutually dependent in the unwrapped graph on figure 4.17-(a), it can be verified that no changes are made to existing patterns of conditional independences.*

The heuristic aspect of the approach concerns the selection of the $\mathbf{\Lambda}$-motion to be composed. Another option would have been to exhibit $\mathbf{\Lambda}_4^6$ or $\mathbf{\Lambda}_3^5$. The rational here is to summarize 'old' motions first in order to open to the opportunity to remove older measurements and grouping them together into one motion (by then approximating the virtual measurements/errors as 'real') for long term navigation purposes.

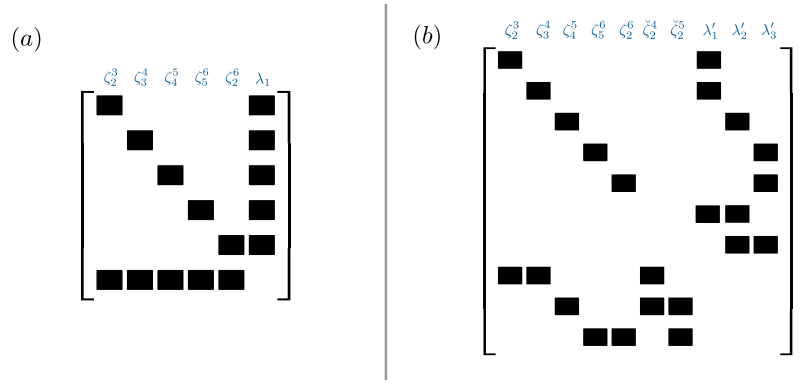


Figure 4.18: Block pattern of the M matrices of the unwrapped SCM (4.45) and the wrapped SCM (4.46).

The effect of this wrapping on the adjustment process is portrayed on figure 4.17. The form of the matrices M are depicted for both unwrapped and wrapped cases. In the case of figure 4.17-(b), note the addition of the columns $\left\{\breve{\zeta}_2^4, \breve{\zeta}_2^5\right\}$ with no value (zero blocks) in the $A^\top A$ block, illustrating the fact that they have flat priors and will not change the end result of the inference process (they are free parameters).

Furthermore, the single constraint that existed in the unwrapped SCM (4.45) has been split into 3 constraints in the wrapped SCM (4.46). First, the constraint available from the unwrapped SCM (4.45) is, following our explanations section 4.3:

$$\mathbf{z}_2^6 \circ \mathrm{Exp}\left(-\zeta_2^6\right) = \mathbf{z}_2^3 \circ \mathrm{Exp}\left(-\zeta_2^3\right) \circ \mathbf{z}_3^4 \circ \mathrm{Exp}\left(-\zeta_3^4\right) \circ \mathbf{z}_4^5 \circ \mathrm{Exp}\left(-\zeta_4^5\right) \circ \mathbf{z}_5^6 \circ \mathrm{Exp}\left(-\zeta_5^6\right) \ , \quad (4.47)$$

while the constraints stemming from the wrapped SCM (4.46) are:

$$\begin{cases} \mathbf{z}_2^6 \circ \mathrm{Exp}\left(-\zeta_2^6\right) = \breve{\mathbf{z}}_2^5 \circ \mathrm{Exp}\left(-\breve{\zeta}_2^5\right) \circ \mathbf{z}_5^6 \circ \mathrm{Exp}\left(-\zeta_5^6\right), \\ \breve{\mathbf{z}}_2^5 \circ \mathrm{Exp}\left(-\breve{\zeta}_2^5\right) = \breve{\mathbf{z}}_2^4 \circ \mathrm{Exp}\left(-\breve{\zeta}_2^4\right) \circ \mathbf{z}_4^5 \circ \mathrm{Exp}\left(-\zeta_4^5\right), \\ \breve{\mathbf{z}}_2^4 \circ \mathrm{Exp}\left(-\breve{\zeta}_2^4\right) = \mathbf{z}_2^3 \circ \mathrm{Exp}\left(-\zeta_2^3\right) \circ \mathbf{z}_3^4 \circ \mathrm{Exp}\left(-\zeta_3^4\right). \end{cases} \quad (4.48)$$

Consequently, each constraint only involves 3 errors, which materializes downstream into 3 Jacobian blocks per block row in the constraint matrix C of M (4.44), whatever the span range of the loop-closure. From a complexity/cost perspective, it has a positive impact on the sparsity pattern, while the trade-off are the increases of (a) the size of the sparse matrix, and (b) the number of non zeros blocks.

In this toy example, there are limited gains to be expected: in both wrapped and unwrapped cases, the treewidth is 2 following the decomposition of matrix M. The real interest of the wrapping approach appears when loop closures overlap each other. Hereafter, we apply a candidate wrapping procedure, formally presented in algorithm 4, to two examples of §4.5.2 and §4.5.3.

Algorithm 4 proceeds by pairing consecutive $\mathbf{\Lambda}$ motions together, starting from the relative motions $\mathbf{\Lambda}_{i-1}^i$. For each pair $\left\{\mathbf{\Lambda}_{i-i}^i, \mathbf{\Lambda}_i^{i+1}\right\}$, an intermediary motion is created if both nodes have the same set of children, i.e., if they are in the same loop-closures. The pairs $\left\{\mathbf{\Lambda}_{i-i}^i, \mathbf{\Lambda}_i^{i+1}\right\}$ are served in a queue and processed one-by-one; details are given in algorithm 5. Whenever an intermediary motion is created, it can in turn be paired with the next motion and join the

Algorithm 4: Wrapping procedure of a causal graph in Carlit.

Input: Unwrapped Causal Graph G
Output: Wrapped Causal Graph G_w

$G_{\mathbf{\Lambda}} \leftarrow$ `RemoveTrajectory(`G`)` ;
$\mathcal{G} \leftarrow$ `SplitIndependentMotionBlocks(`$G_{\mathbf{\Lambda}}$`)` ; `// Split graph by` $\mathbf{\Lambda}$ `sets`
for $G_{\mathbf{\Lambda}}^{(i)} \in \mathcal{G}$ **do**
  $k \leftarrow 0$;
  $Q_0 \leftarrow$ `ConsecutivePairs(`$G_{\mathbf{\Lambda}}^{(i)}$`)`;
  while $Q_k \neq \emptyset$ **do**
    $Q_{k+1}, G_{\mathbf{\Lambda}}^{(i)} \leftarrow$ `ProcessQueue(`$Q_k, k, G_{\mathbf{\Lambda}}^{(i)}$`)` ; `// See alg 5`
    $k \leftarrow k + 1$;
  end
end
$G_w \leftarrow$ `Merge(`$\mathcal{G}$`)`;
return G_w ;

queue. When the queue of pairs that have the same set of children (noted Q_0) is ended, the process continues with pairs that have some difference(s) in their children set, noted Q_k with k increasing from 1 on. This is iterated until all summarized and intermediary motions have two parents.

Additional details such as the functions used in algorithm 5 are provided in appendix B.

Algorithm 5: ProcessQueue: Given a queue of consecutive pairs of $\mathbf{\Lambda}$-motions of a motion block graph $G_{\mathbf{\Lambda}}$, wrap the elements of the queue whose *setdiff* score does not exceed k. This results in a new queue Q_{k+1} and modified graph $G_{\mathbf{\Lambda}}$.

Input: Queue Q_k , Graph $G_{\mathbf{\Lambda}}$, integer k
Output: next queue Q_{k+1}, Graph $G_{\mathbf{\Lambda}}$ recomposed

$Q_{k+1} \leftarrow \emptyset$;
updateFirst $\leftarrow$ False;
previousComposition $\leftarrow \emptyset$;
StagedPair $\leftarrow \emptyset$;
while $Q_k \neq \emptyset$ **do**
  pair $\leftarrow$ `Take`(Q_k) ; // pair $= \{\mathbf{\Lambda_i^j}, \mathbf{\Lambda_j^k}\}$
  if updateFirst == True **then**
    pair $\leftarrow$ `Join`(previousComposition, pair);
    updateFirst $\leftarrow$ False;
  end
  $\mathrm{I}_{ijk} \leftarrow$`ChildrenIntersection`$(G_{\mathbf{\Lambda}}, \mathbf{\Lambda_i^j}, \mathbf{\Lambda_j^k})$;
  $\mathrm{U}_{ijk} \leftarrow$`ChildrenUnion`$(G_{\mathbf{\Lambda}}, \mathbf{\Lambda_i^j}, \mathbf{\Lambda_j^k})$;
  if $\mathrm{I}_{ijk} = \emptyset$ **then**
    continue ; // Do nothing, continue to next element in queue
  else if $|\mathrm{U}_{ijk}| - |\mathrm{I}_{ijk}| > k$ **then**
    // the pair should be processed in the next queue: stage it first
    if StagedPair $\neq \emptyset$ **then** // commit prior staged pair to next Q
      $Q_{k+1} \leftarrow Q_{k+1} \cup$ StagedPair ;
    end
    StagedPair $\leftarrow$ pair;
  else
    if $|\mathrm{I}_{ijk}| > 1$ or $\mathbf{\Lambda_i^k} \notin G_{\mathbf{\Lambda}}$ **then** // if more than 1 common child
      $\mathbf{\Lambda_i^k}$, $G_{\mathbf{\Lambda}} \leftarrow$ `Recompose`$(G_{\mathbf{\Lambda}}$, pair);
      previousComposition $\leftarrow$ $\mathbf{\Lambda_i^k}$;
      if StagedPair $\neq \emptyset$ **then**
        StagedPair $\leftarrow$ `Join`$(\mathbf{\Lambda_i^k}$, StagedPair);
      end
      updateFirst $\leftarrow$ True;
    end
  end
  previousPair $\leftarrow$ pair;
end
if StagedPair $\neq \emptyset$ **then**
  $Q_{k+1} \leftarrow Q_{k+1} \cup$ StagedPair ;
end
return $Q_{k+1}, G_{\mathbf{\Lambda}}$;

4.5.2 Motion Recomposition of Mini Manhattan

The first case study is a cut of the first 15 steps of the synthetic dataset experiment Manhattan 3500, denoted MiniManhattan. The case is interesting in that it depicts 4 loop-closures which overlap each other in an inconvenient way over the robot pose variables $\{\mathbf{x_7}, \mathbf{x_8}, \mathbf{x_9}, \mathbf{x_{10}}\}$. Figure 4.19 portrayed a probabilistic SLAM view of the problem with the factor graph and

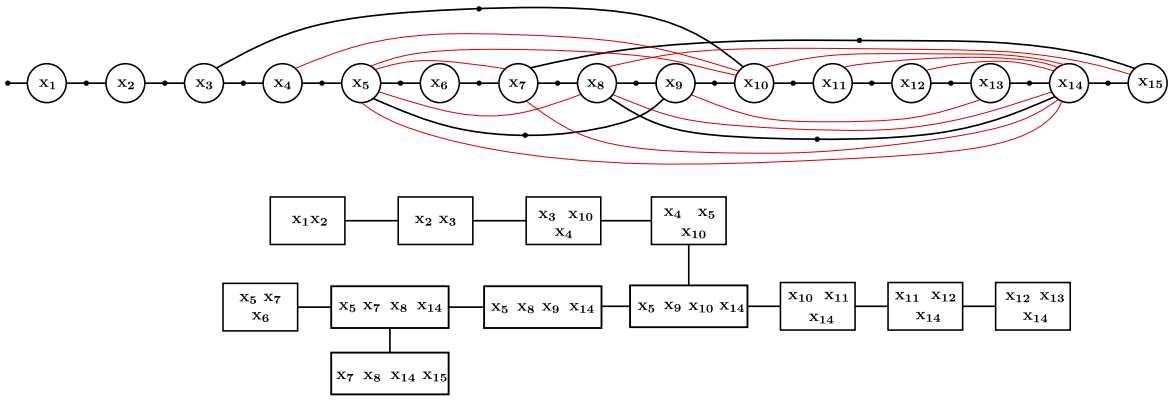


Figure 4.19: Factor graph and clique tree of the MiniManhattan case study. Red edges in factor graph correspond to the fill-in edges of the elimination process. The treewidth (size of the maximum clique) of the clique tree is 4.

the corresponding clique tree. Red edges are the fill-in edges appearing in the elimination process, giving rise to the clique tree. The treewidth, 4, is rather high for such a small problem.

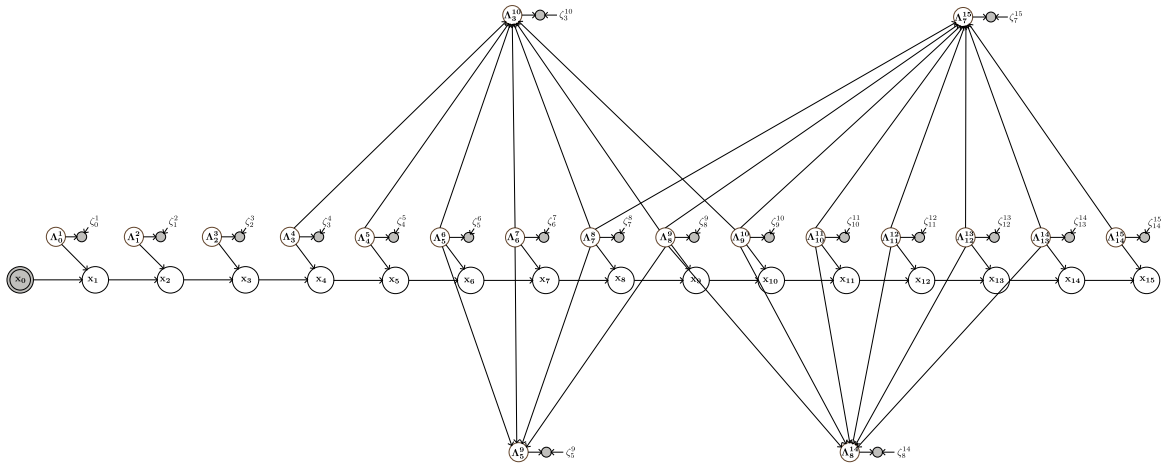


Figure 4.20: Causal graph of the unwrapped SCM of MiniManhattan.

For the CARLIT approach, the causal graph is depicted on figure 4.20. It is built according to the principles delineated in section 4.2, where the four observed summarized motions $\{\mathbf{\Lambda_3^{10}}, \mathbf{\Lambda_5^9}, \mathbf{\Lambda_8^{14}}, \mathbf{\Lambda_7^{15}}\}$ are the loop-closure. The SCM is not explicitly given here, for it bears no interest in this section.

In the adjustment process of the causal approach of MiniManhattan, the constraint matrix

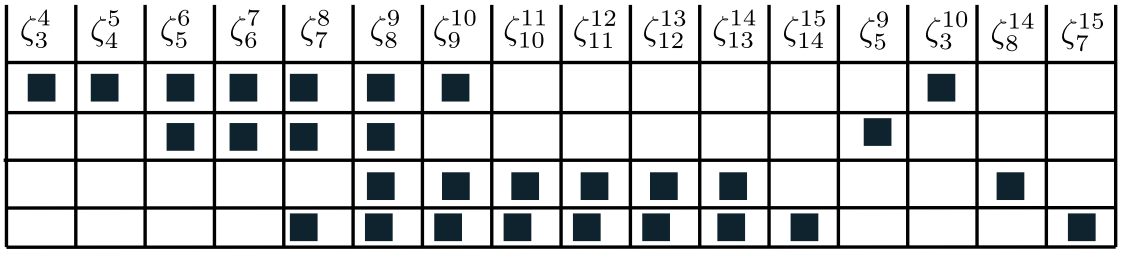


Figure 4.21: Constraint matrix of the unwrapped SCM of the MiniManhattan case study, whose causal graph is shown figure 4.20.

C is produced. Its block pattern is depicted on figure 4.21. The four block rows correspond to the four closures.

The application of the wrapping procedure (algorithm 4) leads to the wrapped causal graph on figure 4.22. The pattern of the new constraint matrix is given on figure 4.23. Ten new block columns (for ten virtual error variables) and ten new block lines (for ten new constraints) have been added to the previous constraint matrix. Finally, figure 4.24 represents the pattern of the full M matrix and its $LDL^\top$ Cholesky decomposition.

The main result from the analysis of the patterns of the various matrices, in the unwrapped and wrapped case, is that the treewidth of the decomposed matrix LD gets decreased from 4 to 3. Note that the treewidth in the wrapped case now scores better (i.e., lower) than the one of probabilistic SLAM: see the clique tree, figure 4.19, which has a treewidth of 4.

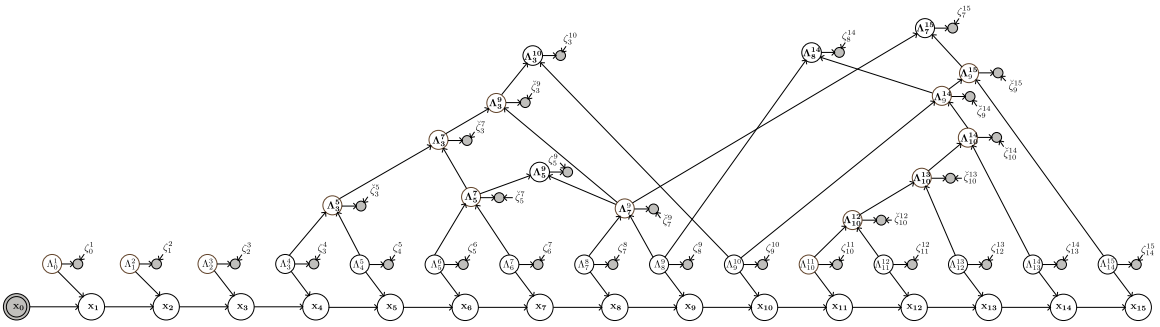


Figure 4.22: Causal graph of the MiniManhattan problem, after application of the proposed wrapping procedure. Ten virtual measurements and virtual errors have been added to the problem compared to the unwrapped version from figure 4.20.

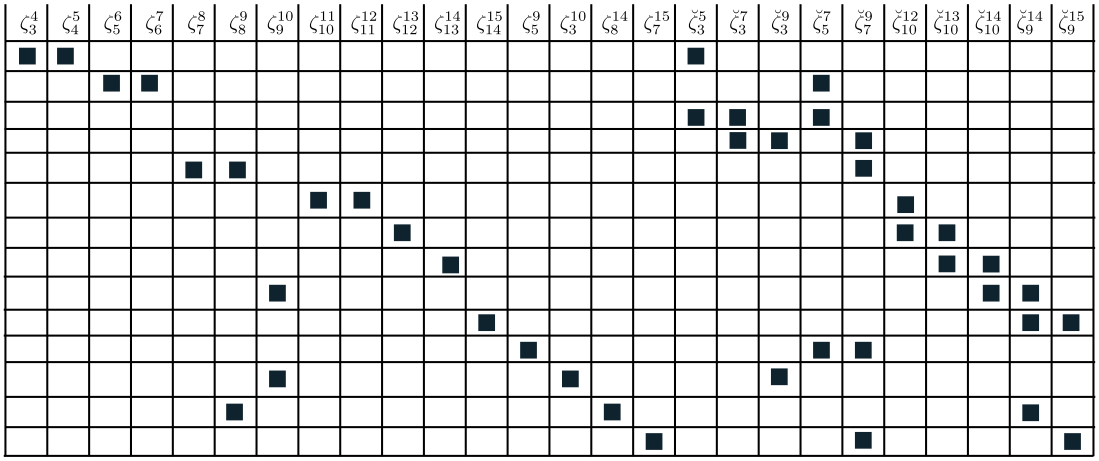


Figure 4.23: Constraint matrix of the wrapped MiniManhattan problem. The unwrapped version is given on figure 4.21.

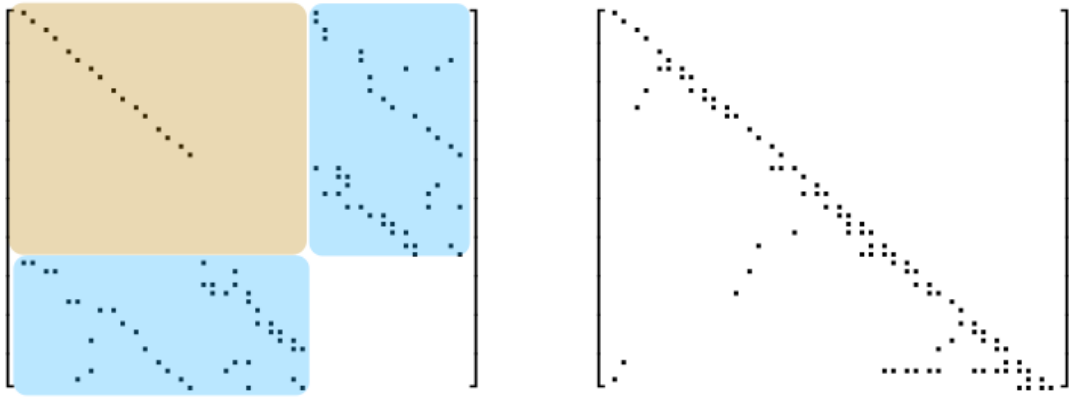


Figure 4.24: Pattern of the full M matrix (left) after the wrapping procedure is applied, and (right) pattern of the lower triangular matrix from the $LDL^\top$ decomposition of M. The blue area in M denotes the constraint matrix (figure 4.23); the diagonal in the beige area, called $A^\top A$ block in eq (4.44) contains the priors on the errors. Note the zeros in lower right of the diagonal, which illustrates the fact that the virtual errors introduced have flat priors.

4.5.3 Motion Recomposition of Mini Victoria Park

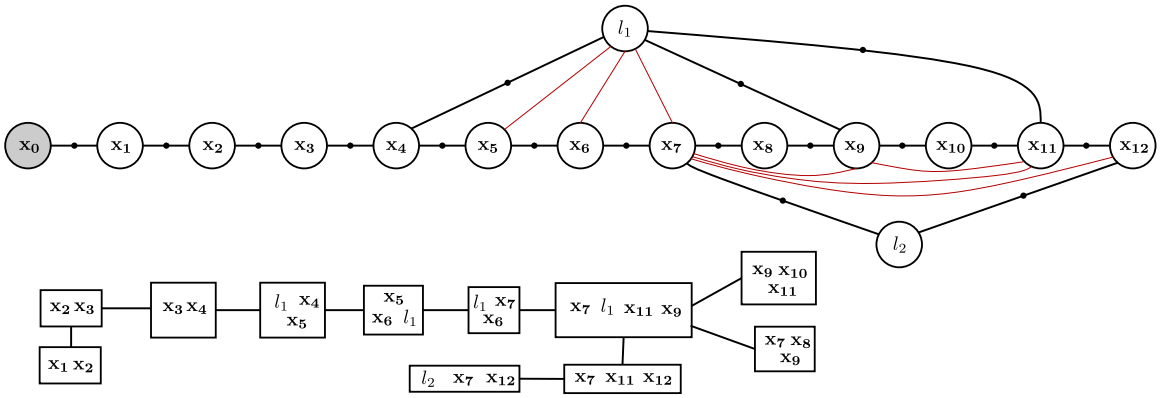


Figure 4.25: Factor graph and clique tree of the MiniVictoriaPark case study. Red edges in factor graph correspond to the fill-in edges of the elimination process. The treewidth (size of the maximum clique) of the clique tree is 4.

The second case study concerns of first twelve steps of the Victoria Park datasets, denoted MiniVictoriaPark in the remainder, in which the re-observations of two landmarks overlap.

The probabilistic view (factor graph and clique tree) is depicted on figure 4.25. The clique tree has a treewidth of four. The causal view is depicted on figures 4.26 (causal graph) and 4.27 (constraint matrix and clique tree of the decomposition). The equivalent clique tree (from the causal approach) on figure 4.27 also has a treewidth of four.

Following the application of the wrapping procedure, the resulting causal graph is portrayed on figure 4.28, and the pattern of the constraint matrix as well as the clique tree corresponding to the Cholesky decomposition of M, are shown on figure 4.29.

Compared to the unwrapped case, the clique tree post-wrap procedure has a larger amount of cliques, but its treewidth has been decreased to 3, from 4. Note that this treewidth scores also better than the clique tree of the probabilistic SLAM (figure 4.25).

The results shown on these examples are promising, and the effects of wrapping will be studied in future work to large scale maps[10]. As takeaway from this section, representing a given spatial inference problem with more variables does not necessarily create adverse effects on the complexity. Recall indeed that, usually, there are more errors than poses and landmarks (or, in probabilistic SLAM terms, there are more factors than there are poses+landmarks). Yet, pursuing the conjunction of (a) representing decision variables as the errors, and (b) proposing a causal view of the problem, actually creates the opportunity to leverage other aspects of structure. On such opportunity concerns the fact that motions can be composed and decomposed according to a heuristic.

Concerning the selected heuristic of the wrapping procedure, at this stage and due to its novelty, it is difficult to evaluate its adequacy: it is not known how well it performs compared to a potential 'optimal' heuristic. A parallel with the mature elimination ordering algorithms could be made here: theoretical results are available, such as the theorem that finding the optimal ordering is NP-hard (Arnborg et al., 1987) but several quasi-optimal heuristics have been well covered in the literature (AMD, Methis, Scotch, etc.), some dedicated to SLAM

[10]So far, it has not been implemented.

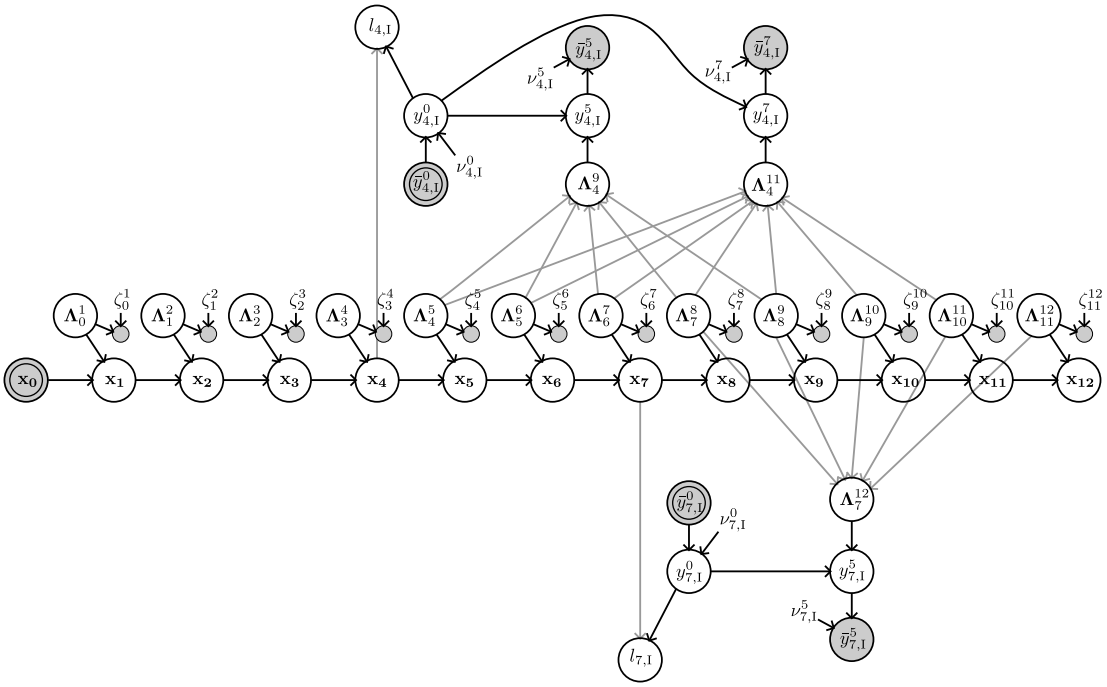


Figure 4.26: Causal graph of the unwrapped SCM of MiniVictoriaPark.

(Agarwal and Olson, 2012).

Finally, note that these compositions are exact, and new mechanics can emerge from there for long term navigation, which were barely touched upon in this section, to make approximate inference by better managing/removing measurements.

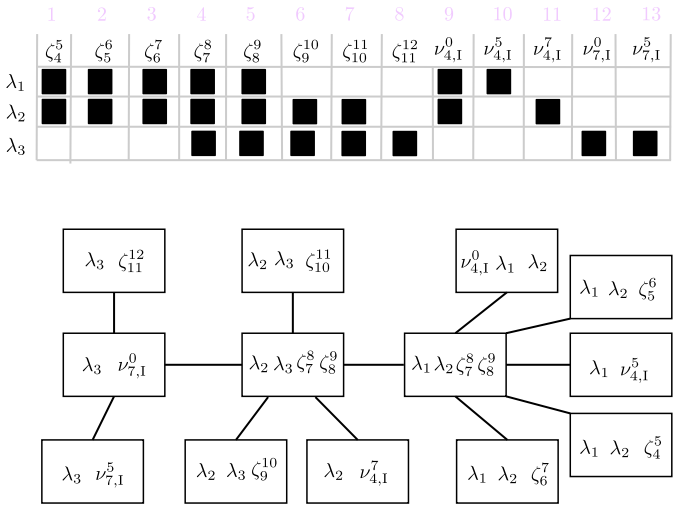


Figure 4.27: Constraint matrix (top) of the unwrapped SCM of the MiniVictoriaPark case study, which causal graph is shown figure 4.26. The clique tree (bottom) is inferred from the (symbolic) matrix decomposition of the matrix M, which contains the constraint block above. Its treewidth is 4.

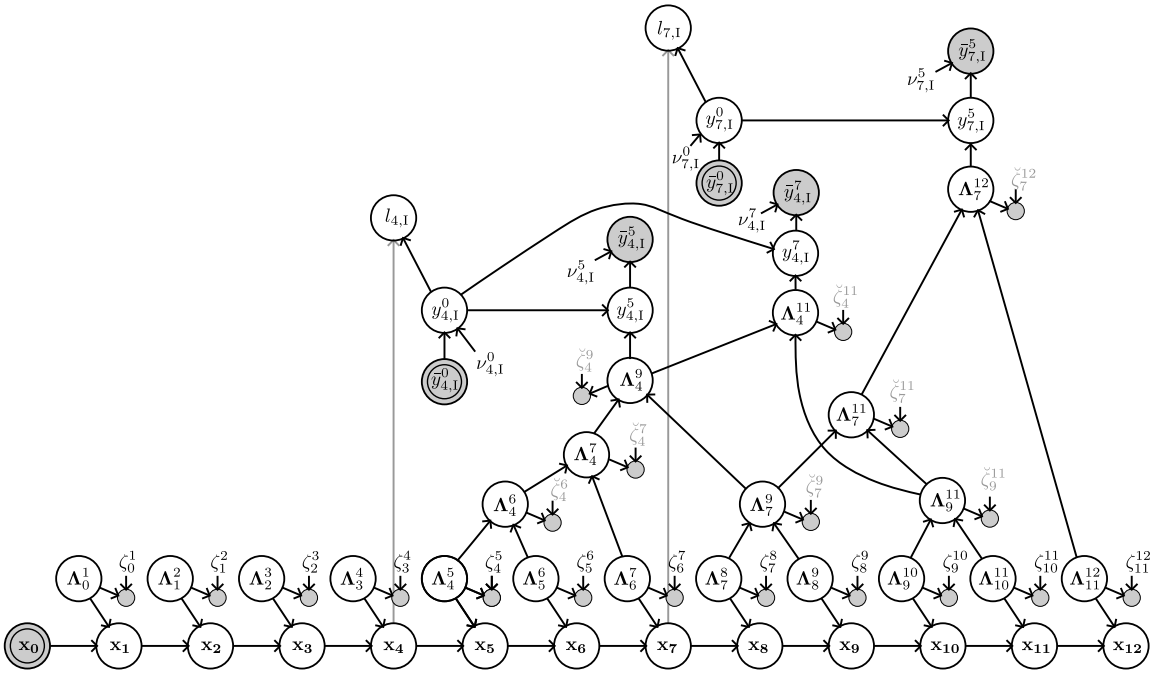


Figure 4.28: Causal graph of the MiniVictoriaPark problem, after application of the proposed wrapping procedure. 8 virtual measurements and virtual errors have been added to the problem compared to the unwrapped version from figure 4.26.

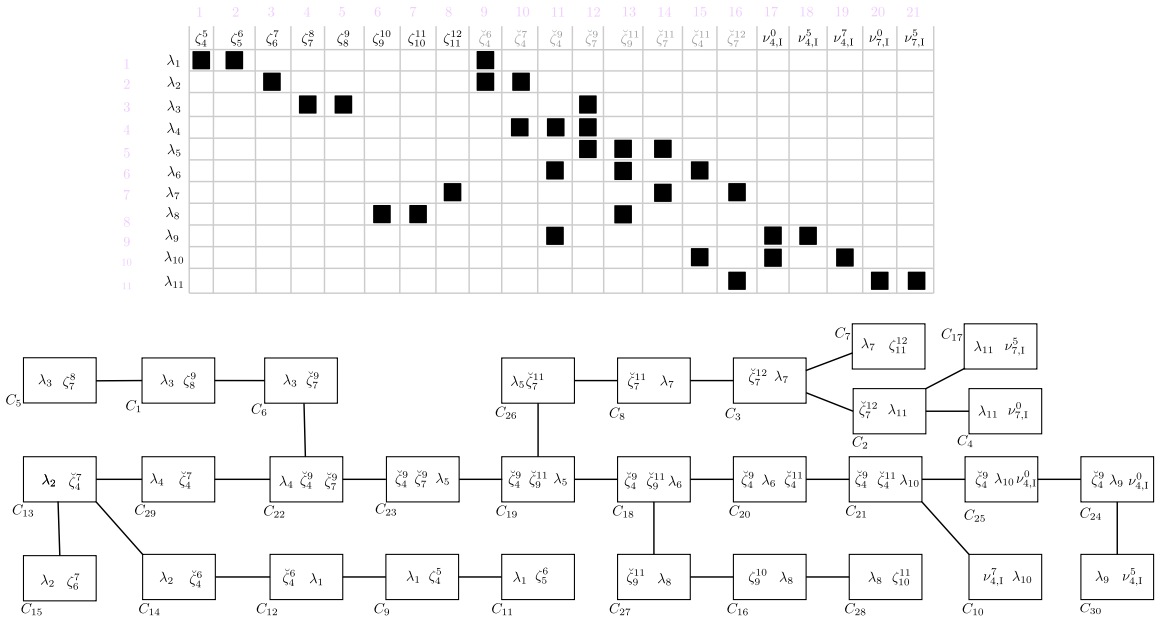


Figure 4.29: Constraint matrix of the wrapped MiniVictoriaPark problem (top) and clique tree induced from the replacement of this constraint matrix in M (bottom). The unwrapped version is given figure 4.27.

4.6 Covariance Recovery

This section addresses the problem of covariance query. So far the proposed CARLIT approach has only tackled the MAP estimate. Covariance query is, often, both (a) necessary in view of the robot task and decision-making; and (b) costly to recover. Typically, in MLE families of SLAM (as per our review of this method in §2.1), the query of the full (and exact) covariance requires the inverse of the information matrix, which is very costly in time and in memory[11]. More astute techniques leverage the Bayes Tree/Clique tree structure (Kaess et al., 2011, 2010). The performance of the query depends on how far away the most distant variables (in the queried set) are from each other in the tree. In CARLIT, the issues that arise when tackling this question are also conceptual: first, do we want a covariance over the error(s) $\mathcal{E}$ or over the robot poses/landmarks?

Our approach to covariance recovery is presented in a 4-variable problem in a 1-dimensional world, on figure 4.30. Three motions are made from $\mathbf{x_0}$ to $\mathbf{x_3}$, of which the measures are corrupted by additive Gaussian noise. At $\mathbf{x_3}$, the distance from $\mathbf{x_0}$ is measured, so that a loop-closure appears. All processes are linear. The noise variances are noted as σ^2_{rm} for the relative motions, and σ^2_{sm} for the single loop-closure measurement.

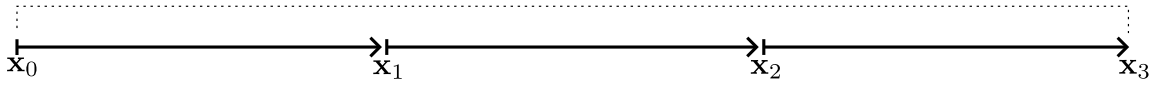


Figure 4.30: Simple 1-dimensional pose-graph SLAM (robot on a rail).

Below, subsections 4.6.1 and 4.6.2 address respectively covariance recovery in the error space and in the robot pose space.

4.6.1 Covariance query on a subset of errors

The SCM of the problem on figure 4.30 writes as:

$$\begin{cases} \mathbf{z}_0^1 := \mathbf{\Lambda}_0^1 + \zeta_0^1, \\ \mathbf{z}_1^2 := \mathbf{\Lambda}_1^2 + \zeta_1^2, \\ \mathbf{z}_2^3 := \mathbf{\Lambda}_2^3 + \zeta_2^3, \\ \mathbf{\Lambda}_0^3 := \mathbf{\Lambda}_0^1 + \mathbf{\Lambda}_1^2 + \mathbf{\Lambda}_2^3, \\ \mathbf{z}_0^3 := \mathbf{\Lambda}_0^3 + \zeta_0^3, \\ \mathbf{x_1} := \mathbf{\Lambda}_0^1 + \mathbf{x_0}, \\ \mathbf{x_2} := \mathbf{\Lambda}_1^2 + \mathbf{x_1}, \\ \mathbf{x_3} := \mathbf{\Lambda}_2^3 + \mathbf{x_2}, \end{cases} \tag{4.49}$$

Our approach consists in writing the first four equations of (4.49) in a posterior joint pdf of the errors, with a Dirac delta function in place of the deterministic relation between the errors:

[11] While the information matrix is sparse, its inverse is dense.

$$p(\mathcal{E}|\mathcal{Z}) = \mathcal{N}\left(\zeta_0^1;\ 0, \sigma_{\text{rm}}^2\right)\mathcal{N}\left(\zeta_1^2;\ 0, \sigma_{\text{rm}}^2\right)\mathcal{N}\left(\zeta_2^3;\ 0, \sigma_{\text{rm}}^2\right)\mathcal{N}\left(\zeta_0^3;\ 0, \sigma_{\text{sm}}^2\right) \times \delta\left(\zeta_0^3 - (\mathbf{z}_0^3 - \mathbf{z}_0^1 + \mathbf{z}_1^2 + \mathbf{z}_2^3 + \zeta_0^1 + \zeta_1^2 + \zeta_2^3)\right) . \quad (4.50)$$

It is noticed that no valid full covariance over $\mathcal{E}$ can be studied, for the Dirac delta $\delta(\cdot)$ would lead to a singular matrix.

Consider the covariance query over $\mathcal{E}_q \triangleq [\zeta_0^1, \zeta_1^2, \zeta_2^3]^\top$. The joint pdf (4.50) marginalizes using the Dirac identity (Dirac elimination):

$$\int_{\zeta_0^3} p(\mathcal{E}|\mathcal{Z})\, d\zeta_0^3 = \mathcal{N}\left(\zeta_0^1;\ 0, \sigma_{\text{rm}}^2\right)\mathcal{N}\left(\zeta_1^2;\ 0, \sigma_{\text{rm}}^2\right)\mathcal{N}\left(\zeta_2^3;\ 0, \sigma_{\text{rm}}^2\right)\phi_q\left(\mathcal{E}_q;\ \mathcal{Z}\right) , \quad (4.51)$$

with the potential ϕ_q proportional to

$$\phi_q\left(\mathcal{E}_q;\ \mathcal{Z}\right) \propto \exp\left\{-\frac{1}{2}\left(A_q\mathcal{E}_q - b_q\right)^\top\left(A_q\mathcal{E}_q - b_q\right)\right\} , \quad (4.52)$$

with

$$A_q \triangleq \begin{bmatrix} \sigma_{\text{rm}} & & & \\ & \sigma_{\text{rm}} & & \\ & & \sigma_{\text{rm}} & \\ \sigma_{\text{sm}} & \sigma_{\text{sm}} & \sigma_{\text{sm}} & \sigma_{\text{sm}} \end{bmatrix} \quad \text{and} \quad b_q \triangleq \begin{bmatrix} 0 \\ 0 \\ 0 \\ \mathbf{z}_0^3 - \mathbf{z}_0^1 - \mathbf{z}_1^2 - \mathbf{z}_2^3 \end{bmatrix} . \quad (4.53)$$

The covariance over $\mathcal{E}_q$ is given by

$$\Sigma_{\mathcal{E}_q} = \left(A_q^\top A_q\right)^{-1} \quad (4.54)$$

Remark 4.10. *Note that $A_q^\top A$ is not sparse, which can be problematic for large loop closures. This reinforces the motivation for compositions described in the section 4.5, as (1) more Dirac terms would be introduced in the expression of $p(\mathcal{E}|\mathcal{Z})$ in (4.50), (2) the Dirac elimination step is symbolic (no computations). However, more implementation efforts are required to support this technique.*

4.6.2 Covariance query on a subset of robot poses

We determine the variance over a single robot pose, noted $\mathbf{x}_q$, by writing its deduced value, according to (4.8, p114):

$$\mathbf{x}_q = \mathbf{x}_0 + \sum_{i=1}^{q}\left(\mathbf{z}_{i-1}^i - \zeta_{i-1}^i.\right) \quad (4.55)$$

For the queried pose $\mathbf{x}_q = \mathbf{x}_3$, this yields:

$$\mathbf{x}_3 = \mathbf{x_0} + \mathbf{z}_0^1 + \mathbf{z}_1^2 + \mathbf{z}_2^3 - \zeta_0^1 - \zeta_1^2 - \zeta_2^3 , \quad (4.56)$$

the covariance of which is deduced by the standard computation

$$\Sigma_{\mathbf{x}_q} = [-1, -1, -1]\,\Sigma_{\mathcal{E}_q}\,[-1, -1, -1]^\top . \quad (4.57)$$

Now, considering the joint queried pose as $\mathcal{X}_q \triangleq [\mathbf{x}_1, \mathbf{x}_2]^\top$, the deduced value is:

$$\begin{cases} \mathbf{x}_1 = \mathbf{x}_0 + \mathbf{z}_0^1 - \zeta_0^1, \\ \mathbf{x}_2 = \mathbf{x}_0 + \mathbf{x}_1 + \mathbf{z}_0^1 + \mathbf{z}_1^2 - \zeta_0^1 - \zeta_1^2, \end{cases} \tag{4.58}$$

which leads to the covariance

$$\Sigma_{\mathbf{x}_1,\mathbf{x}_2} = \begin{bmatrix} -1 & 0 & 0 \\ -1 & -1 & 0 \end{bmatrix} \Sigma_{\mathcal{E}_q} \begin{bmatrix} -1 & 0 & 0 \\ -1 & -1 & 0 \end{bmatrix}^\top . \tag{4.59}$$

Numerical application: We consider a large sample application of the on 1-dimensional pose-graph. Let the value of the genuine hidden motions be defined by $\mathbf{\Lambda}_0^1 = \mathbf{\Lambda}_1^2 = \mathbf{\Lambda}_2^3 = 5\,m$ and $\sigma_{\text{sm}} = \sigma_{\text{rm}} = 1$. Then, one gets

$$\Sigma_{\mathcal{E}_q} = \frac{1}{4} \begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix} . \tag{4.60}$$

Then, from (4.57) and (4.59):

$$\Sigma_{\mathbf{x}_3} = \frac{3}{4} \quad \text{and} \quad \Sigma_{\mathbf{x}_1,\mathbf{x}_2} = \frac{1}{4} \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \tag{4.61}$$

Let us now compare this result with the (matrix-based) probabilistic view of SLAM. It was noted earlier that this type of pose-graph, which has a loop-closure on $\mathbf{x}_0$, suffers from a dilemma of representation (u2 in §2.3.2). The choice between choosing $\mathbf{x_0}$ as a decision variable or as an observation was ambiguous. Let A_{fixed} denote the measurement matrix of the probabilistic SLAM system (MLE SLAM) for which $\mathbf{x_0}$ is fixed, and A_{prior} for the case when a prior is set on $\mathbf{x_0}$ (which becomes a decision variable).

We have:

$$A_{\text{fixed}} = \begin{bmatrix} 1 & & \\ -1 & 1 & \\ & -1 & 1 \\ & & 1 \end{bmatrix} \implies \Sigma_{\text{fixed}} = (A_{\text{fixed}}^\top A_{\text{fixed}})^{-1} = \frac{1}{4} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 3 \end{bmatrix} , \tag{4.62}$$

and, let us set the prior on $\mathbf{x_0}$ to $\sigma_{\mathbf{x_0}}^2 \triangleq 1$:

$$A_{\text{prior}} = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \\ -1 & & & 1 \end{bmatrix} \implies \Sigma_{\text{prior}} = \frac{1}{4} \begin{bmatrix} 4 & 4 & 4 & 4 \\ 4 & 7 & 6 & 5 \\ 4 & 6 & 8 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix} \tag{4.63}$$

We can see that (4.61) and (4.62) reach the same values. This suggests that the specification for the gauge freedom dilemma fixing $\mathbf{x_0}$ is compatible with the CARLIT framework.

In tables 4.3 and 4.4, we test the values of the posterior marginal variances for 100 000 runs to see if a predicted portion of those realizations fall within a $1\,\sigma$ distance from the ground truth. As the tested marginal variables are scalars, this portion should be around 68.27%.

Approach	Posterior variance σ_{post}	Sample count of $\lvert\hat{\mathbf{x}}_1 - \mathbf{x}_1\rvert < \sigma_{\text{post}}$	Sample count of $\lvert\hat{\mathbf{x}}_2 - \mathbf{x}_2\rvert < \sigma_{\text{post}}$	Sample count of $\lvert\hat{\mathbf{x}}_3 - \mathbf{x}_3\rvert < \sigma_{\text{post}}$
MLE SLAM with fixed $\mathbf{x_0}$	$\sigma^2_{\mathbf{x_1}\text{ post}} = 0.75$ $\sigma^2_{\mathbf{x_2}\text{ post}} = 1$ $\sigma^2_{\mathbf{x_3}\text{ post}} = 0.75$	68281	67963	68091
MLE SLAM with prior on $\mathbf{x_0}$	$\sigma^2_{\mathbf{x_1}\text{ post}} = 1.75$ $\sigma^2_{\mathbf{x_2}\text{ post}} = 2$ $\sigma^2_{\mathbf{x_3}\text{ post}} = 1.75$	87357	84128	87181
CARLIT	$\sigma^2_{\mathbf{x_1}\text{ post}} = 0.75$ $\sigma^2_{\mathbf{x_2}\text{ post}} = 1$ $\sigma^2_{\mathbf{x_3}\text{ post}} = 0.75$	68281	67963	68091

Table 4.3: 1σ confidence interval tests for the simple 1-dimensional pose-graph SLAM experiment. The approach CARLIT is tested, along with the two MLE SLAM approach, one which fixes $\mathbf{x_0}$ and the other one which sets a prior on $\mathbf{x_0}$. Out of 100 000 samples, we count the number which falls under a $1-\sigma$ distance of the true value. Parameters: $\sigma_{\text{sm}} = \sigma_{\text{rm}} = 1$.

Approach	Posterior variance σ_{post}	Sample count of $\lvert\hat{\mathbf{x}}_1 - \mathbf{x}_1\rvert < \sigma_{\text{post}}$	Sample count of $\lvert\hat{\mathbf{x}}_2 - \mathbf{x}_2\rvert < \sigma_{\text{post}}$	Sample count of $\lvert\hat{\mathbf{x}}_3 - \mathbf{x}_3\rvert < \sigma_{\text{post}}$
MLE SLAM with fixed $\mathbf{x_0}$	$\sigma^2_{\mathbf{x_1}\text{ post}} = 0.67$ $\sigma^2_{\mathbf{x_2}\text{ post}} = 0.67$ $\sigma^2_{\mathbf{x_3}\text{ post}} = 0.01$	68299	68378	68495
MLE SLAM with prior on $\mathbf{x_0}$	$\sigma^2_{\mathbf{x_1}\text{ post}} = 1.67$ $\sigma^2_{\mathbf{x_2}\text{ post}} = 1.67$ $\sigma^2_{\mathbf{x_3}\text{ post}} = 1.01$	88709	88649	100000
CARLIT	$\sigma^2_{\mathbf{x_1}\text{ post}} = 0.67$ $\sigma^2_{\mathbf{x_2}\text{ post}} = 0.67$ $\sigma^2_{\mathbf{x_3}\text{ post}} = 0.01$	68299	68378	68495

Table 4.4: 1σ confidence interval tests for the simple 1-dimensional pose-graph SLAM experiment. Compared to table 4.3, the parameters have been set at $\sigma_{\text{rm}} = 1$ and $\sigma_{\text{sm}} = 0.1$.

For each draw, the measurements $\mathbf{z}_0^1, \mathbf{z}_1^2, \mathbf{z}_2^3, \mathbf{z}_0^3$ are produced, and the data of each draw is fed to the three tested approaches. One experiment is generated with the noise parameters $\sigma_{\text{sm}} = \sigma_{\text{rm}} = 1$ (table 4.3), and the other with $\sigma_{\text{sm}} = 1$, $\sigma\text{rm} = 0.1$ (table 4.4).

The CARLIT approach and the MLE SLAM with fixed $\mathbf{x_0}$ have exactly the same results, for their covariance matrix is the same, and the MAP estimate turns out to be the same as well. Additionally, the sample counts of falling within $1\,\sigma$ distance is about $\pm 0.3\%$ from the predicted 68.27%, across all tests. In contrast, for the MLE SLAM with a prior on $\mathbf{x_0}$ (selected to be $\sigma_{\mathbf{x_0}} = 1$), more samples than expected fall within the $1\,\sigma$ distance, suggesting that the posterior variance too conservative. This suggests that, in probabilistic SLAM, the correct specification when facing a gauge freedom dilemma is to fix the first variable $\mathbf{x_0}$.

4.7 Conclusion and Perspectives

In this chapter, we explored a new statement of SLAM via Structural Causal Models (SCMs). The motivations were given in section 4.1. The approach, named CARLIT (Causal Approach to Represent Localization and Internalized Topography), relies on a series of causal statements leading to a representation that supports the conditional independences discussed in chapter

2. Notably, causal statements lead to the modeling of loop closures as sequential sums of relative motions, rather than pairwise relations between robot poses, as described in section 4.3. Our representation and processing of landmarks is analyzed in section 4.4, and tested on a real scale experiment, with tags as landmarks. It is analyzed that, conceptually, landmarks are addressed through the viewpoint of internalized topography rather than through a world modelling. The errors $\mathcal{E}$ are the decision variables of the adjustment process (§4.3 and §4.4), which is implemented as a nonlinear constrained least squares problem. Importantly, the iterative algorithm to find the MAP estimate of the errors is bootstrapped at $\mathcal{E} = 0$. Anticipating on the complexity induced by loop closures that overlap each other, we consider structural compositions of intermediary motions to decrease the treewidth (a performance metric for the underlying adjustment procedure), in section 4.5. These compositions are done heuristically and proof of concepts are given on small scale problems. Finally, as it is important to be able to evaluate the posterior uncertainty in inference problems, an approach to answer arbitrary covariance queries is explored in section 4.6.

Although the CARLIT approach is based on different assumptions than the probabilistic view of SLAM (as per key assumptions 2.1 and 2.2, p53), no differences were so far noticed on the MAP estimates of the poses and landmarks. No definitive conclusions are proposed so far in that regard, as it is not impossible that the differences in assumptions materialize in distinct MAP estimates on some other scenarios or real case applications. However, we argue that the main interest of the approach comes from the novel mechanics that the structure can offer, some which are specific to the SLAM problem (such as the astute management of data for long term navigation through compositions), while other novel mechanics could come from the incorporation of relatively recent advances in causal inference. For instance, we briefly explore the notion of informative missing data (Mohan et al., 2013) on appendix A.

Before that, much remains to be done with the theoretical material proposed in this chapter. This includes: a more thorough study of the compositions in different scenarios, and their associated heuristics; the adaptation to incremental SLAM; the study of the algorithm in challenging maps where the initialization process is generally tricky. As mentioned in section 2.2 (p. 61), a lot has been achieved in the SLAM literature. Thus, in particular, the processing of IMU measurements should be proposed in this framework, as well as the modeling of bias for this purpose (or for calibration purposes), and the incorporation of absolute GNSS measurements. We also believe there are interesting implications in addressing the data correspondence problem in a causal framework.

Finally, on a conceptual level, this work is the product of several refinements in understanding after the previous chapter 3: (a) spatial inference problems (such as SLAM) can be stated with a full-fledged causal framework, (b) there are advantages in selecting the errors as the decision variables of the adjustment problem (though the set of errors becomes greater than the set of poses and landmarks), and, (c) most relations in the system are deterministic.

The historical analysis in chapter 1 has been instrumental in that regard.

Chapter 5

Epilogue: Quasi-Deterministic Triangulations

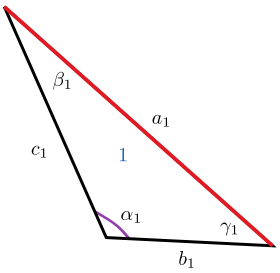


Figure 5.1: A simple triangle, with the red side denoting the baseline. To avoid overloading the next figures, we standardize the nomenclature in the displays. We will only make visible the purple angular symbol and go by the rule that it corresponds to α, and that β is the next angle in the *clockwise* direction. Additionally, a, b and c are the sides opposite to, respectively, α, β and γ.

The main body of this document begins with the study of geodetic adjustment in triangulations (chapter 1), before addressing the modern question of SLAM. In this epilogue, we shall return once more to the problem of geodetic adjustment, not in order to tackle an outdated late 18th/early 19th century problem, but instead to show that the concepts emphasized in the causal SLAM contribution (chapter 4) are well suited to address the unresolved issues noted in pre-robotics adjustment methods. By the end of chapter 1, many questions arose such as: how can we specify the assumptions geometers did when tackling their triangulations, other than verbosely ? What was the proper way of applying corrections to the observation errors (*the errors to be feared*) ? Were the formulations employed at the time correct ?

Surely, if SLAM and geodetic adjustment are problems that share many structural commonalities, then we should in turn be able to employ the causal framework described in the previous chapter to answer those questions. Hence, in this epilogue, armed with the concepts developed and the SCM framework, we provide a causal representation of the triangulation operations and an approach to adjust the measurement errors following closure events (triangle closures, baseline closures & chain closures).

Incidentally, this epilogue aims to deliver a corroboration for several important points highlighted in the preceding SLAM chapter. These points concern (a) the causal graph allowing a finer exploitation of pattern of conditional independences, (b) the relevance of the distinction between interventions and observations for the same physical process (see our treatment of landmark discovery §4.4), (c) the closure as an event that blocks already existing open colliders (see §4.2 for CARLIT), and (d) the constrained optimization approach to adjustment.

The approach is introduced in ascending order of difficulty. Section 5.1 shows the approach for one single triangulation; section 5.2 introduces it for the chain of triangles, with and without the measurement of second baseline; section 5.3 then generalizes the representation for a network of triangles. Finally, in section 5.4, we discuss various inadequacies that had prevented the development of causal inference at the time of the early metric system: although historical, we believe these last remarks are not devoid of epistemic value for the question of spatial inference in modern robotics. We limit the scope of the representations to cases of planar triangulations (i.e., no triangle-on-sphere considerations).

5.1 Representation of a single triangulation

Consider a basic geodesic operation to build the simple triangle shown figure 5.1. The process of triangulation is performed via the measurements of one baseline length and two angles adjacent to the baseline. All measurements are considered subject to errors. The goal is to represent how these measurements combine with each other so that opportunities of adjust-

ment can be identified. The challenge is to pose this problem in terms of a Structural Causal Model (SCM).

Let $\bar{a}_1$ denote the measured length of the side a_1, a.k.a. the baseline. Let τ_{a_1} be the exogenous variable capturing the measurement error. Let $\bar{\beta}_1$ and $\bar{\gamma}_1$ be the measurements of the angles β_1 and γ_1 respectively. Relating to the angular measurements errors, ϵ_{β_1} and ϵ_{γ_1} are the respective exogenous variables.

In this basic problem, the three following equations enter our SCM:

$$\begin{cases} a_1 & := \bar{a}_1 + \tau_{a_1} \\ \beta_1 & := \bar{\beta}_1 + \epsilon_{\beta_1} \\ \gamma_1 & := \bar{\gamma}_1 + \epsilon_{\gamma_1} \end{cases} \tag{5.1}$$

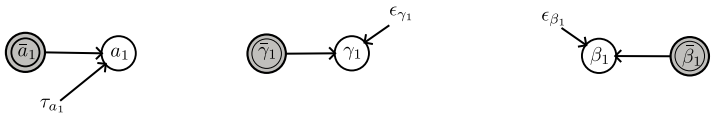


Figure 5.2: Causal graph for the basic 1-triangle geodesic operation. Note that all measurements are cast as 'interventions'. Furthermore, the errors are disconnected from one another.

The corresponding causal graph is shown figure 5.2. Firstly, note that all 3 measurements are actually cast as 'interventions', not observations. The argument here is similar to the one made with the first landmark observation in our causal SLAM (§4.4): the viewpoint of topography is adopted whereby the characteristics of the triangle are determined by the length of one side and two angles, which are set these 3 measurements.

Secondly, note the disconnection between the variables. Without more structural equations, it is impossible to establish a dependency between the three errors. Hence, at this point, nothing can be done to *adjust* the SCM. This last fact is inline with our intuition, but other probabilistic frameworks (which would consider the vertices as decision variables) would not be able to convey that useful information.

Since we have sufficient information to define the triangle as a mathematical object, we can deduce variables that have not been measured: γ_1, b_1 and c_1. This is done via deterministic trigonometric equations (the law of sines) that complete the SCM (5.1).

$$\begin{cases} a_1 & := \bar{a}_1 + \tau_{a_1} \\ \beta_1 & := \bar{\beta}_1 + \epsilon_{\beta_1} \\ \gamma_1 & := \bar{\gamma}_1 + \epsilon_{\gamma_1} \\ \alpha_1 & := \pi - \beta_1 - \gamma_1 \\ b_1 & := a_1 \frac{\sin \beta_1}{\sin(\beta_1 + \gamma_1)} \\ c_1 & := a_1 \frac{\sin \gamma_1}{\sin(\beta_1 + \gamma_1)} \end{cases} \tag{5.2}$$

The expanded causal graph is represented on figure 5.3. Note that the three deduced variables are unobserved colliders. Thus, they do not create a path of influence between the parents. The independences between the errors, as noted in figure 5.2, is still present. We have deduced more geometric variables with existing information does not create any type of closure. This is again, inline with basic intuition one can have about this problem.

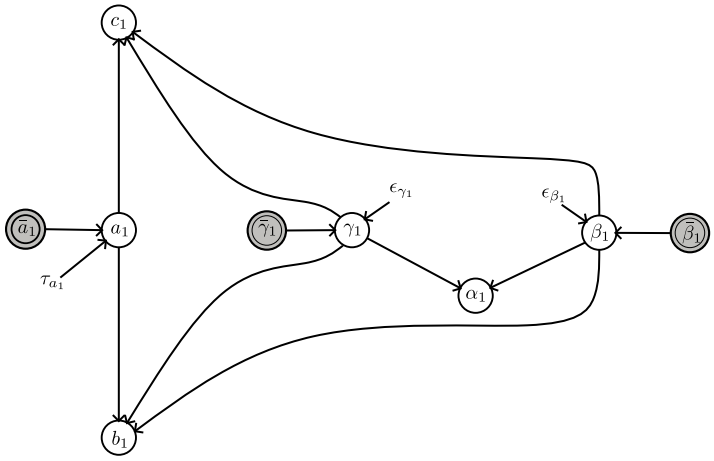


Figure 5.3: Causal graph of the basic triangle, where, compared to figure 5.2, additional geometric entities are deduced. These entities are the length b_1 and c_1. The graph is an abstraction of SCM (5.2). Since deduced the entities are open colliders, it is graphically verified that the errors $\{\tau_{a_1}, \epsilon_{\beta_1}, \epsilon_{\gamma_1}\}$ remain here independent of one another.

Suppose, as practiced by all observers since the 17th century (see chapter 1), that we measure the third angle α_1 of the triangle. In other words, we create a *triangle closure*. This implies a mismatch between the deduced value α_1 and measured value noted $\bar{\alpha}_1$ that can only be explained by the errors during the measurement process of α_1, β_1 and γ_1. Although, physically, there is no doubt that the measurement process for getting $\bar{\alpha}_1$ is the same as for $\bar{\beta}_1$ and $\bar{\gamma}_1$, the causal approach compels us to treat it as an observation rather than an intervention. The functional of the measurement is added in SCM (5.3).

$$\begin{cases} a_1 & := \bar{a_1} + \tau_{a_1} \\ \beta_1 & := \bar{\beta_1} + \epsilon_{\beta_1} \\ \gamma_1 & := \bar{\gamma_1} + \epsilon_{\gamma_1} \\ \alpha_1 & := \pi - \beta_1 - \gamma_1 \\ b_1 & := a_1 \frac{\sin \beta_1}{\sin(\beta_1 + \gamma_1)} \\ c_1 & := a_1 \frac{\sin \gamma_1}{\sin(\beta_1 + \gamma_1)} \\ \bar{\alpha}_1 & := \alpha_1 + \epsilon_{\alpha_1} \end{cases} \tag{5.3}$$

We see in figure 5.4 the causal graph of SCM (5.3). According to the rules of propagation of influence in the graph, the errors ϵ_{α_1}, ϵ_{β_1} and ϵ_{γ_1} become statistically linked. On the other hand, the error on the length of the first baseline remains independent of them. As understood by most observers at the time, and established by Laplace (1820) (see subsection 1.4.2) using probability theory, the triangle opens the opportunity to bring corrections to the angular error terms $\{\epsilon_{\alpha_1}, \epsilon_{\beta_1}, \epsilon_{\gamma_1}\}$. Furthermore, after the adjustment process, it is expected that the precision of these errors be improved (or the 'weights' as referred in the period).

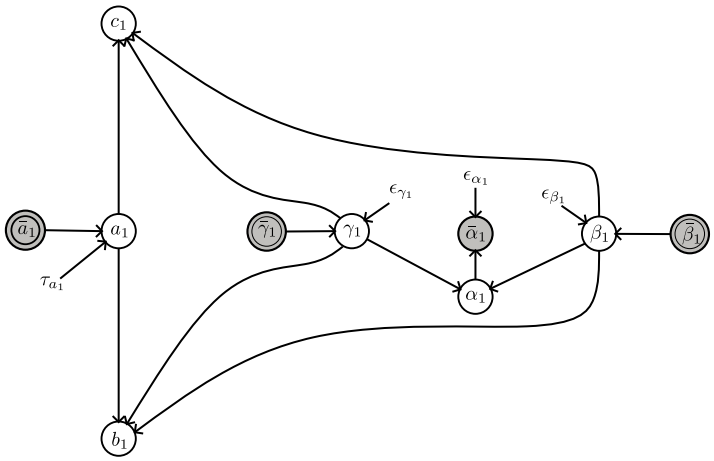


Figure 5.4: Causal graph of the basic triangulation, where all angles have been measured. Compare to the previous figure 5.1, note the addition of the closed collider on $\bar{\alpha}_1$ which creates a dependency between the errors ϵ_{α_1}, ϵ_{β_1} and ϵ_{γ_1}.

5.1.1 Triangle closure adjustment

We have established that corrections were applicable to the errors $\xi \triangleq [\epsilon_{\alpha_1},\ \epsilon_{\beta_1},\ \epsilon_{\gamma_1}]^\top$ in the SCM (5.3) following the measurement of the third angle $\bar{\alpha}_1$. We presently address the question on how to provide the corrections to the errors, i.e., how to adjust the error terms. It is easy to see that the first four equations of (5.3) lead to the formulation of the following constraint:

$$\bar{\alpha}_1 - \epsilon_{\alpha_1} = \pi - \bar{\beta}_1 - \epsilon_{\beta_1} - \bar{\gamma}_1 - \epsilon_{\gamma_1}\ . \tag{5.4}$$

Let us define, from eq (5.4), the following quantities:

$$\begin{cases} d_{\mathrm{tc},1} & \triangleq \bar{\alpha}_1 + \bar{\beta}_1 + \bar{\gamma}_1 - \pi \\ C_{\mathrm{tc},1}(\xi) & \triangleq \epsilon_{\alpha_1} - \epsilon_{\beta_1} - \epsilon_{\gamma_1}, \end{cases} \tag{5.5}$$

where subscript 'tc, 1' stands for 'triangle closure' of triangle 1.

It is assumed that the measurements errors in ξ are i.i.d. Gaussians of variance σ^2. The assumption is especially well justified with the CLT if an instrument such as the repeating circle is used (see figure 1.2). The constrained minimization problem, as in CARLIT (see §4.3) writes as:

$$\underset{\xi}{\operatorname{argmin}}\ \frac{1}{2}\left(\|\epsilon_{\alpha_1}/\sigma\|_2^2 + \|\epsilon_{\beta_1}/\sigma\|_2^2 + \|\epsilon_{\gamma_1}/\sigma\|_2^2\right) \quad \text{s.t.}\ \ C_{\mathrm{tc},1}(\xi) = d_{\mathrm{tc},1}. \tag{5.6}$$

Similarly as in the CARLIT method, we solve this problem using the Lagrangian formulation, by introducing a Lagrange parameter λ_1:

$$\mathcal{L}(\xi, \lambda_1) \triangleq \frac{1}{2}\left(\|\epsilon_{\alpha_1}/\sigma\|_2^2 + \|\epsilon_{\beta_1}/\sigma\|_2^2 + \|\epsilon_{\gamma_1}/\sigma\|_2^2\right) + \lambda_1(C_{\text{tc},1}(\xi) - d_{\text{tc},1}) \ . \tag{5.7}$$

Classically, this leads to the matrix system:

$$H \begin{bmatrix} \epsilon_{\alpha_1} \\ \epsilon_{\beta_1} \\ \epsilon_{\gamma_1} \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ d_{\text{tc},1} \end{bmatrix}, \tag{5.8}$$

where

$$H \triangleq \begin{bmatrix} 1/\sigma^2 & & & 1 \\ & 1/\sigma^2 & & -1 \\ & & 1/\sigma^2 & -1 \\ 1 & -1 & -1 & 0 \end{bmatrix} \tag{5.9}$$

Solving the system eq (5.8) is trivial since the function $C_{\text{tc},1}(\xi)$ is linear in $\xi \triangleq [\epsilon_{\alpha_1}, \epsilon_{\beta_1}, \epsilon_{\gamma_1}]^\top$.

Once the errors have been adjusted, i.e. once that system (5.8) has been solved, the others variables of the problem (angles, lengths) can be deduced easily from the SCM (5.3).

5.1.2 Comparison with Laplace's 1818 work

In §1.4.2 and §1.5.3, we investigated the Laplacian paradigm for probability theory and how his views were employed for geodetic adjustment. It consisted in first stating causal assertions about the problem, which provided the foundation to probabilistic inference.

Given that the Laplacian paradigm was an inspiration in our work, a natural question arises: is Laplace treatment of geodetic adjustment equivalent to the describes above, using contemporary causal inference tools (Pearl, 2009) ?

In the second supplement of 'Théorie Analytique des Probabilités', Laplace (1820, p537) considers the case of a triangle where the 3 angles are measured. He states that the probability of the 3 errors is (with slight modification of notations to maintain consistency with the rest of the section):

$$e^{-h\epsilon_{\alpha_1}^2 - h\epsilon_{\beta_1}^2 - h\epsilon_{\gamma_1}^2} \ , \tag{5.10}$$

with h being some notion of precision (back then referred as weight), which is linked to the more familiar concept (and recent) of variance by $h = \frac{1}{2\sigma^2}$. Thus, equation (5.10) is the joint of the measurement priors, considered independently. Then, the constraint pdf between the angles of the triangle is noted as:

$$\epsilon_{\alpha_1} + \epsilon_{\beta_1} + \epsilon_{\gamma_1} = T, \tag{5.11}$$

where T is the excess of the 3 measured angles on π. Firstly, note that, in the SCM (5.3), we express the relations between errors and measurements differently. Indeed, for angles γ_1 and β_1, the structural relation is $\gamma_1 := \bar{\gamma}_1 + \epsilon_{\gamma_1}$ and $\beta_1 := \bar{\beta}_1 + \epsilon_{\beta_1}$. On the other hand, for α_1, the relation is $\bar{\alpha}_1 := \alpha_1 + \epsilon_{\alpha_1}$ which implies that $\alpha_1 = \bar{\alpha}_1 - \epsilon_{\alpha_1}$.

This difference in the structural equations has arisen because we treat the measurements of γ_1 and β_1 as interventions which participate in the building of the triangle, while the measurement on α_1 is treated as observation of an already existing quantity (since its value can be deduced from: sum of angles equals π). This is why we have our constraint in the

form $\epsilon_{\alpha_1} - \epsilon_{\beta_1} - \epsilon_{\gamma_1} = T$ (see eq. (5.5)) rather than $\epsilon_{\alpha_1} + \epsilon_{\beta_1} + \epsilon_{\gamma_1} = T$. The distinction we do between observation/intervention for geometric problem is not made by Laplace, nor to our knowledge by anyone else: posing a different structural equation to the same physical process on the basis of qualitative causal considerations is novel[1].

Nonetheless, eq. (5.10) implicitly states that the measurements are i.i.d. Gaussian, and submitted to the constraint (5.11).

The next step is more contrived: the expression of ϵ_{γ_1} from eq (5.11) is substituted into (5.10) and rearranged, leading to

$$e^{-2h(\epsilon_{\beta_1}+\frac{1}{2}\epsilon_{\alpha_1}-\frac{1}{2}T)^2-\frac{3h}{2}(\epsilon_{\alpha_1}-\frac{1}{3}T)^2-\frac{h}{3}T^2} \ . \tag{5.12}$$

Then, the term (5.12) is integrated over ϵ_{β_1} to result in (Laplace, 1820, p537):

$$\int_{-\infty}^{+\infty} e^{-2h(\epsilon_{\beta_1}+\frac{1}{2}\epsilon_{\alpha_1}-\frac{1}{2}T)^2-\frac{3h}{2}(\epsilon_{\alpha_1}-\frac{1}{3}T)^2-\frac{h}{3}T^2} d\epsilon_{\beta_1}$$
$$= e^{-\frac{3h}{2}(\epsilon_{\alpha_1}-\frac{1}{3}T)^2-\frac{h}{3}T^2} \int_{-\infty}^{+\infty} e^{-2h(\epsilon_{\beta_1}+\frac{1}{2}\epsilon_{\alpha_1}-\frac{1}{2}T)^2} d\epsilon_{\beta_1} \ . \tag{5.13}$$

It is then stated that the probability of ϵ_{α_1} is the factor $e^{-\frac{3h}{2}(\epsilon_{\alpha_1}-\frac{1}{3}T)^2-\frac{h}{3}T^2}$, and that therefore, the most probable correction of ϵ_{α_1} is the one which sets to zero the quantity $\epsilon_{\alpha_1} - \frac{1}{3}T$.

However, this is not correct. While it is true that the most probable correction is $\hat{\epsilon}_{\alpha_1} = \frac{1}{3}T$ (an intuitive result), their remains an ϵ_{α_1} term inside the integral. This case highlights the difficulty of working with system of multivariate errors, without notions of correlations to rely upon (a concept, like covariances etc..., that would be introduced much later).

Consequently, Laplace deduces from his factor that $\frac{3h}{2}$ is the *a posteriori* weight. And since $\frac{3h}{2} > h$, the intended purpose was to show that the weight of the result rises after the correction. The fact that the precision of the error is improved a posteriori should also be intuitively true, but by how much ? Is the value given by Laplace correct, or under/overconfident ?

Instead, let us determine the covariance matrix from the SCM framework using the Dirac technique seen in the previous chapter. For comparison with the substitution, we also Dirac-eliminate ϵ_{γ_1} The joint probability distribution inside the triangle is:

$$p(\xi|\bar{\alpha}_1, \mathrm{do}(\bar{\beta}_1), \mathrm{do}(\bar{\gamma}_1)) \propto$$
$$\exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon_{\alpha_1}^2 + \epsilon_{\beta_1}^2 + \epsilon_{\gamma_1}^2\right)\right\} \ \delta\left(\epsilon_{\gamma_1} - \left(\pi - \bar{\beta}_1 - \epsilon_{\beta_1} - \bar{\gamma}_1 - \bar{\alpha}_1 + \epsilon_{\alpha_1}\right)\right) \tag{5.14}$$

Then, ϵ_{γ_1} is integrated out to remove the Dirac term δ. This results in:

$$p(\epsilon_{\beta_1}, \epsilon_{\alpha_1}|\bar{\alpha}_1, \mathrm{do}(\bar{\beta}_1), \mathrm{do}(\bar{\gamma}_1)) \propto$$
$$\exp\left\{-\frac{1}{2\sigma^2}\left(\epsilon_{\alpha_1}^2 + \epsilon_{\beta_1}^2 + \left(\pi - \bar{\beta}_1 - \epsilon_{\beta_1} - \bar{\gamma}_1 - \bar{\alpha}_1 + \epsilon_{\alpha_1}\right)^2\right)\right\}. \tag{5.15}$$

[1]The same logic was applied for landmarks in §4.4, whereby the structural equation is not the same if the landmark is discovered, or re-observed.

The covariance is retrieved via the identification of the form $\exp\left\{-\frac{1}{2}\left\|A\begin{bmatrix}\epsilon_{\alpha_1}\\ \epsilon_{\beta_1}\end{bmatrix}-b\right\|_2^2\right\}$, where A and b are commonly known as measurement and measurement matrix and vector. The identification leads to:

$$A \triangleq \frac{1}{\sigma}\begin{bmatrix}1 & \\ & 1\\ 1 & -1\end{bmatrix}, \quad b \triangleq \begin{bmatrix}0\\ 0\\ \bar{\gamma}_1+\bar{\alpha}_1+\bar{\beta}_1-\pi\end{bmatrix}. \tag{5.16}$$

Thus,

$$\Sigma \triangleq (A^\top A)^{-1} = \frac{\sigma^2}{3}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix} = \frac{1}{6h}\begin{bmatrix}2 & 1\\ 1 & 2\end{bmatrix}. \tag{5.17}$$

The marginal weight of ϵ_{α_1} is thus in fact $\left(\Sigma_{(1,1)}\right)^{-1} = \frac{3}{2}h$, which is the same as Laplace. However, since Laplace shunned the correlation between the errors, the equivalent covariance used in the next phase of the reasoning should be (by extrapolating the logic) $\Sigma_{Laplace} \triangleq \frac{1}{3h}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$. Therefore, since the determinants have $|\Sigma| < |\Sigma_{Laplace}|$, the weight of the result does improve following a triangle closure adjustment but his result was underconfident when the errors $\epsilon_{\alpha_1}, \epsilon_{\beta_1}$ are considered jointly. Next, we will see how it influences the result on the adjustment of a chain of adjacent triangles.

5.2 Representation of a chain of triangles

Having been able to fully formulate a basic triangulation (one triangle) in the SCM framework, we now seek to represent a chain of triangle. Historically, the chain of triangles was in geodesy the mathematical object to build in order to measure the arc-length of a meridian, or a parallel. An iconic example, that we have covered in chapter 1 is the 1792-1799 Dunkerque-Paris-Barcelona meridian.

5.2.1 Open chain of triangles

Consider the chain of three triangles on figure 5.5. Triangle (1) is the same as in the previous section. Triangle (2) is built adjacently to triangle (1) (we have two notations, b_1 and a_2, covering the same variable depending on which triangle is considered). Triangle (3) is built in turn adjacently to the second triangle. Consider directly a scenario where the observers have measured all angles of all triangles.

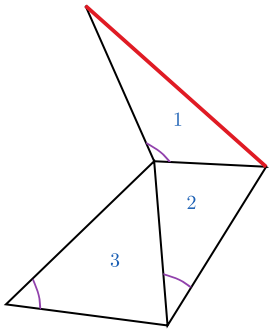


Figure 5.5: Small chain of triangles.

The following SCM (5.18) describes the system:

$$
\begin{cases}
& \text{First baseline:} \\
a_1 & := \bar{a_1} + \tau_{a_1} \\
& \text{First triangle:} \\
\beta_1 & := \bar{\beta_1} + \epsilon_{\beta_1} \\
\gamma_1 & := \bar{\gamma_1} + \epsilon_{\gamma_1} \\
\alpha_1 & := \pi - \beta_1 - \gamma_1 \\
\bar{\alpha}_1 & := \alpha_1 + \epsilon_{\alpha_1} \\
b_1 & := a_1 \frac{\sin \beta_1}{\sin(\beta_1 + \gamma_1)} \\
& \text{Second triangle:} \\
\beta_2 & := \bar{\beta_2} + \epsilon_{\beta_2} \\
\gamma_2 & := \bar{\gamma_2} + \epsilon_{\gamma_2} \\
\alpha_2 & := \pi - \beta_2 - \gamma_2 \\
\bar{\alpha}_2 & := \alpha_2 + \epsilon_{\alpha_2} \\
c_2 & := a_2 \frac{\sin \gamma_2}{\sin(\beta_2 + \gamma_2)} \\
& \text{Third triangle:} \\
\beta_3 & := \bar{\beta_3} + \epsilon_{\beta_3} \\
\gamma_3 & := \bar{\gamma_3} + \epsilon_{\gamma_3} \\
\alpha_3 & := \pi - \beta_3 - \gamma_3 \\
\bar{\alpha}_3 & := \alpha_3 + \epsilon_{\alpha_3} \\
c_3 & := a_3 \frac{\sin \gamma_3}{\sin(\beta_3 + \gamma_3)}
\end{cases}
\tag{5.18}
$$

and figure 5.6 plots the corresponding causal graph. For convenience, we do not show the variables c_1, b_2, b_3 which would all be graphical colliders (if they were to be explicitly represented). The same remark that was made in CARLIT (chapter 4) can be repeated here: when the problem size grows, the number of variables that can optionally be plotted becomes large. In the case of CARLIT, these were all possible i, j combinations of relative motions $\mathbf{\Lambda_i^j}$. But since those variables are all *open* colliders, unless the small subset of them that are observed, they do not create any additional influence path between errors (and thus no additional simultaneous corrections). They can therefore not be explicitly shown in order to declutter the graphs.

Previously in this chapter 5, we have shown that the pattern of conditional independences is compatible with our (human) intuition. As the size of the triangulation problem grows, it becomes progressively more difficult to maintain clairvoyance on which parts of the triangulation are (in)dependent from one another[2]. Causal graphs offer an algorithmic answer to that question thanks to d-separation. But arguably, even without an algorithmic query, it is easy to read directly on the graph which areas of the graph are d-separated. In figure 5.6, we can see that a correction from a triangle closure does not 'influence' other triangles in view of the open colliders in b_1, c_2 and c_3. This can be found evidently without a causal graph by qualitative considerations (we highlighted this in chapter 1), but as the size and

[2]One might argue however, that this type of human intuition is not guaranteed in small problems either. These problems are sometimes referred as *paradoxes* in statistics. For an analysis of them through the lens of causal inference, see Pearl and Mackenzie (2020), in particular, the Monty Hall problem.

complexity grows, the causal graph becomes the most reliable tool to reveal those conditional independences.

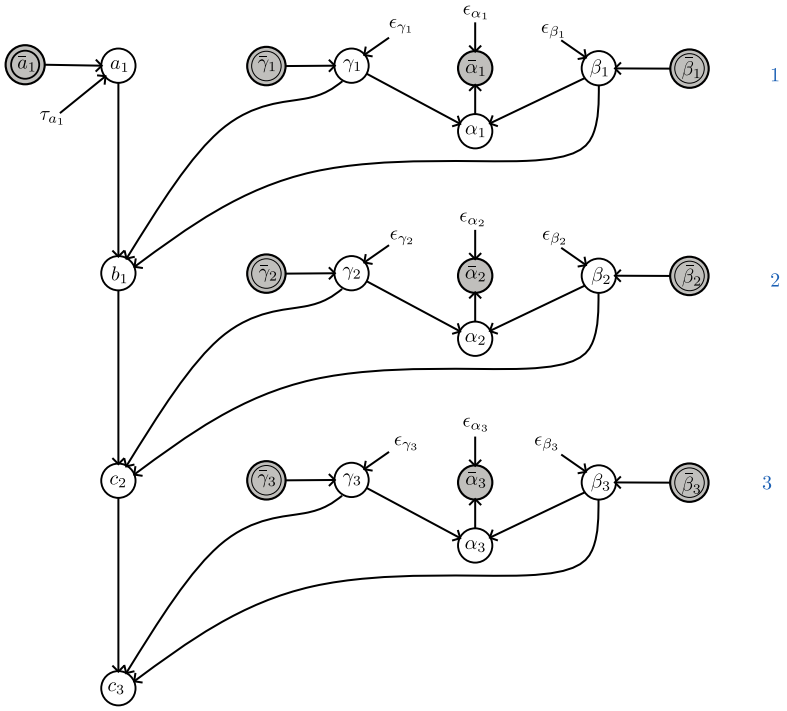


Figure 5.6: Causal graph of the SCM 5.3 corresponding to the chain of triangles on figure 5.5.

Moreover, observe in figure 5.6 that b_1 is affected by the errors τ_{a_1}, ϵ_{γ_1}, ϵ_{α_1} and ϵ_{β_1}. In turn, c_2 is affected by the same errors affecting b_1 plus new errors: ϵ_{γ_2}, ϵ_{α_2} and ϵ_{β_2}. In turn, c_3 is affected by the errors of c_2 and new errors, etc... The causal graph is able to visually convey the notion of *drifting*: the sides of the triangles become less and less precise as the number of triangles in the chain grows. We mentioned in section 1.4.2 that Laplace also identified this notion of drifting in a chain of triangles, which motivated the need to measure a side length (a.k.a. a second baseline) at the extremity of the chain. The causal graph shows indeed that, if c_3 were to be given, new statistical links would exist between all the errors. These insights are available to us thus far without any data.

Remark 5.1. *From our experience, it is essential to stress that obtaining a model exhibiting these insights is nearly only possible if the graph is constructed causally, and the designer follows a Laplacian quasi-deterministic mindset. Indeed, while other probabilistic graphical model frameworks, like the Bayesian Networks (rung 1 in the PCH), or even the Causal Bayesian Networks (rung 2 in the PCH) can in theory deliver as many insightful conditional independences as a causal graph does, we have not seen much applications of them recently in SLAM. The primary reason for this limited adoption is that although they can conceivably reproduce the conditional independences of a causal graph at the associational level, there are still conceptual problems. As graphical representations of joint probability distributions, the BN and CBN frameworks promote a probabilistic approach. In this view, writing a product of, e.g., Gaussians and Dirac distributions (as a cheap proxy for deterministic relations) does*

not bring forward immediately a clear method to conduct inference, i.e. there are no general efficient solver to rely upon. On the other hand, the designer that employs a BN or CBN (or an undirected representation, but that is besides the point here) is implicitly encouraged to treat all available data as measurement to form a product of likelihood functions (or product of factors), without reasoning about the opportunity of distinguishing between observation and intervention (like we in contrast do in, both for CARLIT and causal triangulations). This is because the product of likelihood leads to a seducing and easy way to formulate least squares optimization problem; this is known since Gauss (1809), see equation (1.1). This may be the main reason why approaching a problem containing both stochastic and deterministic components (e.g. spatial inference problems) with the willingness to (a) maximally exploit domain knowledge and (b) use a strictly probabilistic framework is not impossible, but a risky proposition. In contrast, the SCM offers no conceptual resistance when incorporating completely deterministic functionals. And as it happens, models in robotics are inherently largely populated by those deterministic functions.

This remark is orthogonal to the other possibilities that SCM offers over probabilistic frameworks: contemplated actions via the do-operator and counterfactual queries. We merely point out that, even if the problem concerns 'only' the inference of measurement error terms (i.e., a seemingly rung 1 query), our takeaway message is that the SCM framework should still be preferred. Furthermore, as we have shown both with landmark SLAM problems and triangulations, the appropriate structure for spatial inference is actually of rung 2 of the PCH because of the necessity to use interventions (do-operators and double circles in the causal graph).

5.2.2 Baseline closure

Consider now that a second baseline is measured at c_3. Keeping similar notations as above, the new functional $\bar{c}_3 := c_3 + \tau_{c_3}$ enters our previous SCM, which becomes:

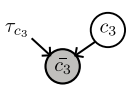


Figure 5.7: Representation of a second baseline measurement (baseline closure) at side c_3, to be added to figure 5.6.

$$
\begin{cases}
& \text{First baseline:} \\
a_1 & := \bar{a_1} + \tau_{a_1} \\
& \text{First triangle:} \\
\beta_1 & := \bar{\beta_1} + \epsilon_{\beta_1} \\
\gamma_1 & := \bar{\gamma_1} + \epsilon_{\gamma_1} \\
\alpha_1 & := \pi - \beta_1 - \gamma_1 \\
\bar{\alpha}_1 & := \alpha_1 + \epsilon_{\alpha_1} \\
b_1 & := a_1 \frac{\sin \beta_1}{\sin(\beta_1+\gamma_1)} \\
& \text{Second triangle:} \\
\beta_2 & := \bar{\beta_2} + \epsilon_{\beta_2} \\
\gamma_2 & := \bar{\gamma_2} + \epsilon_{\gamma_2} \\
\alpha_2 & := \pi - \beta_2 - \gamma_2 \\
\bar{\alpha}_2 & := \alpha_2 + \epsilon_{\alpha_2} \\
c_2 & := a_2 \frac{\sin \gamma_2}{\sin(\beta_2+\gamma_2)} \\
& \text{Third triangle:} \\
\beta_3 & := \bar{\beta_3} + \epsilon_{\beta_3} \\
\gamma_3 & := \bar{\gamma_3} + \epsilon_{\gamma_3} \\
\alpha_3 & := \pi - \beta_3 - \gamma_3 \\
\bar{\alpha}_3 & := \alpha_3 + \epsilon_{\alpha_3} \\
c_3 & := a_3 \frac{\sin \gamma_3}{\sin(\beta_3+\gamma_3)} \\
& \text{Baseline closure at } c_3\text{:} \\
\bar{c}_3 & := c_3 + \tau_{c_3}
\end{cases}
, \qquad (5.19)
$$

The figure 5.6 can be completed by binding a closed collider to c_3 (figure 5.7).

Graphically, we can deduce from the closed collider $\bar{c_3}$ that new dependencies are created between the errors $\xi = \{\tau_{a_1}, \tau_{c_3}, \epsilon_{\beta_1}, \epsilon_{\gamma_1}\epsilon_{\beta_2}, \epsilon_{\gamma_2}\epsilon_{\beta_2}, \epsilon_{\gamma_2}\}$. In terms of new corrections, this new pathway of influence can be translated to a new constraint and added in the solver.

We now show how to infer the most probable corrections to the error following this baseline closure.

Similarly as in §5.1.1 (and inspired by CARLIT §4.3), we pose a constrained minimization problem aiming at "restoring harmony":

$$
\underset{\xi}{\operatorname{argmin}} \frac{1}{2} \left(\|\tau_{a_1}/\sigma_l\|_2^2 + \|\tau_{c_3}/\sigma_l\|_2^2 + \sum_{i=1}^{3} \left(\|\epsilon_{\alpha_i}/\sigma\|_2^2 + \|\epsilon_{\beta_i}/\sigma\|_2^2 + \|\epsilon_{\gamma_i}/\sigma\|_2^2 \right) \right)
$$
$$
\text{s.t.} \quad C(\xi) = d, \qquad (5.20)
$$

where it is assumed that the baseline measurement errors τ_{a_1} and τ_{c_3} are distributed according to $\mathcal{N}(0, \sigma_l^2)$, and that the angular measurement errors are distributed according to $\mathcal{N}(0, \sigma^2)$. $C(\xi) = d$ are the constraints which arise from the closures.

Let us first list the constraints between the errors. The category of constraints already dealt with in §5.1.1 are the triangle closures. There are 3 of such constraints in the SCM

(5.19), i.e. one for each triangle:

$$\begin{cases} \epsilon_{\alpha_1} - \epsilon_{\beta_1} - \epsilon_{\gamma_1} &= \bar{\alpha}_1 + \bar{\beta}_1 + \bar{\gamma}_1 - \pi \\ \epsilon_{\alpha_2} - \epsilon_{\beta_2} - \epsilon_{\gamma_2} &= \bar{\alpha}_2 + \bar{\beta}_2 + \bar{\gamma}_2 - \pi \\ \epsilon_{\alpha_3} - \epsilon_{\beta_3} - \epsilon_{\gamma_3} &= \bar{\alpha}_3 + \bar{\beta}_3 + \bar{\gamma}_3 - \pi \end{cases} . \tag{5.21}$$

The constraint which concerns the baseline closure writes as:

$$(\bar{a}_1 + \tau_{a_1}) \frac{\sin(\bar{\beta}_1 + \epsilon_{\beta_1})}{\sin(\bar{\beta}_1 + \bar{\gamma}_1 + \epsilon_{\beta_1} + \epsilon_{\gamma_1})} \frac{\sin(\bar{\gamma}_2 + \epsilon_{\gamma_2})}{\sin(\bar{\beta}_2 + \bar{\gamma}_2 + \epsilon_{\beta_2} + \epsilon_{\gamma_2})} \frac{\sin(\bar{\gamma}_3 + \epsilon_{\gamma_3})}{\sin(\bar{\beta}_3 + \bar{\gamma}_3 + \epsilon_{\beta_3} + \epsilon_{\gamma_3})} + \tau_{c_3} = \bar{c}_3. \tag{5.22}$$

$$\begin{cases} C_{\text{bc}}(\xi) &= (\bar{a}_1 + \tau_{a_1}) \frac{\sin(\bar{\beta}_1 + \epsilon_{\beta_1})}{\sin(\bar{\beta}_1 + \bar{\gamma}_1 + \epsilon_{\beta_1} + \epsilon_{\gamma_1})} \frac{\sin(\bar{\gamma}_2 + \epsilon_{\gamma_2})}{\sin(\bar{\beta}_2 + \bar{\gamma}_2 + \epsilon_{\beta_2} + \epsilon_{\gamma_2})} \frac{\sin(\bar{\gamma}_3 + \epsilon_{\gamma_3})}{\sin(\bar{\beta}_3 + \bar{\gamma}_3 + \epsilon_{\beta_3} + \epsilon_{\gamma_3})} + \tau_{c_3} \\ d_{\text{bc}} &= \bar{c}_3 \ . \end{cases} \tag{5.23}$$

Equations (5.22) and (5.21) constitute the components of $C(\xi) = d$ in the minimization problem (5.20).

Unfortunately, due to the form of the constraint (5.22), the system cannot be solved directly but rather by succesive linearizations. As in CARLIT, the natural initialization point is $\tilde{\xi} = 0$.

We thus solve for $\{\delta\xi, \lambda\}$ around $\tilde{\xi}$ (null errors). The Lagrangian writes as:

$$\mathcal{L}(\delta\xi, \lambda) = \frac{1}{2} \left\| A\delta\xi - \tilde{b} \right\|_2^2 + \lambda^\top \left(\left. \frac{\partial C}{\partial \xi} \right|_{\tilde{\xi}} \delta\xi - \tilde{d} \right) , \tag{5.24}$$

where

$$\begin{cases} A &\triangleq \text{Diag}\left([1/\sigma_l, 1/\sigma_l, 1/\sigma, \ldots, 1/\sigma]\right) \\ \tilde{b} &\triangleq -\text{Diag}\left([1/\sigma_l, 1/\sigma_l, 1/\sigma, \ldots, 1/\sigma]\right) \tilde{\xi} \\ \tilde{d} &\triangleq d - C(\tilde{\xi}) \end{cases} . \tag{5.25}$$

Define the matrix $C_{\tilde{\xi}} \triangleq \left. \frac{\partial C}{\partial \xi} \right|_{\tilde{\xi}}$ as the Jacobian matrix of the constraints:

$$C(\tilde{\xi} + \delta\xi) \approx C(\tilde{\xi}) + C_{\tilde{\xi}}\, \delta\xi \ . \tag{5.26}$$

In this present case study, $C_{\tilde{\xi}} \in \mathbb{R}^{4\times 11}$ and is expressed as:

$$C_{\tilde{\xi}} = \begin{bmatrix} & & 1 & -1 & -1 & & & & & & \\ & & & & & 1 & -1 & -1 & & & \\ & & & & & & & & 1 & -1 & -1 \\ J_{\tilde{\tau}_{a_1}} & 1 & 0 & J_{\tilde{\epsilon}_{\beta_1}} & J_{\tilde{\epsilon}_{\gamma_1}} & 0 & J_{\tilde{\epsilon}_{\beta_2}} & J_{\tilde{\epsilon}_{\gamma_2}} & 0 & J_{\tilde{\epsilon}_{\beta_3}} & J_{\tilde{\epsilon}_{\gamma_3}} \end{bmatrix} , \tag{5.27}$$

where the column order in $C_{\tilde{\xi}}$ is $[\delta\tau_{a_1}, \delta\tau_{c_3}, \delta\epsilon_{\alpha_1}, \delta\epsilon_{\beta_1}, \delta\epsilon_{\gamma_1}, \delta\epsilon_{\alpha_2}, \delta\epsilon_{\beta_2}, \delta\epsilon_{\gamma_2}, \delta\epsilon_{\alpha_3}, \delta\epsilon_{\beta_3}, \delta\epsilon_{\gamma_3}]$ and where the J terms are:

$$\begin{cases} J_{\tilde{\tau}_{a_1}} & = \frac{1}{\bar{a}_1+\tilde{\tau}_{a_1}} \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \\ & \text{Errors at the numerators of sin functions in } C_{\mathrm{bc}}\text{:} \\ J_{\tilde{\epsilon}_{\beta_1}} & = \frac{\sin(\bar{\gamma}_1+\tilde{\epsilon}_{\gamma_1})}{\sin(\bar{\beta}_1+\tilde{\epsilon}_{\beta_1})\sin(\bar{\beta}_1+\tilde{\epsilon}_{\beta_1}+\bar{\gamma}_1+\tilde{\epsilon}_{\gamma_1})} \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \\ J_{\tilde{\epsilon}_{\gamma_2}} & = \frac{\sin(\bar{\beta}_2+\tilde{\epsilon}_{\beta_2})}{\sin(\bar{\gamma}_2+\tilde{\epsilon}_{\gamma_2})\sin(\bar{\beta}_2+\tilde{\epsilon}_{\beta_2}+\bar{\gamma}_2+\tilde{\epsilon}_{\gamma_2})} \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \\ J_{\tilde{\epsilon}_{\gamma_3}} & = \frac{\sin(\bar{\beta}_3+\tilde{\epsilon}_{\beta_3})}{\sin(\bar{\gamma}_3+\tilde{\epsilon}_{\gamma_3})\sin(\bar{\beta}_3+\tilde{\epsilon}_{\beta_3}+\bar{\gamma}_3+\tilde{\epsilon}_{\gamma_3})} \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \\ & \text{Errors exclusively at the denominator of } C_{\mathrm{bc}}\text{:} \\ J_{\tilde{\epsilon}_{\gamma_1}} & = -\mathrm{cotan}(\bar{\beta}_1 + \tilde{\epsilon}_{\beta_1} + \bar{\gamma}_1 + \tilde{\epsilon}_{\gamma_1}) \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \\ J_{\tilde{\epsilon}_{\beta_2}} & = -\mathrm{cotan}(\bar{\beta}_2 + \tilde{\epsilon}_{\beta_2} + \bar{\gamma}_2 + \tilde{\epsilon}_{\gamma_2}) \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \\ J_{\tilde{\epsilon}_{\beta_3}} & = -\mathrm{cotan}(\bar{\beta}_3 + \tilde{\epsilon}_{\beta_3} + \bar{\gamma}_3 + \tilde{\epsilon}_{\gamma_3}) \left(C_{\mathrm{bc}}(\tilde{\xi}) - \tilde{\tau}_{c_3}\right) \end{cases} . \tag{5.28}$$

The $\tilde{d}$ term in the Lagrangian (5.24) is deduced:

$$\tilde{d} \triangleq \begin{bmatrix} \bar{\alpha}_1 - \bar{\beta}_1 - \bar{\gamma}_1 - \pi \\ \bar{\alpha}_2 - \bar{\beta}_2 - \bar{\gamma}_2 - \pi \\ \bar{\alpha}_3 - \bar{\beta}_3 - \bar{\gamma}_3 - \pi \\ \bar{c}_3 \end{bmatrix} - C(\tilde{\xi}) \ . \tag{5.29}$$

The saddle point $\left\{\widehat{\delta\xi}, \hat{\lambda}\right\}$ for the linearized system around $\tilde{\xi}$ is sought for:

$$\begin{bmatrix} \widehat{\delta\xi} \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} A^\top A & C_{\tilde{\xi}}^\top \\ C_{\tilde{\xi}} & 0 \end{bmatrix}^{-1} \begin{bmatrix} A^\top \tilde{b} \\ \tilde{d} \end{bmatrix} \ . \tag{5.30}$$

As in CARLIT (§4.3), the system (5.30) is solved, iteratively, by updating the value of $C_{\tilde{\xi}}, \tilde{d}, \tilde{b}$ until some stopping condition is achieved (e.g., $\left\|\widehat{\delta\xi}\right\|_2$ converges). Algorithm 6 summarizes the process, which mostly echoes the Algorithm 3 we proposed for causal SLAM.

Covariance Query

In the previous paragraph, we have determined a locally optimal adjustment of the errors $\hat{\xi}$. In this paragraph, we propose to establish the variance of the length c_3 (the second baseline), and we investigate how this quantity evolves with or without the second baseline measurement.

According to SCM (5.19), we can recover $\hat{c}_3$ by using a function of the errors, noted $f(\xi_{c_3})$. $\xi_{c_3} \subset \xi$ is the set of errors which intervene in f. Besides, c_3 is given at the adjusted value of the errors $\hat{\xi}_{c_3}$.

$$\begin{aligned} c_3 &= f(\hat{\xi}_{c_3}) \\ c_3 &= (\bar{a_1} + \hat{\tau}_{a_1}) \frac{\sin(\bar{\beta}_1 + \hat{\epsilon}_{\beta_1})}{\sin\left(\bar{\beta}_1 + \bar{\gamma}_1 + \hat{\epsilon}_{\beta_1} + \hat{\epsilon}_{\gamma_1}\right)} \frac{\sin(\bar{\gamma}_2 + \hat{\epsilon}_{\gamma_2})}{\sin\left(\bar{\beta}_2 + \bar{\gamma}_2 + \hat{\epsilon}_{\beta_2} + \hat{\epsilon}_{\gamma_2}\right)} \frac{\sin(\bar{\gamma}_3 + \hat{\epsilon}_{\gamma_3})}{\sin\left(\bar{\beta}_3 + \bar{\gamma}_3 + \hat{\epsilon}_{\beta_3} + \hat{\epsilon}_{\gamma_3}\right)} \end{aligned} \tag{5.31}$$

Algorithm 6: Iterative Adjustment of a Chain of Triangles.

Input: Normed matrix A, vector d & functional of constraints $C(\xi)$
Output: Most probable adjustment for the errors $\hat{\xi}$

```
/* Initialize the linearization point to null errors                    */
ξ̃ ← 0;
while !StopCondition() do
    // Determine the necessary quantities from eq (5.25), (5.27)
    d̃ ← d − C(ξ̃);
    C_ξ̃ ← ∂C/∂ξ |_ξ̃;
    b̃ ← −Diag([1/σ_l, 1/σ_l, 1/σ, ..., 1/σ]) ξ̃ ;
    // Apply eq (5.30)
    {δξ̂, λ̂} ← SaddlePoint(A, C_ξ̃, b̃, d̃);
    // Update the linearization point of the errors
    ξ̃ ← ξ̃ + δξ̂;
end
ξ̂ ← ξ̃;
return ξ̂
```

The variance associated to $\hat{c}_3$, in the neighborhood of the corrections $\hat{\xi}$, is approximated by:

$$\sigma_{c_3}^2 = \left.\frac{\partial f}{\partial \xi_{c_3}}\right|_{\hat{\xi}_{c_3}} \Sigma_{\xi_{c_3}} \left.\frac{\partial f}{\partial \xi_{c_3}}\right|_{\hat{\xi}_{c_3}}^{\top} . \tag{5.32}$$

The expression of the partial derivatives of f can be found by recycling the derivative of the constraint C_{bc}, see equations (5.28). The next step is to determine the covariance matrix $\Sigma_{\xi_{c_3}}$ of the estimation error $\xi_{c_3} - \hat{\xi}_{c_3}$. This covariance is not the same depending on which case we are considering: with or without the measurement of baseline closure. As we have already shown how to establish the covariance of two angles measurement errors within a triangle, we show how to deal with several triangles, and baseline measurements. Again, similarities with CARLIT in the philosophy of the approach may be noticed (see §4.6).

Consider firstly the case when no second baseline is measured. In this situation, the covariance $\Sigma_{\xi_{c_3}}$ is easily established by noticing the d-separation in the graph (figure 5.6) between 4 groups of errors: (a) the measurement error on the first baseline τ_{a_1}, and (b, c, d) the measurement errors $\epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma$. This tells us that the covariance is block-diagonal. In each distribution which concerns the errors $\epsilon_\alpha, \epsilon_\beta, \epsilon_\gamma$ of one of the triangles, we repeat the sequence of steps from equations (5.14) to (5.17). The difference from the previous step is that instead of δ-eliminating ϵ_γ from (5.14) to (5.15), we integrate out ϵ_α terms.

This yields:

$$A_{(\beta_1,\gamma_1)} = \frac{1}{\sigma}\begin{bmatrix} 1 & \\ & 1 \\ 1 & 1 \end{bmatrix} \implies \Sigma_{(\beta_1,\gamma_1)} = \Sigma_{(\beta_2,\gamma_2)} = \Sigma_{(\beta_3,\gamma_3)} = \frac{\sigma^2}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \tag{5.33}$$

and the block-diagonal covariance, in the no-baseline-closure case:

$$\Sigma_{\xi_{c_3}} = \begin{bmatrix} \sigma_l^2 & & & & & & \\ & \frac{\sigma^2}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} & & & & & \\ & & \frac{\sigma^2}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} & & & & \\ & & & \frac{\sigma^2}{3}\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} & & & \end{bmatrix}, \tag{5.34}$$

where the variable ordering in the matrix is $\mathcal{O} = [\tau_{a_1}, \epsilon_{\beta_1}, \epsilon_{\gamma_1}, \epsilon_{\beta_2}, \epsilon_{\gamma_2}, \epsilon_{\beta_3}, \epsilon_{\gamma_3}]$.

Secondly, consider now the situation where a second baseline has been measured in c_3 (baseline closure). The error on the new measurement has a prior $p(\tau_{c_3})$ and is linked to the other errors in ξ by the constraint $C_{\text{bc}}(\xi) = d_{\text{bc}}$, of eq (5.22). This can be represented by the product:

$$p(\tau_{c_3})\delta\left(\tau_{c_3} - \left(\bar{c}_3 - (\bar{a}_1 + \tau_{a_1})\frac{\sin(\bar{\beta}_1 + \epsilon_{\beta_1})}{\sin\left(\bar{\beta}_1 + \bar{\gamma}_1 + \epsilon_{\beta_1} + \epsilon_{\gamma_1}\right)}\frac{\sin(\bar{\gamma}_2 + \epsilon_{\gamma_2})}{\sin\left(\bar{\beta}_2 + \bar{\gamma}_2 + \epsilon_{\beta_2} + \epsilon_{\gamma_2}\right)}\frac{\sin(\bar{\gamma}_3 + \epsilon_{\gamma_3})}{\sin\left(\bar{\beta}_3 + \bar{\gamma}_3 + \epsilon_{\beta_3} + \epsilon_{\gamma_3}\right)}\right)\right). \tag{5.35}$$

Next, τ_{c_3} is δ-eliminated. We can write the joint distribution of the remaining errors:

$$\begin{aligned} p(\xi_{c_3}|\dots) = p(\tau_{a_1})p(\epsilon_{\beta_1}, \epsilon_{\gamma_1}|\dots)p(\epsilon_{\beta_2}, \epsilon_{\gamma_2}|\dots)p(\epsilon_{\beta_3}, \epsilon_{\gamma_3}|\dots) \\ \times p\left(\bar{c}_3 - (\bar{a}_1 + \tau_{a_1})\frac{\sin(\bar{\beta}_1 + \epsilon_{\beta_1})}{\sin\left(\bar{\beta}_1 + \bar{\gamma}_1 + \epsilon_{\beta_1} + \epsilon_{\gamma_1}\right)}\frac{\sin(\bar{\gamma}_2 + \epsilon_{\gamma_2})}{\sin\left(\bar{\beta}_2 + \bar{\gamma}_2 + \epsilon_{\beta_2} + \epsilon_{\gamma_2}\right)}\frac{\sin(\bar{\gamma}_3 + \epsilon_{\gamma_3})}{\sin\left(\bar{\beta}_3 + \bar{\gamma}_3 + \epsilon_{\beta_3} + \epsilon_{\gamma_3}\right)}\right). \end{aligned} \tag{5.36}$$

All terms in eq (5.36) are Gaussian. Since there are no more Dirac expressions (which we used to mimicked deterministic relations), we can proceed to determine a non-singular covariance matrix. First, establish the measurement matrix $A_{\tilde{\xi}_{c_3}}$ around the linearization point $\tilde{\xi}_{c_3}$:

$$A_{\tilde{\xi}_{c_3}} = \begin{bmatrix} 1/\sigma_l & & & & & & \\ & 1/\sigma & & & & & \\ & & 1/\sigma & & & & \\ & 1/\sigma & 1/\sigma & & & & \\ & & & 1/\sigma & & & \\ & & & & 1/\sigma & & \\ & & & 1/\sigma & 1/\sigma & & \\ & & & & & 1/\sigma & \\ & & & & & & 1/\sigma \\ & & & & & 1/\sigma & 1/\sigma \\ A_{\tilde{\sigma}_{a_1}} & A_{\tilde{\epsilon}_{\beta_1}} & A_{\tilde{\epsilon}_{\gamma_1}} & A_{\tilde{\epsilon}_{\beta_2}} & A_{\tilde{\epsilon}_{\gamma_2}} & A_{\tilde{\epsilon}_{\beta_3}} & A_{\tilde{\epsilon}_{\gamma_3}} \end{bmatrix}, \tag{5.37}$$

where

$$\begin{cases} h(\tilde{\xi}_{c_3}) & \triangleq (\bar{a}_1 + \tilde{\tau}_{a_1}) \frac{\sin(\bar{\beta}_1 + \tilde{\epsilon}_{\beta_1})}{\sin(\bar{\beta}_1 + \bar{\gamma}_1 + \tilde{\epsilon}_{\beta_1} + \tilde{\epsilon}_{\gamma_1})} \frac{\sin(\bar{\gamma}_2 + \tilde{\epsilon}_{\gamma_2})}{\sin(\bar{\beta}_2 + \bar{\gamma}_2 + \tilde{\epsilon}_{\beta_2} + \tilde{\epsilon}_{\gamma_2})} \frac{\sin(\bar{\gamma}_3 + \tilde{\epsilon}_{\gamma_3})}{\sin(\bar{\beta}_3 + \bar{\gamma}_3 + \tilde{\epsilon}_{\beta_3} + \tilde{\epsilon}_{\gamma_3})} \\ A_{\tilde{\tau}_{a_1}} & = -\frac{1}{\sigma_l} \frac{1}{\bar{a}_1 + \tilde{\tau}_{a_1}} h(\tilde{\xi}_{c_3}) \\ A_{\tilde{\epsilon}_{\beta_1}} & = -\frac{1}{\sigma_l} \frac{\sin(\bar{\gamma}_1 + \tilde{\epsilon}_{\gamma_1})}{\sin(\bar{\beta}_1 + \tilde{\epsilon}_{\beta_1}) \sin(\bar{\beta}_1 + \tilde{\epsilon}_{\beta_1} + \bar{\gamma}_1 + \tilde{\epsilon}_{\gamma_1})} h(\tilde{\xi}_{c_3}) \\ A_{\tilde{\epsilon}_{\gamma_2}} & = -\frac{1}{\sigma_l} \frac{\sin(\bar{\beta}_2 + \tilde{\epsilon}_{\beta_2})}{\sin(\bar{\gamma}_2 + \tilde{\epsilon}_{\gamma_2}) \sin(\bar{\beta}_2 + \tilde{\epsilon}_{\beta_2} + \bar{\gamma}_2 + \tilde{\epsilon}_{\gamma_2})} h(\tilde{\xi}_{c_3}) \\ A_{\tilde{\epsilon}_{\gamma_3}} & = -\frac{1}{\sigma_l} \frac{\sin(\bar{\beta}_3 + \tilde{\epsilon}_{\beta_3})}{\sin(\bar{\gamma}_3 + \tilde{\epsilon}_{\gamma_3}) \sin(\bar{\beta}_3 + \tilde{\epsilon}_{\beta_3} + \bar{\gamma}_3 + \tilde{\epsilon}_{\gamma_3})} h(\tilde{\xi}_{c_3}) \\ A_{\tilde{\epsilon}_{\gamma_1}} & = \frac{1}{\sigma_l} \text{cotan}(\bar{\beta}_1 + \tilde{\epsilon}_{\beta_1} + \bar{\gamma}_1 + \tilde{\epsilon}_{\gamma_1}) h(\tilde{\xi}_{c_3}) \\ A_{\tilde{\epsilon}_{\beta_2}} & = \frac{1}{\sigma_l} \text{cotan}(\bar{\beta}_2 + \tilde{\epsilon}_{\beta_2} + \bar{\gamma}_2 + \tilde{\epsilon}_{\gamma_2}) h(\tilde{\xi}_{c_3}) \\ A_{\tilde{\epsilon}_{\beta_3}} & = \frac{1}{\sigma_l} \text{cotan}(\bar{\beta}_3 + \tilde{\epsilon}_{\beta_3} + \bar{\gamma}_3 + \tilde{\epsilon}_{\gamma_3}) h(\tilde{\xi}_{c_3}) \end{cases} . \tag{5.38}$$

The relevant linearization point to pick is $\tilde{\xi}_{c_3} = \hat{\xi}_{c_3}$, i.e. so that the covariance is given at the tangent space around the most probable errors. This linearization point $\hat{\xi}_{c_3}$ is determined after processing the Lagrangian (5.24).

Finally, note that it is not inconceivable to create virtual measurements of the baselines at b_1 and c_2 nodes, as was done in CARLIT, in order to improve the sparsity patterns of the matrix $C_{\tilde{\xi}}$ and bring down the complexity of the saddle point computation of equation (5.30). This can also be used advantageously to improve the process of covariance query, for the same reasons. However, the small problems presented in this epilogue chapter are not worth the effort.

In the next paragraphs, we show in a numerical application how a baseline closure improves the confidence in the results of the corrected errors, given different values of parameters. Then, in another case study, we further analyse the pattern of independencies offered by the causal graph that represents a chain of triangle whose baselines are not located in the extreme ends of the chain. In both applications, we also discuss the results in comparison to the writings of Laplace and Delambre on the subject.

Numerical Application

We consider a simplified triangulation problem modeled for a chain of $n = 26$ equilateral triangles:

$$\begin{cases} \bar{a}_1 + \tilde{\tau}_{a_1} = 25 \text{ km} \\ \forall i \in 1, \dots, 26, \ \bar{\beta}_i + \tilde{\epsilon}_{\beta_i} = \frac{\pi}{3} \\ \forall i \in 1, \dots, 26, \ \bar{\gamma}_i + \tilde{\epsilon}_{\gamma_i} = \frac{\pi}{3} \\ \forall i \in 1, \dots, 26, \ \bar{\beta}_i + \bar{\gamma}_i + \tilde{\epsilon}_{\gamma_i} + \tilde{\epsilon}_{\beta_i} = \frac{2\pi}{3} \end{cases} \tag{5.39}$$

For various values of the variance of the angle measurement error, σ^2, we investigate the effect of the baseline closure on the posterior variance $\sigma^2_{c_n}$ of the length of the second baseline

c_n. c_n is the side located at the end the chain of triangles, similar to c_3 in the showcase figure (5.5).

The number of triangles chosen ($n = 26$) corresponds to the operations done for the extension of the Dunkerque-Paris-Barcelona meridian to the Balearic Islands (figure 1.8). The length of the triangle sides (25 km) is roughly representative of the dimension of the triangles[3].

Tables 5.1 and 5.2 show the results of the posterior standard deviations σ_{c_n} and σ_{post} for various input (or prior) values of standard error distribution σ and σ_l. The prior standard deviation on the baseline measurement error $\sigma_l = 0$ corresponds to the consideration made by Delambre *et al.* that the baseline measurement were exact, or near exact. $\sigma_l = 0.5$ m corresponds on the other hand to a more realistic value. Regarding the prior errors on the angle measurement, $\sigma = 1''$ corresponds to the advertised precision of the Borda's repeating circle. $\sigma = 4''$ would be a more realistic value given the analysis of excesses of the 3 angles from π on the 115 triangles of the Dunkerque-Paris-Barcelona meridian[4]. $\sigma = 40''$ is more faithful for the instruments used in the mid 18th century by figureheads such as Cassini, La Condamine and Maupertuis. Table 5.1 show these results without taking into account the baseline closure (no baseline measurement at c_{nj}). On the other hand, table 5.2 serves to appreciate the effect of baseline closure. We also want to gauge the implication of Laplace method, which difficulty for dealing with multivariate errors was noted in the previous paragraph 14, we give the results in two columns: one for the SCM method and one for Laplace's method (Laplace, 1820).

Table 5.1: Chain of $n = 26$ equilateral triangles, without baseline closure. σ_{post} is the standard deviation of the error a posteriori on the angle measurement; and σ_{c_n} is the posterior standard deviation of c_n, the length of the side c of the n-th triangle of the chain, given also a posteriori. 'A posteriori' refers here to the improvement of precision following the adjustment procedure with n triangle closures.

		SCM		Laplace	
σ_l value	σ value	σ_{c_n} (m)	σ_{post}	σ_{c_n} (m)	σ_{post}
$\sigma_l = 0.5$ m	$\sigma = 1''$	0.8150	$0.816''$	0.8212	$0.816''$
	$\sigma = 4''$	2.079	$3.266''$	2.653	$3.266''$
	$\sigma = 40''$	20.1906	$32.66''$	26.0627	$32.66''$
	$\sigma = 1'$	30.28	$48.99''$	39.09	$48.99''$
	$\sigma = 1^\circ$	1816.60	$48.99'$	2345.21	$48.99'$
$\sigma_l \approx 0$ m	$\sigma = 1''$	0.5046	$0.816''$	0.651	$0.816''$
	$\sigma = 4''$	2.018	$3.266''$	2.6057	$3.266''$
	$\sigma = 40''$	20.18	$32.66''$	26.06	$32.66''$
	$\sigma = 1'$	30.28	$48.99''$	39.09	$48.99''$
	$\sigma = 1^\circ$	1816.60	$48.99'$	2345.21	$48.99'$

First, consider the value of the variance σ_l^2 on the baseline measurements error. We note no significant effect of choosing $\sigma_l = 0$ over $\sigma_l = 0.5$, except when the variance on the angle error is the smallest ($\sigma = 1''$) without closure. In this case, the standard deviation of the unmeasured side c_n at the end of the chain is $0.815m$ (if the baseline measurement is considered perfect), versus $0.5046m$ (for $\sigma_l = 0.5$). Above $\sigma = 4''$, no change is perceptible.

[3] One side between the station 'Desierto' in the Spanish coast and the station 'Campvey' in Island of Ibiza was actually 160 km in length.

[4] This is treated in Laplace's 2nd supplement (Laplace, 1820).

Table 5.2: Incorporation of the baseline closure. σ_{post} is the standard deviation of the error on the angle measurement, given a posteriori; and σ_{c_n} is the standard deviation of c_n, the length of the side c of the n-th triangle of the chain, given also a posteriori. 'A posteriori' refers here to the improvement of precision following the adjustment procedure employing both the triangle closures and the baseline closure.

		SCM		Laplace	
σ_l value	σ value	σ_{c_n} (m)	σ_{post}	σ_{c_n} (m)	σ_{post}
$\sigma_l = 0.5$ m	$\sigma = 1''$	0.409	$0.816''$	0.427	$0.815''$
	$\sigma = 4''$	0.486	$3.266''$	0.4914	$3.254''$
	$\sigma = 40''$	0.4998	$32.66''$	0.4999	$32.534''$
	$\sigma = 1'$	0.4999	$48.99''$	0.4999	$48.801''$
	$\sigma = 1^\circ$	0.4999	$48.99'$	0.4999	$48.801'$
$\sigma_l \approx 0$m	$\sigma = 1''$	≈ 0	$0.816''$	≈ 0	$0.813''$
	$\sigma = 4''$	≈ 0	$3.2641''$	≈ 0	$3.2516''$
	$\sigma = 40''$	≈ 0	$32.66''$	≈ 0	$32.53''$
	$\sigma = 1'$	≈ 0	$48.99''$	≈ 0	$48.80''$
	$\sigma = 1^\circ$	≈ 0	$48.99'$	≈ 0	$48.801'$

Remark 5.2. *The standard deviation of $\sigma_{c_n} = 0.651$ m corresponds more or less to the results given by Laplace in the third supplement of* Théorie Analytique des Probabilités, *where it is stated that "if a second baseline were to be measured exactly at the end of chain (at c_n), you can bet one to one that the error would stay within a third of meter". Translated in more modern jargon for the 68% confidence interval, it is stated that $\sigma_{c_n} \approx 0.5$ m for the scenario without baseline closure. The difference with $\sigma_{c_n} = 0.651$ m stems undeniably from the fact that the triangles were not equilateral like in this simplified case, and other considerations must be taken into account in the real historical case (elevation, earth shape, etc...). Additional details on these calculations were not given in the book. Note finally that the SCM method produce more confident results at $\sigma_{c_n} = 0.5046$ m.*

In the scenario without baseline closure (table 5.1), since we know by design that the error τ_{a_1} (which distributes according to the parameter σ_l) is independent of the errors on the angle, the values of σ_{post} are unchanged according to the value of σ_l, as expected. Obviously, in the baseline closure situation, the choice of $\sigma_l \approx 0$ has a direct impact on $\sigma_{c_n} \approx 0$ from the logical implication that the second baseline c_n is measured with perfection.

Secondly, concerning the effect the baseline closure on the posterior standard deviation σ_{post} of the angle measurement errors, we found no noticeable improvement from $\sigma \to \sigma_{\text{post}}$ attributed to the baseline closure. The appreciable difference between σ and σ_{post} is essentially due to the triangle closures, whereby $\sigma_{\text{tr,post}} = \sqrt{\frac{2}{3}}\sigma$. Even for the largest standard deviation of $\sigma = 1^\circ$, we found no change attributable to baseline closure above the order of 10^{-7}. This suggests that to appreciably improve the posterior precision of the angle measurement, more baseline measurements are needed along the chain.

.

Case Study : Longer Chain of Triangles

The larger problem which was presented figures 1.1 (p15) and 1.6 (p25)gives us the opportunity to develop additional insights on the exploitation of conditional independences, thanks to the causal graph representation. The chain of triangles is reproduced for convenience on the left of figure 5.8. The right of figure 5.8 shows the truncated causal graph. For brevity, we do not write down the full SCM model: it is similar in pattern to SCM (5.19).

The measurement of $\bar{b}_{10}$ is the second baseline, and represents a baseline closure event. This yields the functional $b_{10} := \bar{b}_{10} + \tau_{b_{10}}$, where $\bar{b}_{10}$ is the measurement value and $\tau_{b_{10}}$ the exogenous variable (noise). In the 1799 commission report that established the value of the unit meter (figure 1.7 and quote subsection 1.4.1), it was noted that the two baselines (Melun and Perpignan) were "linked" [sic. liées] with each other by a chain of 53 triangles in-between. In the book *Basis*, Delambre (1807) postulated that the mismatch between the calculated and measured length of the second baseline should be removed by correcting the errors in the angles of the 53 triangles. Implicitly, Delambre qualitative consideration was that only 53 out of the 115 triangles should be entitled for corrections.

Can the causal graph figure 5.8 depict such an insight ? To answer the question, we have to see which errors are influenced by the closed collider $\bar{b}_{10}$. Let $\mathcal{E}_G$ denote all the observations in the causal graph 5.8 (named G) and let $\mathcal{A}_G$ denote all the actions (interventions). We have:

$$\mathcal{A} = \left\{\bar{\beta}_i\right\}_{i=1,\dots,13} \cup \{\bar{\gamma}_i\}_{i=1,\dots,13} \cup \{\bar{a}_1\}\,, \qquad \mathcal{E} = \{\bar{\alpha}_i\}_{i=1,\dots,13} \cup \left\{\bar{b}_{10}\right\}. \tag{5.40}$$

- Due to the open colliders b_{11}, b_{12} and b_{13}, there are no possibilities to correct the angles in the triangles (11), (12), (13) thanks to the second baseline:

$$\forall i \in \{11, 12, 13\}\,, \quad \begin{cases} \tau_{b_{10}} \perp\!\!\!\perp \epsilon_{\alpha_i} | \mathcal{E}_G, \mathcal{A}_G \\ \tau_{b_{10}} \perp\!\!\!\perp \epsilon_{\beta_i} | \mathcal{E}_G, \mathcal{A}_G \\ \tau_{b_{10}} \perp\!\!\!\perp \epsilon_{\gamma_i} | \mathcal{E}_G, \mathcal{A}_G \end{cases}. \tag{5.41}$$

- Due to the open colliders c_1 and c_3, there are no possibilities to correct the angles in triangle (3) (the triangle above the first baseline):

$$\begin{cases} \tau_{b_{10}} \perp\!\!\!\perp \epsilon_{\beta_3} | \mathcal{E}_G, \mathcal{A}_G \\ \tau_{b_{10}} \perp\!\!\!\perp \epsilon_{\gamma_3} | \mathcal{E}_G, \mathcal{A}_G \end{cases}. \tag{5.42}$$

- Finally, since $\bar{b}_{10} \in \mathcal{E}_G$, according to the rules of d-separation, all errors which are in the ancestor set of $\bar{b}_{10}$ in G are mutually dependent:

$$\forall i \in \{1, \dots, 10\}\,, \quad \begin{cases} \tau_{b_{10}} \not\perp\!\!\!\perp \epsilon_{\alpha_i} \mid \mathcal{E}_G, \mathcal{A}_G \\ \tau_{b_{10}} \not\perp\!\!\!\perp \epsilon_{\beta_i} \mid \mathcal{E}_G, \mathcal{A}_G \\ \tau_{b_{10}} \not\perp\!\!\!\perp \epsilon_{\gamma_i} \mid \mathcal{E}_G, \mathcal{A}_G \\ \tau_{a_1} \not\perp\!\!\!\perp \epsilon_{\alpha_i} \mid \mathcal{E}_G, \mathcal{A}_G \\ \tau_{a_1} \not\perp\!\!\!\perp \epsilon_{\beta_i} \mid \mathcal{E}_G, \mathcal{A}_G \\ \tau_{a_1} \not\perp\!\!\!\perp \epsilon_{\gamma_i} \mid \mathcal{E}_G, \mathcal{A}_G \\ \tau_{b_{10}} \not\perp\!\!\!\perp \tau_{a_1} \mid \mathcal{E}_G, \mathcal{A}_G \end{cases}, \tag{5.43}$$

  which provides the occasion to make simultaneous corrections (baseline closure).

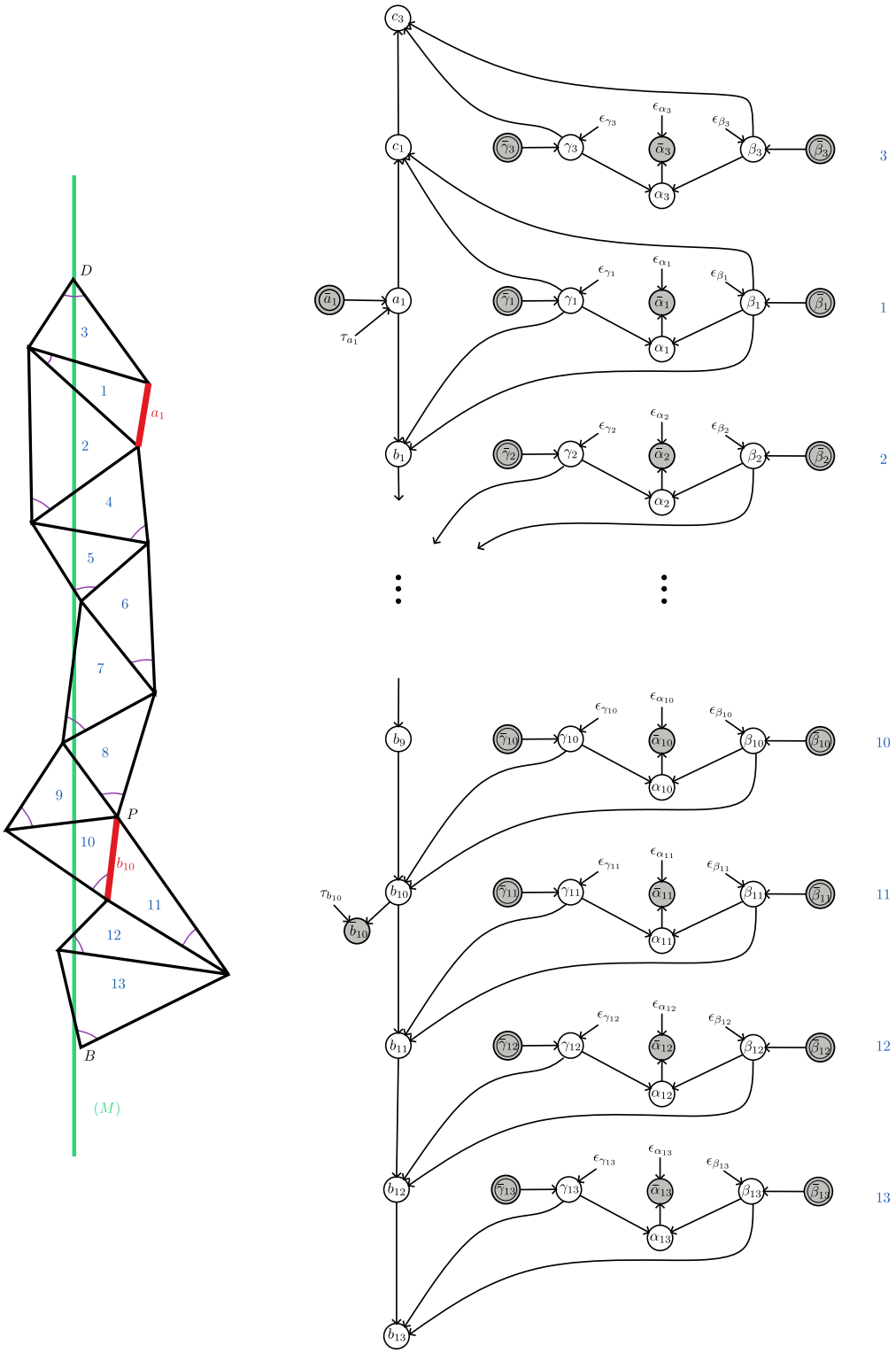


Figure 5.8: A larger chain of triangles (left) represented by a causal graph (right). Two sides (baselines) are measured.

Hence, the fact that the adjustment scope is limited to the triangles in-between the baselines is also validated by the graph. One difference with Delambre though, is that in this example, the errors on the baselines length τ_{a_1} and $\tau_{b_{10}}$ also get adjusted. As mentioned earlier, the measure of the baseline lengths were considered exact back then, which should not have been the case[5]. But it is easy to see, under the assumption that $a_1 \in \mathcal{A}_G$ and $b_{10} \in \mathcal{E}_G$ (no exogenous variables τ_{a_1} and $\tau_{b_{10}}$), that the same conclusion occurs for the scope of the corrections.

5.3 Representation of a network of triangles

For completeness, we now address the problem of representing the 'chain closure'. Chain closure occurs whenever a chain of triangles joins itself. This was typically a challenging problem in 19th century triangulation. An interesting example on the consequences of avoiding this problem can be recognized with the history of the 1830 French map, see remark 5.3.

Remark 5.3. *The 1830 French map was constituted of several chains of triangles which roughly followed meridians and parallels, see figure 1.12. In 1817, Laplace and Delambre had chosen this triangulation pattern likely for geodesic studies (Berthaut, 1898a). Unfortunately, both men died before all the observations were finished in 1830. Laplace (1820), who championed causality (subsection 1.5.3 and 1.5.4), did not attempt to provide an approach to deal with this type of closure. Consequently, and in addition to other problems (box 2), the map was adjusted by various combinations of baselines closures (there were 7 baselines), but not for chain closure type. Anecdotally, this resulted in residual mismatches that polluted this map. The inconsistencies in the map, in the order of tens of meters in the worst places, were moved in the overlapping margin of the published 270 sheets of the map*[6]*. But the problem was still there, and before the second world war, the 'Dépôt' was compelled to produce emergency spreadsheets containing ad-hoc rule-of-thumbs correcting offsets. The map users could then apply these offsets to smoothen out the mismatches between adjacent and overlapping sheets so that these maps were useful (presumably for military planning, artillery, etc...). This is described in Levallois (1988, p190-191). Nevertheless, these issues were identified and a new triangulation of France ('Nouvelle Triangulation' or NT France) was actually decided already by the end of the 19th century with the willingness, this time, to plan more carefully the calculations. Unfortunately again, due to numerous delays in the observations, NT France adjustment would only be finished in 1959 (Levallois, 1988, p216-217)!*

Our representation, based on causal graph, can cope with the chain closure type. We show our approach first on a simple example figure 5.9, which is a closed version of the simple chain of figure 5.5. The approach is then explained on a more realistic case study (figure 5.14) in the after-next paragraph 5.3.

Case Study : Simple Network

Figure 5.9 shows the simple network we work with, which is an extension of the simple chain figure 5.5. We assume that all angles of all four triangles have been measured.

[5] Contemporary GPS precision experiments reported a 10^{-5} relative error (Vincent, 1998), Puissant (1842) also noted that there were errors when the baselines were measured again for the 1830 French map.

[6] An assemblage of the polished 1830 map can be seen in https://www.geoportail.gouv.fr/carte, by selecting "Carte de l'Etat Major".

The main insight in the treatment of this kind of problem is to consider the fourth triangle as already induced by the other 3. In other words, there should be no causal intervention (*do-operator*) to 'build' the fourth triangle. First, remark that the angle α_4 can be deduced as the sum of angles around the central vertex should sum to 2π. Then, having access one angle and 2 sides, the fourth triangle is fully defined.

The question is then how to deduce astutely the angles β_4 and γ_4. A first approach consists in using the Al-Kashi theorem to get a_4, and then to use Al-Kashi again to get β_4. γ_4 is then deduced easily by summing the angles to π.

The corresponding truncated SCM model is:

$$\begin{cases} \alpha_4 := 2\pi - \beta_3 - \beta_2 - \alpha_1 \\ a_4 := \sqrt{b_4^2 + c_4^2 - 2b_4c_4\cos\alpha_4} \\ \beta_4 := \arccos\left(\frac{c_4^2+a_4^2-b_4^2}{2a_4c_4}\right) \\ \gamma_4 := \pi - \alpha_4 - \beta_4 \end{cases} \tag{5.44}$$

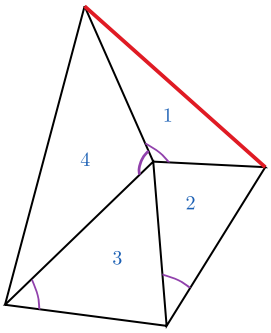


Figure 5.9: Simple network of triangles. The fourth triangle can be understood as induced by the other ones.

To get the full SCM model, append the truncated model to the simple chain SCM (5.18) which describes the functionals the first 3 triangles. It is not reproduced here for brevity. The causal graph of the concatenation of SCM (5.44) and (5.18) is given on figure 5.10. The additions from the partial SCM (5.44) are shown in blue. These are all open colliders for now.

To incorporate the measurements, we need to add them as observations in the graph (not as interventions). This is shown figure 5.11 (blue elements). The crucial aspect of the causal interpretation materializes in the fact that (1) for the triangles in a chain, 2 of the 3 angle measurements are cast as interventions; (2) in contrast for some triangles in a *network* (such as triangle 4 in figure 5.10), all 3 angles are already determined, therefore measurement data describing them can only be incorporated as observation.

One remark should be made from analyzing the pattern of newly created dependencies. In figure 5.11, the error τ_{a_1} correlates with the angular errors ϵ (all of them), which was not the case before (see figure 5.10). However, intuitively, acting on the value of the baseline should not influence the relation between the angles. Visually, one can imagine geometrically scaling up or down the size of this same network of triangles (figure 5.9) while preserving the angles. The only action available to achieve this size shift is acting upon the baseline length via $\bar{a}_1$.

The point of these considerations is as follows: the implications of an SCM/causal graph design can be checked a posteriori (and potentially criticized) to highlight logical flaws or inefficiencies, in order to then come up with an improved or more astute SCM.

To propose another solution to this problem, we will (1) remove the functional on a_4; (2) change the functional on β_4.

Our second approach goes as follows: similar to what was described figure (1.11) for the Hanover triangulation, we propose to multiply the ratio of sides in the circuit (1) $\to$ (2) $\to$ (3) $\to$ (4) $\to$ (1). This product equals one:

$$\frac{c_1}{b_1}\cdot\frac{a_2}{c_2}\cdot\frac{a_3}{c_3}\cdot\frac{c_4}{b_4} = 1, \tag{5.45}$$

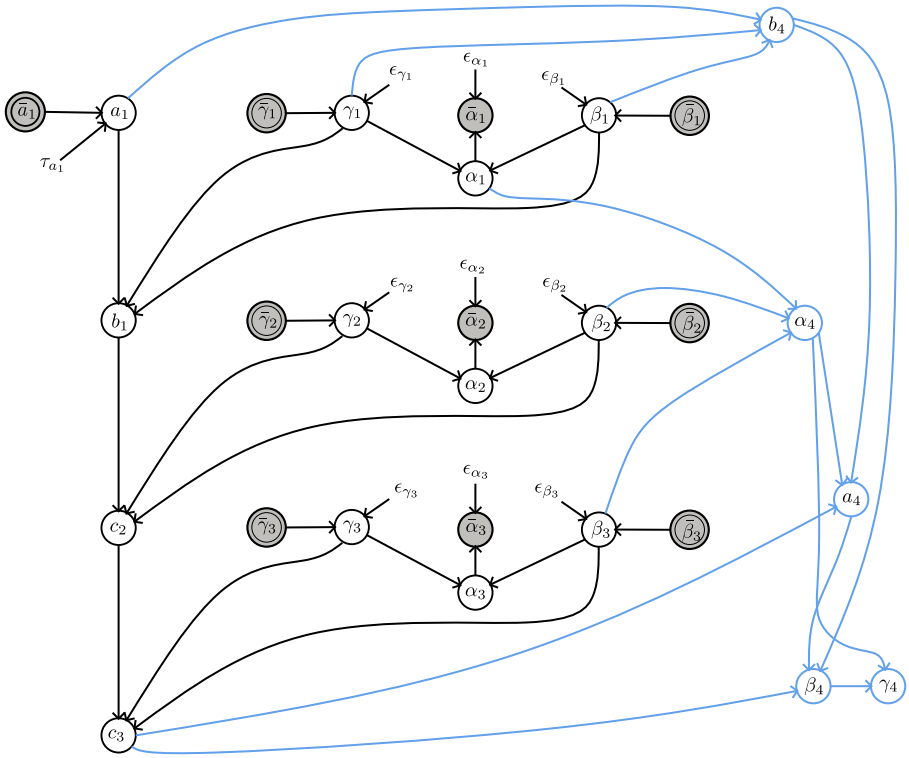


Figure 5.10: Causal graph of the simple network of triangles (figure 5.9) according to SCM (5.44). Additions from the partial SCM (5.44) are shown in blue.

which is true since $b_1 = a_2,\ c_2 = a_3,\ c_3 = c_4,\ c_4 = b_1$ and $b_4 = c_1$. Using the law of sines, replace the ratios of sides by ratio of sines:

$$\frac{\sin\gamma_1}{\sin\beta_1}\cdot\frac{\sin\alpha_2}{\sin\gamma_2}\cdot\frac{\sin\alpha_3}{\sin\gamma_3}\cdot\frac{\sin\gamma_4}{\sin\beta_4} = 1\ . \tag{5.46}$$

Let us replace $\sin\gamma_4$ by $\sin(\alpha_4+\beta_4)$:

$$\implies \cos\alpha_4 + \sin\alpha_4\cot\beta_4 = \frac{\sin\beta_1\sin\gamma_2\sin\gamma_3}{\sin\gamma_1\sin\alpha_2\sin\alpha_3}\ , \tag{5.47}$$

$$\implies \cot\beta_4 = \frac{\frac{\sin\beta_1\sin\gamma_2\sin\gamma_3}{\sin\gamma_1\sin\alpha_2\sin\alpha_3} - \cos\alpha_4}{\sin\alpha_4}\ . \tag{5.48}$$

We have identified our new functional on β_4:

$$\beta_4 := \operatorname{acot}\left(\frac{\frac{\sin\beta_1\sin\gamma_2\sin\gamma_3}{\sin\gamma_1\sin\alpha_2\sin\alpha_3} - \cos\alpha_4}{\sin\alpha_4}\right) \tag{5.49}$$

Recall that the other functionals for the chain closure event are $\alpha_4 := 2\pi - \beta_3 - \beta_2 - \alpha_1$, $\gamma_4 := \pi - \alpha_4 - \beta_4$. This yields the full SCM (5.50). The corresponding causal graph is

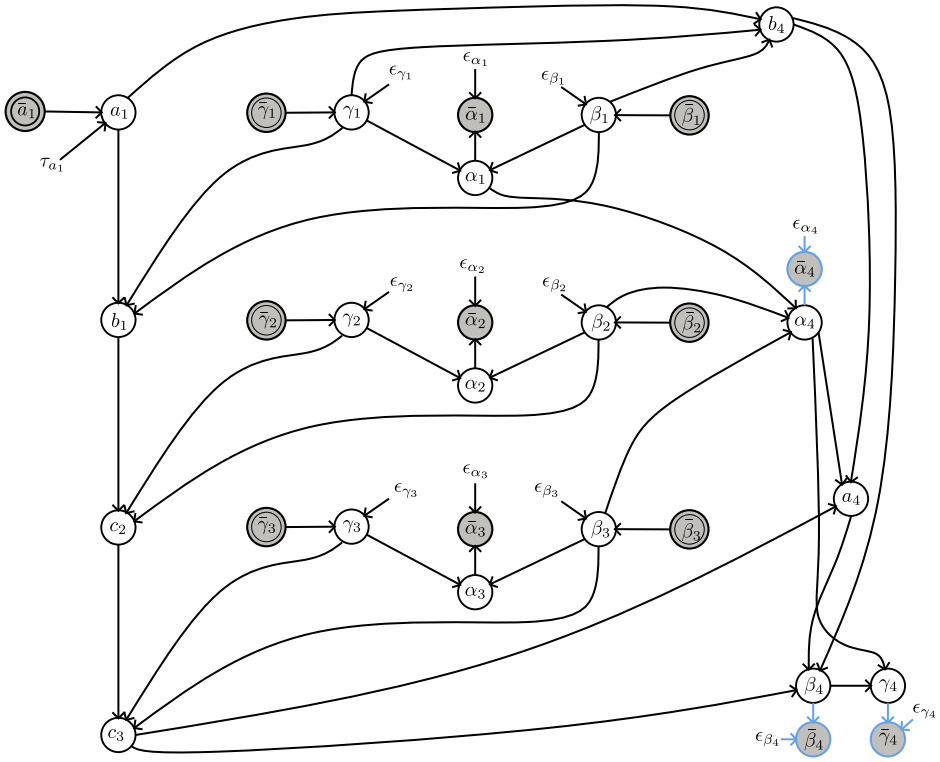


Figure 5.11: Causal graph from figure 5.9 of the simple network of triangles, completed with the measurements as observations (blue elements).

portrayed figure 5.12 without the angular measurements in the fourth triangle. Note that this is a cleaner graph compared to figure 5.10.

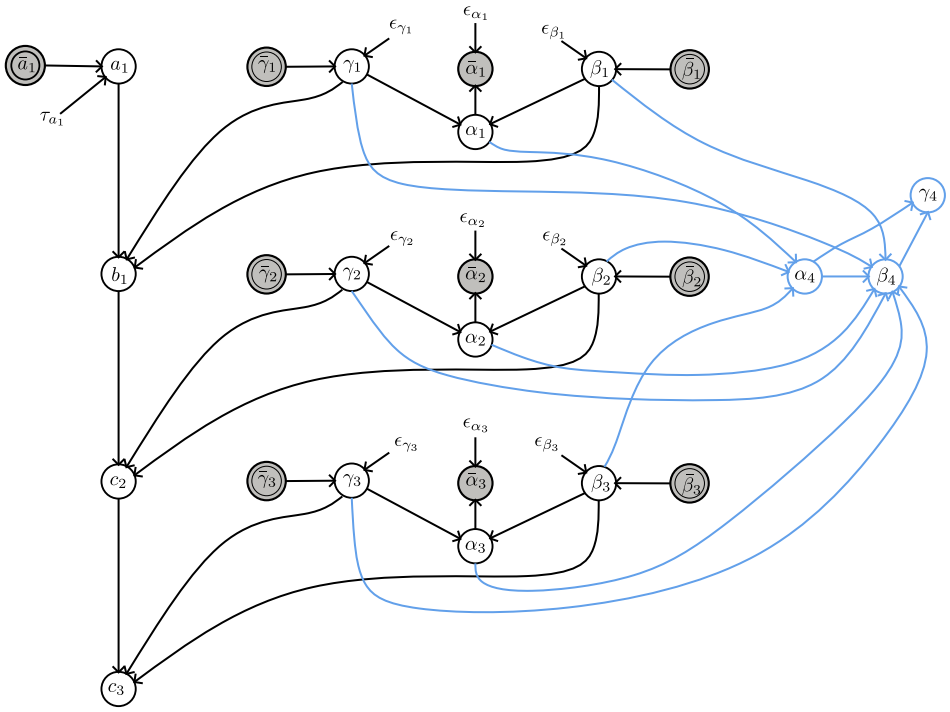


Figure 5.12: The second approach for the simple network problem of figure 5.9. New elements relevant to this problem are in blue. The main insight to solve the chain closure is to recognize that the 4th triangle is already implicitly defined by the existing geometry of the other 3 triangles.

$$
\left\{
\begin{array}{ll}
& \text{First triangle:} \\
a_1 & := \bar{a_1} + \tau_{a_1} \\
\beta_1 & := \bar{\beta_1} + \epsilon_{\beta_1} \\
\gamma_1 & := \bar{\gamma_1} + \epsilon_{\gamma_1} \\
\alpha_1 & := \pi - \beta_1 - \gamma_1 \\
\bar{\alpha}_1 & := \alpha_1 + \epsilon_{\alpha_1} \\
b_1 & := a_1 \frac{\sin\beta_1}{\sin(\beta_1+\gamma_1)} \\
& \text{Second triangle:} \\
\beta_2 & := \bar{\beta_2} + \epsilon_{\beta_2} \\
\gamma_2 & := \bar{\gamma_2} + \epsilon_{\gamma_2} \\
\alpha_2 & := \pi - \beta_2 - \gamma_2 \\
\bar{\alpha}_2 & := \alpha_2 + \epsilon_{\alpha_2} \\
c_2 & := a_2 \frac{\sin\gamma_2}{\sin(\beta_2+\gamma_2)} \\
& \text{Third triangle:} \\
\beta_3 & := \bar{\beta_3} + \epsilon_{\beta_3} \\
\gamma_3 & := \bar{\gamma_3} + \epsilon_{\gamma_3} \\
\alpha_3 & := \pi - \beta_3 - \gamma_3 \\
\bar{\alpha}_3 & := \alpha_3 + \epsilon_{\alpha_3} \\
c_3 & := a_3 \frac{\sin\gamma_3}{\sin(\beta_3+\gamma_3)}
\end{array}
\right.
\bigcup
\left\{
\begin{array}{ll}
& \text{Chain closure of the network:} \\
\alpha_4 & := 2\pi - \beta_3 - \beta_2 - \alpha_1 \\
\beta_4 & := \operatorname{acot}\left(\frac{\frac{\sin\beta_1 \sin\gamma_2 \sin\gamma_3}{\sin\gamma_1 \sin\alpha_2 \sin\alpha_3} - \cos\alpha_4}{\sin\alpha_4} \right) \\
\gamma_4 & := \pi - \alpha_4 - \beta_4
\end{array}
\right. . \qquad (5.50)
$$

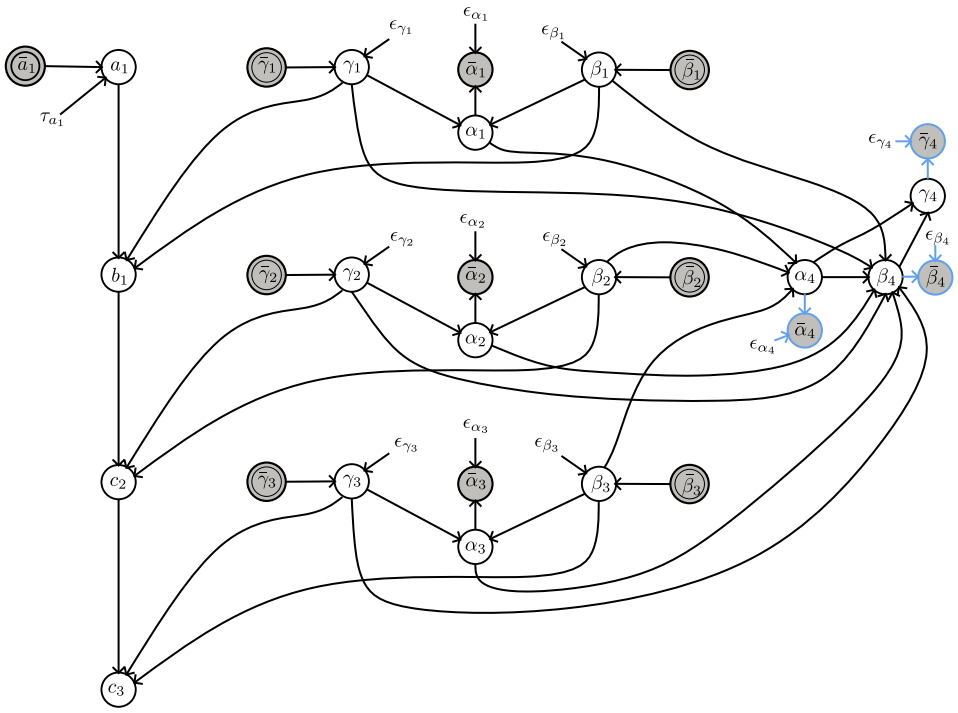


Figure 5.13: The second approach for the simple network problem of figure 5.9, with measurements (blue elements).

Adding the angular measurements from the fourth triangle as observation, we can see that, reading the independences on the graph figure 5.13, the corrections made from the measurements in blue do not involve the error on the baseline τ_{a_1}. This makes this model more suitable than the one presented in figures 5.10 and 5.11.

To incorporate the chain closure as a constrain in our solver, we use equation (5.47). In this document, we only set the constraint formulation:

$$\begin{cases} C_{\text{cc},4}(\epsilon) & \triangleq \cos(\bar{\alpha}_4 - \epsilon_{\alpha_4}) + \sin(\bar{\alpha}_4 - \epsilon_{\alpha_4}) \cot(\bar{\beta}_4 + \epsilon_{\beta_4}) - \frac{\sin(\bar{\beta}_1+\epsilon_{\beta_1}) \sin(\bar{\gamma}_2+\epsilon_{\gamma_2}) \sin(\bar{\gamma}_3+\epsilon_{\gamma_3})}{\sin(\bar{\gamma}_1+\epsilon_{\gamma_1}) \sin(\bar{\alpha}_2-\epsilon_{\alpha_2}) \sin(\bar{\alpha}_3-\epsilon_{\alpha_3})} \\ d_{\text{cc},4} & = 0 \end{cases} \tag{5.51}$$

As expected, the chain closure event introduce a constraint between a larger number of error terms (all angles in the case of model (5.50)).

Finally, in the previous SCM, the value of a_4 in figure 5.9 can be deduced easily since all other elements of the 4th triangle are available:

$$a_4 := b_4 \frac{\sin \alpha_4}{\sin \beta_4} \ . \tag{5.52}$$

Remark 5.4. *In figure 5.13, note that the elements of fourth triangle, in particular the variables $\bar{\beta}_4$ and $\bar{\gamma}_4$, differ from the elements of the other 3 triangles. In the induced fourth triangle, $\bar{\beta}_4$ and $\bar{\gamma}_4$ are not interventions (double-circles). This remark gives us the occasion to insist once more on the merits of distinguishing between seemingly similar variables according to their causal interpretation, for yet the same physical process (angle measurement). This*

subtlety would arguably be difficult to underscore in a purely probabilistic framework (see remark 5.1). On the other hand, this distinction naturally surfaces in the causal approach.

Case Study : Parallels-Meridians Network

A case study is considered on figure 5.14, which mimics the representation of overlapping chain of triangles. Two chains follows their respective meridian, while two others follow their respective parallel. The meridian and parallels are represented by green dashed lines.

The question at hand, which is posed by the 1830 French map (see remark 5.3) is: how to properly represent and adjust this type of triangulation ?

The main insight consists in identifying this type of problem as a generalization of the chain closure problem treated in the simple network of triangles problem (figure 5.9).

The first step involves the removal of the last triangles, numbered 27 and 28, so that the problem simplifies to a single chain of triangle (or rather, a single tree since it can branch out). This incomplete network is represented on figure 5.15. Its causal representation, and the method by which it is adjusted, is no different from what was described sections 5.1 and 5.2.

We then make use of the notion of induced triangles introduced earlier in this section for the simplest network problem. The goal is to rebuild the triangles 27 and 28, from existing triangles. However, this is not directly possible as there would be no avenue to deduce the angles of these triangles. Hence, this is why it is necessary to first deduce additional triangles $t_{v_1}, t_{v_2}, t_{v_3}, t_{v_4}$ in the inner part of the structure. This is shown figure 5.16. Each t_{v_i} triangle is built similarly as the 4th triangle in case study 5.3. Then, triangles 27 and 28 can be deduced in turn. Overall, the order in which these triangles are built is $\mathcal{O} \triangleq [t_{v_1}, t_{v_2}, t_{v_3}, t_{v_4}, t_{27}, t_{28}]$. Hence, the corresponding causal graph would be extended with open colliders of the variables of the deduced triangles. Finally, the measurements available for triangles 27 and 28 creates mismatches between the deduced elements and the measured ones, creating new opportunities for adjustment. In the scope of this document, we only address this last problem qualitatively, to show that there is no impediment to tackle this problem using the SCM framework.

The induced triangles $t_{v_1}, t_{v_2}, t_{v_3}, t_{v_4}$ represented in figure 5.16 are not necessarily the only possible combination to build the triangles 27 and 28. This suggests that, in order to automate the process, a heuristic should be proposed that picks the induced triangles to be built. But this goes out of the scope of this work.

It is not cleared to us at this time, whether there exists a map which makes use of the methods described in this chapter, or whether they reached equivalent results but in a non-explicit manner.

The practice described in Puissant (1842) treatise, if applied to this network of triangles, was to solve for a baseline closure problems in a chain, using various combinations of paths from baseline to baseline. Thus, one baseline closure was conducted via the path $p_1 = (1, 2, 4, 5, 6, 7, 8, 9, 10)$, and another proceed via the path $p_2 = (1, 2, 4, 5, 28, 27, 26, 25, 23, 22, 21, 17, 16, 15, 14, 9)$ as if in both cases the other path had not existed. However, as noted by Levallois (1988), this produced a mismatch since the correction via one path generated a mismatch from the point of view of the other path.

As mentioned in the chapter 1 for the Hanover network (figure 1.9), another practice introduced by Gauss & colleagues consisted in identifying closures in the network by considering as many geometric constraints as possible, and using them as constraint equations in an under-determined least square system (Grcar, 2011). In principle, the method could be

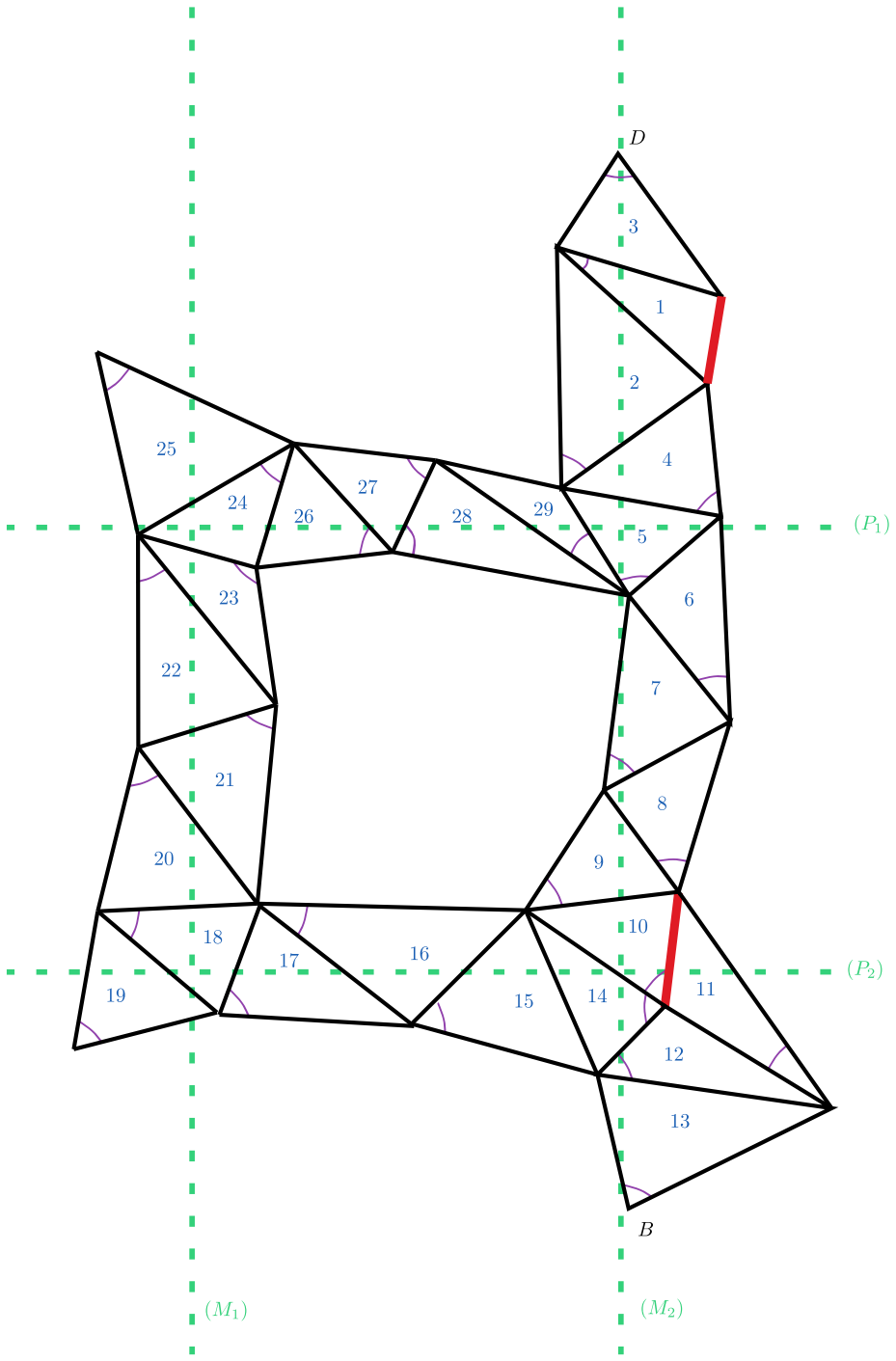


Figure 5.14: Network of triangles representative of the difficulty of closed chain triangulation. This type of problem was not solved in the 19th century, due to the lack of adequate mathematical tools to represent causal assumptions one might have about the problem.

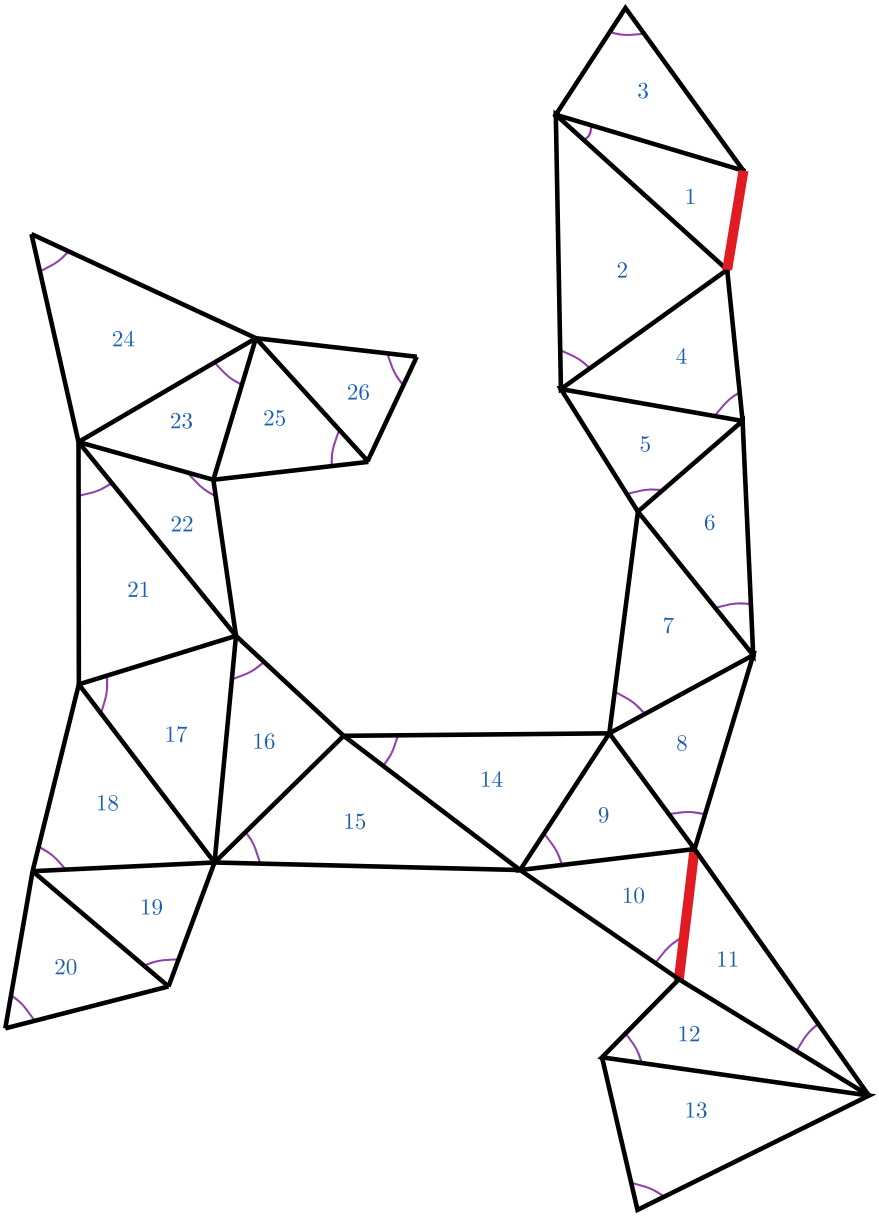


Figure 5.15: Incomplete representation of the triangulation problem from figure 5.14, for which the triangles numbered 27 and 28 have been removed.

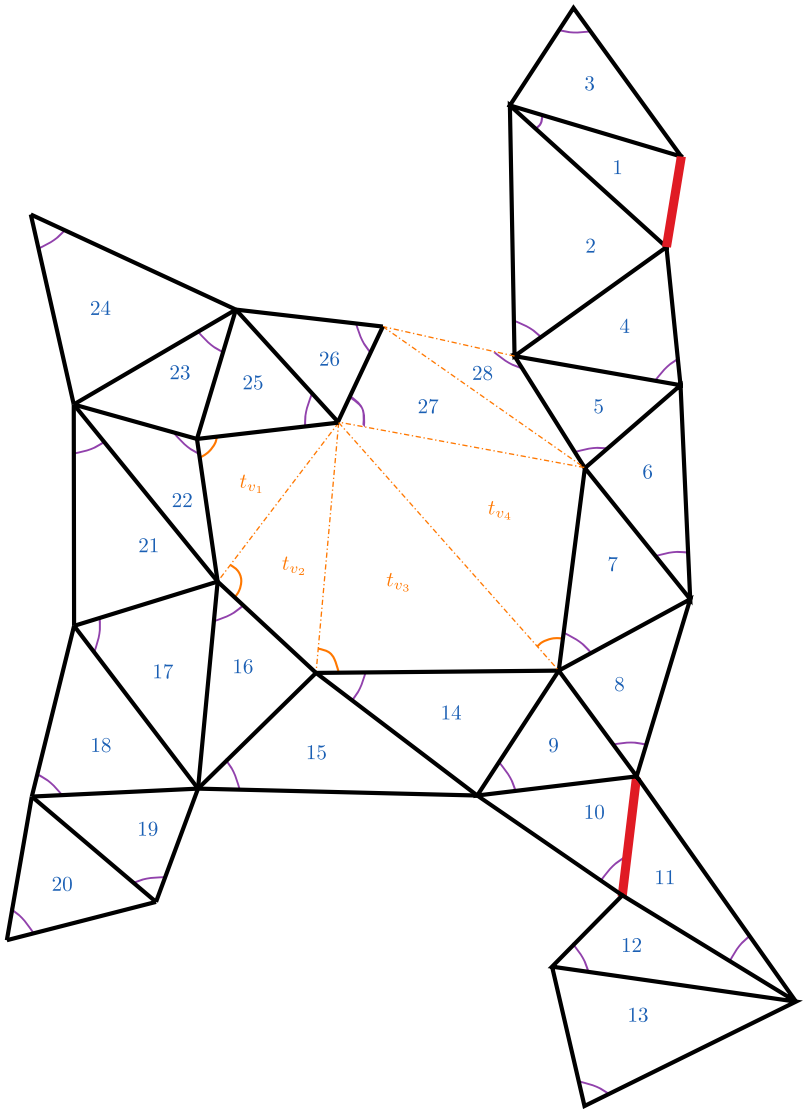


Figure 5.16: Representation of the triangulation problem from figure 5.14, with the induced triangles $t_{v_1}, t_{v_2}, t_{v_3}, t_{v_4}$ highlighted. These induced triangles are necessary to build causally triangles 27 and 28.

equivalent to the one proposed here, but it seems that there was no systematic way to ensure that all relevant closures were considered, or that some were not implicitly double counted.

As noted later by Tardi (1934, chap. 10), a system that uses directly the geographical coordinates of the vertices as decision variables was easier to set[7]. This may explain, why, both in modern SLAM and geodetic adjustment, the errors are typically not the decision variables: it is not because its scientific grounds are better or worse, but it leads more easily, and without ambiguity, to the formulation of a (least squares) optimization problem.

5.4 Final remarks on the early metric system

In this last section, we take a step back from the theoretical developments in this manuscript in order to shed a new light in the link between (a) the modern day scientific question of spatial understanding in robotics, and (b) the geodetic adjustment problem which occurred during the creation of the early metric system (late 18th century). Problem (b) is in essence similar to the chain of triangles presented figure 5.8, since, with the help of some additional geometry, its SCM can be expanded to encompass the distance $[DB]$ which is use to find the arc length, and then to deduce the meridian length which was the favored quantity to define the early metric system. More generally any result can be inferred by the addition of open colliders, whether the purpose of the inference process is metrological, or cartographic.

The causal representation applied to (b) in this chapter does refine our views given in chapter 1 about the historical development of probabilistic inference techniques in that period. We observed that there were indeed problems with these approaches due to several factors, including (1) the absence of method to address multivariate problems and (2) the lack of formalism concerning the articulation of causal assumptions. Issue (1) ended up being fixed by contributions in the early 20th century (by, e.g., Pearson, Fisher et al.). However, while we have showed that the tools of Pearl & co-authors can be judiciously employed to tackle issue (2), they are arguably not yet widespread in spatial inference problems. Yet, continuing on issue (2), it was recognized early in *Essay* (Laplace, 1814a, p11) that correctly enunciating the assumptions of a problem was non-trivial : "the theory of probabilities ties in to considerations so delicate, that it is not surprising that, with the same data, two individuals get a different result, particularly in very complex topics". Case in point, the problem of geodetic network adjustment was one of such "complex topic". As shown in §5.3, there was no proper adjustment proposed for a network of triangles (versus just a chain of triangles) because it was not recognized that a different structural formulation was needed for the triangles involved in the chain closure events. Ideally, this type of problem should have highlighted the need to bridge the gap for the development of a better-than-verbose method to specify those "delicate considerations", for they are an essential first step to address inference problems that are causally not trivial (like SLAM or geodetic networks).

We put forward some answers as to why, in the fallout of the early metric system, many opportunities were missed to settle on a somewhat cohesive framework which couples causal statements and analytical probabilities.

[7] This could conceivably be represented by a factors graph were the nodes are the vertices of the triangle. However, in large geographical surveys, it required to commit earlier in the development to one coordinate projection system (Lambert, Mercator etc...), while in contrast the system based on the adjustment the errors could avoid this difficulty during the optimization process.

We classify those answers in two categories: one which explains why the direct contributors of the metric system (Delambre, Legendre, Laplace et al.) did not manage to develop a causal framework (despite causal determinism being a widespread view at the time, as recalled by Van Strien (2014)), and the second category which explains why their successors were not able to tackle the shortcomings leftover from the period of the early metric system.

5.4.1 The role of the contributors

One of the most important theme of our work is the relevance of graphs, particularly DAGs, to represent faithfully the assumptions one has on a given spatial inference problem. More broadly, the benefits of directed graphs in statistical and causal inference topics have been well delineated by Pearl (Pearl and Mackenzie, 2020). Notice however that, although the manipulation of graphs are nowadays generally studied within the field of computer science, their relevance to probability theory is not particularly tied to technological capacities. The association of statistical independences and graphical notions, a.k.a. the graphoid axioms (Pearl and Paz, 1985), were built on the work of conditional independences (Dawid, 1979). These works are relatively recent, but one does not need computers and programming languages to find graphs useful, whether we need them for expressing assumptions of the statistical type (Bayesian networks, MRF, factor graphs, etc...) or of the causal type (causal graphs). Besides, the first publication on graph theory is generally attributed to Euler, back in 1736. Thus, one would think that the concept of graphs was not alien to the scholars of the early 19th century.

In our view, the most likely explanation for the lack of popularity of graphs in early probability theory was the domination of another mode of expression which overshadowed other modes (including geometry and graphs): the analytical expression, in a broad sense. Kožnjak (2015) revealed that the geodesian and polymath Boscovich, who contributed to the doctrine of determinism by being the first writer of the tirade commonly known as the 'Laplace demon' (decades before Laplace, as mentioned in §1.5.3), routinely used unpopular geometric methods (Maire and Boscovich, 1755) which were repudiated by prominent members of the Académie such as D'Alembert and Lagrange. Kožnjak (2022, note 42) recalls that Lagrange was proud of the fact that his book "Mécanique Analitique [sic]" did not contain any graphical representation, and Laplace himself (author of the "*Analytical* theory of probabilities") criticized Boscovich for using geometrical notions (Kožnjak, 2022, note 50). This of course was not limited to the dispute with Boscovich and his adversaries, Fourier's "Théorie *Analytique* de la chaleur" (analytical theory of heat) (Fourier, 1822) and even Legendre presentation of the method of least squares (Legendre, 1805) are other major examples of pure analytical achievements. In a paper commemorating the two hundred years of the publication of the method of least squares, Falguerolles and Pinchon (2006) regrets the absence of graphics in Legendre's presentation. We can thus conclude that this historical period did not favor the outbreak of a variant of causal graphs[8].

5.4.2 The role of the successors

Overall, despite those scientific problems exposed in hindsight, the scientific aspects of the early metric system endeavor were nonetheless initially considered as a great success. These

[8]Pearl attributes the invention of causal graphs (then named 'path diagrams') to biologist Sewall Wright in 1922 (Pearl and Mackenzie, 2020, chap. 2), i.e. more than a century later and far from this environment.

efforts participated in finally solving a rampant societal challenge: establishing a unifying weights and measures system (see §1.1). For the anecdote, it is even reported that when Delambre (1810) presented the third volume of *Basis of the decimal metric system* to Napoléon, the ruler expressed admiration: "Conquests come and go, but these operations will remain." (Alder, 2003). The conception of the metric system was a source for national pride too: its motto, '*For all times, for all people*', shows that the objective was to solve once and for all the proliferation of weight and measure standards, not only in France, but also internationally, to facilitate trade. As such, it can be considered that the metric system was not just a scientific endeavor, but that there were a soft power operation (Guedj, 2000).

Box 11: The schemes of a Big Science project

Since geodesic operations were at the time a discipline dominated by France (Jozeau, 1997), it influenced, at the political level, the decision of selecting the meridian arc length as the basis of the unit meter (Guedj, 2000). But the Académie also had its own agenda: a drive to keep studying of the shape of the earth. Biot (1803, p35-36) admitted a few years later that the true motives for recommending the meridian arc length over the 'seconds pendulum'[a] was concealed: "sometimes to serve men, we must resolve to deceive them". One can also suspect money to be a factor, Borda could promote his repeating circles instruments (figure 1.2), and the grant secured for the project was three times the yearly budget of the Académie (Alder, 2003, chap 3). In contrast, the data & arguments in favor of the seconds pendulum was already available for decades, thanks to Condamine (1748) who knew that measuring a meridian arc length was not the best choice (i.e., 57 years before Legendre's conclusion, see §1.3.1). Ironically, in post analysis, the 1740 measurement of the Paris meridian arc length, done by the famous cartographer Cassini (the 3rd of his name) and calculated by the geometer La Caille, ended up being slightly closer to the aspired value (Guedj, 2000, chap 17) than the value proclaimed by the 1799 commission. While the scope of the project was 'just' to standardize weights & measures, the temptation of making a clean state won out: in the euphoria of the Révolution, the scholars would help to create a new kind of human being, one who would be free from the corrupting influences of the old regime. Some of these scientists (including Monge and Lagrange) helped to push forward in 1793 the extension of the decimal mode of counting to clocks and calendars (i.e., 10 hours watches and 10 days per week, with new names), a taint that survived a few years.
As Bouasse (1919, p274) later commented, the ability to measure angles at $1''$ of precision does not make people of bad company virtuous. The metric system would be temporarily abandoned in 1815, and is no more defined on the meridian arc length, although its nominal value was kept for back-compatibility.

[a] The "seconds pendulum" refers to a simple pendulum with a specific length such that the period of oscillation is 2 seconds. This length was debated to be the basis of the metric system, over the meridian arc length.

Anyway, some careers were greatly boosted, particularly Méchain's, Delambre's and a few years later Arago's. It is then easy to conceive that this prestige also radiated into the novel methods and instruments either employed or invented for the occasion.

More broadly, it seems that the metric system enterprise was afterwards surrounded by an aura of deference, to an unwarranted degree (see box 11). This had adverse effect on the work of the next generations of geometers. In addition, it seems that the theoretical results in probability theory achieved by Laplace, Gauss *et al.* were not understood by their peers (Crepel, 1989; Hald, 2007). Besides, it is recognized that Laplace was difficult to read: his CLT proof was not the clearest (Haggstrom, 2015) and most developments are indeed tedious (see, e.g., Pulskamp's translation notes[9]). Consequently, as mentioned in §1.5, the post-Laplacian treatise on geodesy written by Puissant (1842) only reproduced Laplace's results without much of an explanation. All in all, this resulted in a slower pace of innovation in designing new instruments or researching new methods (especially concerning causality). Several anecdotes signal that an unfortunate gatekeeping attitude settled in. For instance, the repeating circle (figure 1.2) was used by the French geodesians for more than sixty years (Jozeau, 1997) while neighboring countries instruments were steadily improving. A second anecdote highlight a missed opportunity to trigger a research project. In 1900, a new expedition to measure a meridian arc length in Peru/Ecuador was contemplated, more than one century and a half after the tumultuous expedition of La Condamine & colleagues. Asked by a state minister whether or not it would appropriate for the *Académie des Sciences* to send some eminent members there, mathematician Poincaré (1900) (then a figurehead of the Académie) argued in favor of sending instead cartographers from the army (the 'Dépôt de Guerre' organization, i.e. the unit mentioned in §1.4.5 and of which Cholesky was a member). Poincaré was aware of the various challenges encountered by his predecessors and pointed out that the expedition would benefit from having a disciplined, well-trained and cohesive team. The missed opportunity in this anecdote stems from the fact that he considered that there were no more scientific challenges in geodetic adjustment: the correct methods and procedures have already been established following the 1792-1799 Dunkerque-Paris-Barcelona meridian operations and the cartographers of the 'Dépôt' were adequate gatekeepers of the "traditions" of French geodesy. The use of the word "tradition" in the letter clearly signals that not only geodetic adjustment was not an active field of research, but also that the deference towards the work done at the time of the early metric system (and shortly after) persisted. This actually hindered scientific progress and innovation. Essentially, what would remain of the Gauss-Laplace synthesis period were probability theory (mostly misunderstood) and the successes of the least squares technique. Minimizing the least squares of the errors was straightforward, and had been well presented by its inventor (or co-inventor), Legendre (Stigler, 1981; Legendre, 1805). But in itself, the method of least squares is just a data fitting technique, and does not help in the specification of qualitative considerations one has about his or her problem. To the contrary, since a result can be produced via the method of least squares, it may diminish the otherwise necessary effort to pay attention to the subtleties of the starting assumptions. Indeed, getting those subtleties wrong can still produce a credible result such as, e.g., a seemingly compelling cartography, but it is not assessable whether a seemingly good result is obtained thanks to the quality of the method employed or despite its shortcomings. In that regard, the critics of the least squares, who were mentioned in a bad light in §1.4.5 for their defiance vis-à-vis the least squares technique, should be partially rehabilitated. In the 1850s, astronomer Le Verrier indeed push in favor of abandoning the teaching of the least squares so that students can use their *judgements*, which suggests that they were mindlessly applying it without wisdom. We also mentioned the case of 1910s physicist Bouasse from Toulouse (Bouasse, 1919), who did

[9] http://www.probabilityandfinance.com/pulskamp/Laplace/index.html

not take the 'adjustment' part of geodetic adjustment (done via least squares) seriously: it was a way to tinker with the results. The overall message of these *no-nonsense* characters was to favor reasoning. Though anecdotal, it is amusing to notice that the only scholar Bouasse has any respect for is La Condamine (of the tumultuous 1735 Peru/Ecuador expedition to determine a meridian arc length). Despite the fact that La Condamine adjusted his chain of triangles using ad-hoc averaging methods, he was willing to discuss thoroughly his reasoning when it came to the application of corrections he applied to his measurements, unlike the two others scholars in his company. The "most sympathetic among the scholars involved in the study of the shape of earth" (Bouasse, 1919, p274) reported his approach in his journal (Condamine, 1745, art25-26 p455-63). All in all, the only issue we have with the arguments of the least squares opponents, is that they considered the method in itself to be the culprit. Rather, there is no doubt that the least squares method has its place, but causality should be expressed first, then exploited, in order to bring about a tailored least squares problem (e.g., as in (4.16) and/or (5.24)) which abides by the starting assumptions (expressed via SCM and causal graphs). The least squares method simply derives from the normal distribution of error terms, which have strong justification in large sample situations (CLT).

5.4.3 An explainable system

Questionable aspects related to the setup of the metric system were mentioned above, as well as some adverse effects it has had on successors. Nonetheless, it would be fair to emphasize a positive aspect in the design of this system: the fact that it was envisioned as an explainable system, where once the meter value is taken in nature, all other quantities could be deduced from reasoning. Mineralogist René Haüy (1743-1822), secretary of the weights & measures commission emphasized on the cause-and-effect thinking deeply rooted in the design of the metric system (Haüy, 1793, p11-12):

> Everything is linked in this [metric] system, entities are held together via intimate relationships, every result follows from the preceding one, and brings about the following one, so that once established that the unit of measure must be taken in nature [i.e., by measuring an arc length of a meridian], the plan of the system was already plotted in advance, by the precise genealogical order of ideas.

Hence, the early metric system itself was, profoundly, a causal system. Once the chain of triangles is established, that its errors are adjusted, and that the latitudes are measured, then the value of the unit meter can be deduced from the arc length. Then the other quantities of this system (surfaces, volumes and weights) were in turn deduced downstream: meter $\rightarrow$ meter^2 $\rightarrow$ meter^3; (meter^3 + water) $\rightarrow$ kilogramme.

By framing the metric system as an explainable construct, its adoption was significantly facilitated, as humans inherently tend to trust these frameworks over arbitrary quantification. This realization should serve as a valuable inspiration for addressing the societal impact of robotics. Autonomous machines designed as much as possible on principles of cause-and-effect will be more transparent and more reliable than those based on excessively large and opaque algorithms.

Remark 5.5. *In SLAM terminology, appellations such as 'dense SLAM', 'topological SLAM', 'semantic SLAM' or* 'metric *SLAM' are sometimes employed to broadly categorize the hypotheses and data structure a given approach of SLAM is tied to. These designations are*

often seen as slight abuses of language, but are mostly tolerated as commonly used shorthand expressions. Using a modest play on words and an openly anachronistic statement, we will allow ourselves to reverse the logic of 'metric *SLAM', for it is in fact the early metric system which can be in essence categorized as an ancestral causal SLAM system.*

Appendix A

On Magellan's loop-closure

The Magellan shipping expedition, which occurred in the early XVIth century, is universally best known for being the first recorded circumnavigation of Earth. Concerning the understanding of world, the period (a.k.a. the age of discovery) is also associated with a drastic improvement of world maps. Magellan's circumnavigation has the particular property that a previously known and visited territory (South East Asia & the Indonesian Islands) was reached via the Pacific route[1]. Of particular interest to us is the event associated with the loop closure by missing data during the crossing of the Pacific[2]: i.e. how under (1) low confidence on the precision of the available map (see fig A.1), (2) the growing uncertainty on the ships position, and (3) the absence of expected land sighting, an informative map adjustment was nonetheless conducted.

The purpose of this appendix is twofold. A first goal is to highlight the fact that this type of adjustment, although significant in human cartography, is so far not well covered in robotics (as remarked § 2.3.1). This in turn prompts a discussion on an approach for adjusting under similar missing data situations, as this type of corrections should be an essential component in robotics spatial inference. We give our perspective on that prospect, notably in relation to the causal framework discussed in chapters 4 and 5, which constitutes the second purpose of this appendix.

A.1 Background

Before 1500, Asia was very clearly overvalued in its dimensions and its extension towards the East (Duviols and Castro, 2019, p10-11). Between 1500 and 1519, the already visited places (including by Magellan himself) in S.E. Asia, by the route of the Indian Ocean, lead to the adjustment of S.E. Asia to relatively more realistic proportions[3] Nonetheless, the contour of East Asia were vastly inaccurate and Islands such as Japan were only known via secondary accounts[4]. Concerning the position of the Moluccas on the map (archipelago near

[1] See, e.g., the Wikipedia entry https://en.wikipedia.org/wiki/Magellan_expedition. Our main source for this discussion is Castro and Bernand (2010); Duviols and Castro (2019) book on the subject, a reference that reproduced direct testimonies on the expedition and analyzed them.

[2] See the trajectory February and March 1521 in the Wikipedia link provided.

[3] See for comparison with the 1519 map of J. Reinel (fig A.1), the 1492 globe of Martin Behaim (https://gallica.bnf.fr/ark:/12148/btv1b55008737g/f1.medres3d).

[4] The cartographic and navigation considerations are hereby only developed from the limited knowledge available of the sailors (i.e. from Spain and Portugal), not inhabitants of the region, for the point is to study

Figure A.1: Jorge Reinel 1519 map of the world, which is suspected to have been the map employed for the Magellan expedition. The vertical red dashed line, slightly visible in the middle of the map, splits the world in two zones in which newly discovered territories could be claimed either by Spain (left side) and Portugal (right side) under treaty. The left and right edges of the map denote the anti-meridian, i.e. the prolongation of the treaty line on the other side of the earth. The Moluccas archipelago (a.k.a. spice islands, visible far left on the map) was close to the anti-meridian and claimed by both European powers. The Magellan expedition main goal was to reach this strategic location, to the benefit of Spain, by navigating south of Brazil and by crossing the uncharted Pacific Ocean. In reality, the Moluques Islands were in the Portuguese side. In practice, this means that this map underestimated the length of the trip by a few degrees. Document reproduced from Gallica - Bibliothèque Nationale de France (Reinel, 1519).

the left edge of figure A.1), they were positioned to be 5 degrees to the east of anti-meridian, while in reality they should be 5 degrees west of it (Duviols and Castro, 2019, p215). As additional inaccuracies affected the size of the earth, overall, the size of the uncharted Pacific Ocean was slightly underestimated (although not by an order of magnitude). Thus, when crossing the Pacific Ocean from the Magellan strait (south America) to Guam and then to the Philippines, from December 1520 to March 1521, Magellan ships underestimated the distance to be travelled by a few longitudinal degrees. An unfortunate and aggravating factor was that no rich land was encountered during the crossing, which would have allowed to resupply.

Of interest here is the process by which cartographic adjustment were inferred before the end of the crossing. This process was triggered not by 'closing the loop' on known territory, but precisely by the lack of land sighting. To develop this point, we rely on two online accessible accounts of the expedition.

The first account is from Albo's log book (Albo, 1522), who was the pilot of one of the ship. It contains a fair amount of navigation information such as direction of the ship, latitudes and longitudes. The determination of precise latitudes were generally easy in celestial navigation, and Castro and Bernand (2010, p660) note that the errors were low: out of 37 latitude measurements compared, only 1 of them has an error of more than 1 degree. By contrast,

how they coped with their localisation and mapping uncertainties.

the determination of longitudes was unreliable and error-prone, and were only done on land (it would remain a challenge until the invention of Harrison chronometers in the XVIIIth century). (Castro and Bernand, 2010, p664) analyzes that the errors of the longitudinal values established were between 2.5 and 10 degrees, under some conservative hypotheses. Thus, by the beginning of March 1521, before encountering the islands of Guam, the longitudinal component of the ships localization was extremely unreliable (and known to be so) due to both the unreliable process of establishing the longitude values and the long distance travelled since the last longitude measurement.

The second relevant account of the expedition is Pigafetta's book (Pigafetta, 1522). Although not a sailor, he was considered close to Magellan and was aware of the decisions made by him. During the crossing, Pigafetta mentioned the following (Pigafetta, 1522) :

> In going by this course we passed near two very rich islands; one is in twenty degrees latitude in the antarctic pole, and is called Cipanghu [i.e. Japan, which was incorrectly believed to be on the path]; the other, in fifteen degrees of the same pole, is named Sumbdit Pradit. After we had passed the equinoctial line we navigated between west, and north-west and a quarter west, by north-west. Afterwards we made two hundred leagues to westwards, then changed the course to a quarter of south-west, until in thirteen degrees north latitude, in order to approach the land of Cape Gaticara, *which cape (under correction of those who have made cosmography), (for they have never seen it), is not placed where they think, but is towards the north, in twelve degrees or thereabouts.*

As analyzed by Castro and Bernand (2010, p386), the "cape Gaticara" mentioned in the quote was supposed to be encountered by the expedition. It was in fact a non-existent cape on the east-most region of Asia which was based on indirect information ("for they [the cartographers] have never seen it"). The crucial bit of information contained in the above concerns the adjustment of this imaginary cape Gaticara on the map: given all stated uncertainties, concerning both the low confidence on the information on their map and the localization of the ships, it was decided to adjust the position of this landmark. Had the ship been localization been more precise, the adjustment would have occurred earlier, not a couple of days before the first land sight (at Guam).

In other words, in the judgement of these experimented sailors, the lack of expected land sighting was informative, for actionable conclusions could be inferred by attributing the absence of data to a certain disposition of the map or of the ship localization (in that case, the former).

In his romanced biography of Magellan, writer Stefan Sweig would exaggerate the death toll and the sense of hopelessness of the crew during Pacific crossing. He nonetheless fittingly conveyed the idea that it is the absence of land sighting that triggered Magellan to revise pre-established cartographic notions:

> *[...] and still there came no land, nor any sign to give hope that they were drawing near the land. [...] For thousands upon thousands of hours did Magellan's fleet move onward into vacancy.*
> *Since November 28, when Cabo Deseado [near Horn's cape] had faded out of sight on the horizon, they had had neither maps nor means of measurement [this is however untrue, see above]. False had proved Faleiro's estimate of distances [Magellan's astronomer who helped prepare the expedition]. Long since, thought Magellan,*

he must have got beyond Cipangu (as Marco Polo had called it), the Land of the rising Sun. Yet at the time when he believed this, he had not yet traversed as much as a third of the width of the vast ocean which, because it was so peaceful, he called the Pacific.

Magellan by (Zweig, 1938, chap. 10).

As it happens, once the indications of the surviving crew were incorporated into a new world map (see Ribeiro, 1529), the dimensions of the major continents as well as the unexplored areas were then drawn in correct proportions and more precise contours than Reinel's world map of figure A.1. This was an important milestone in the so-called Age of Discovery. In particular, the map of S.E. Asia and Indonesia was greatly improved, despite mainland Asia having not been visited directly by the expedition. One exception is the localization of the Moluccas Islands, which were still located too far east, mostly for geopolitical reasons (it was a contested location where the sought-after cloves could be found).

In the next section, we discuss how this type of reasoning, that should in our opinion be embedded in spatial inference systems, can be emulated for SLAM problems. We mention an existing framework to can be employed and outline the remaining challenges ahead.

A.2 Perspective on the representation of missing observations

What is described in the Magellan story fall under the ability to reason about the *causes* of missing observation. Modern causal inference framework have shown to be appropriate to address the cause of missing data, typically for discrete value problems as developed by Mohan et al. (2013) (see also Mohan et al., 2022). Thus, the CARLIT framework (chap. 4), as a specialization of a causal inference framework for SLAM, is a potential candidate to represent this type of inference. We discuss that perspective on a toy SLAM scenario and the foreseen limitations.

Box 12: SLAM Scenario - Missing Data

From a starting pose, arbitrarily denoted $\mathbf{x_0}$, a robot observes 5 relative motion events $\mathcal{U} \triangleq \{\mathbf{z}_0^1, \mathbf{z}_1^2, \mathbf{z}_2^3, \mathbf{z}_3^4, \mathbf{z}_4^5\}$. A first landmark, denoted l_1 is discovered (no prior knowledge) by a robot exteroceptive sensor at time t_1. The landmark is not re-observed. Robot motion and sensor measurements $\mathcal{Z} \triangleq \{z_0, z_1, z_2, z_3\}$ are affected by noise. Furthermore, we assume that the realization of the noises puts the estimated position of the robot in expected range of the landmark l_1 when in the estimated position $\hat{\mathbf{x}}_\mathbf{5}$ (see figure A.2). We assume no adverse effect due to the robot inertia, false sightings, outliers, etc...

Objective: The inference problem consists in establishing the state estimation of the robot poses $\mathcal{X} \triangleq \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{x_4}, \mathbf{x_5}\}$ and the landmark position $\mathcal{L} \triangleq \{l_1\}$.

Box 12 describes a scenario where only one landmark l_1 is observed and, after a curve observed by poor quality odometric measurements, the landmark l_1 is expected to be seen again. Figure A.2 shows a sketch of this scenario where the ground truth trajectory is in green, and the estimated trajectory is in gray. The sighting area of the embedded landmark

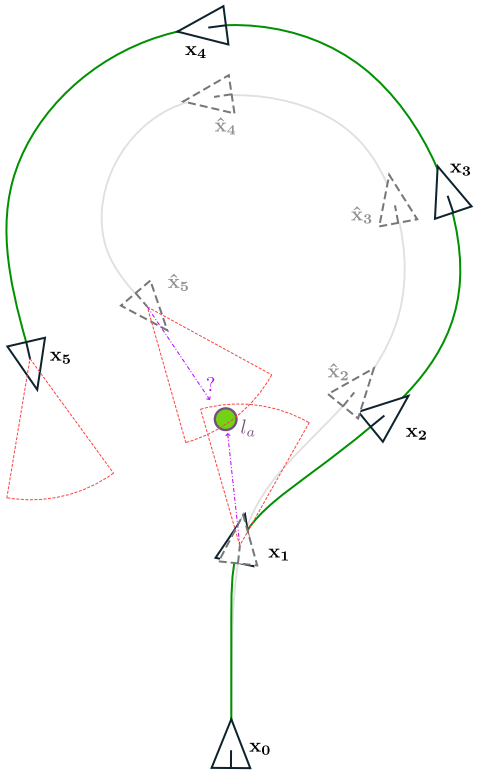


Figure A.2: A toy scenario to motivate the treatment of missing data in SLAM. This figure represents the scenario described box 12. The green path represents the ground truth of the robot trajectory, while the gray path represents the dead-reckoning estimate of the trajectory. The red dashed cones **C** denote the area of sighting at different poses. A landmark l_1 is discovered while the robot is at pose $\mathbf{x_1}$. l_1 is predicted to be re-sighted at the estimated pose $\hat{\mathbf{x}}_\mathbf{5}$. However, no sighting is made (symbol "?") due to the drift accumulated in the trajectory. The issue at stakes here is to consider whether a procedure to bring some corrections to the errors in this setting can and should be proposed, given that, in this scenario, there are no classical loop-closure.

measurement sensor is represented by the red cone **C** (field of view and range). As the estimated pose $\hat{\mathbf{x}}_\mathbf{5}$ is widely inaccurate, no landmark observation of l_1 is possible in those circumstances. Figure A.3 represents the causal graph and, comparatively, factors graph of the problem, without considerations for the missing observation. The causal graph (a) in figure A.3 shows the first landmark sighting cast as an intervention. This representation corresponds to what is described for CARLIT (refer to §4.4). Fig A.3-(a) also exhibits that all errors are statistically independent of one another: therefore no opportunity for adjustment is offered by the model.

However, intuitively, in this situation, assuming the sensor is fully reliable (i.e. no missing-completely-at-random to employ Mohan et al. (2022) terminology) the errors on the odometry measurements of the robot and/or the initial landmark measurement sensor error should be corrected in such a way that the field of view at $\hat{\mathbf{x}}_\mathbf{5}$ does not cover the landmark l_1. In typical state of the art sparse probabilistic SLAM systems, nothing is done to infer such adjustment (as pointed out in chap. 2).

Nonetheless, existing missing-data approach, developed by Mohan et al. (2013) for the SCM framework, should be adapted to warrant the correction of errors, while being mindful of the mechanisms explaining the lack of re-sighting.

To represent this process, the causal graph is completed with additional elements on figure A.4. Let y_1^4 denote the position of landmark l_1 from the robot perspective 4 four steps after its

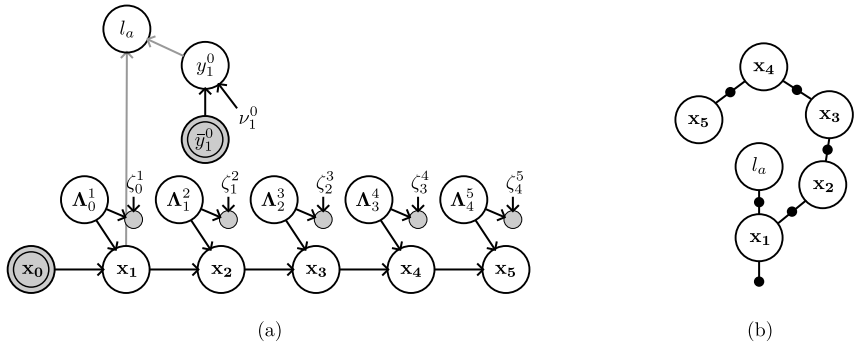


Figure A.3: (a) Causal graph and (b) factors graph representations to the scenario box 12 without considerations to the lack of re-sighting of l_1 (see fig A.2).

discovery. The graph on figure A.4 represents the following additional structural equations:

$$\begin{cases} y_1^4 & := \left(\mathbf{\Lambda_1^5}\right)^{-1} \cdot y_1^0 \\ \bar{y}_1^{4\star} & := \begin{cases} y_1^4 \text{ if } y_1^4 \notin \mathbf{C} \\ m \text{ if } y_1^4 \in \mathbf{C} \text{ and } \bar{y}_1^4 \text{ not available .} \end{cases} \end{cases} \tag{A.1}$$

In equation (A.1), $\bar{y}_1^{4\star}$ is a proxy for y_1^4, the behavior of $\bar{y}_1^{4\star}$ changes depending on whether a sighting should occur. This proxy node appears when a real measurement is not available, i.e. when the structural equation $\bar{y}_1^4 := y_1^4 + \nu_1^4$ does not occur. Then, if the landmark is not predicted to be within the sighting area (i.e. $y_1^4 \notin \mathbf{C}$), the proxy $\bar{y}_1^{4\star}$ 'collapses' onto y_1^4, and y_1^4 is an open collider. In this case, the causal graph is equivalent to figure A.3-(a). Finally, $\bar{y}_1^{4\star}$ takes the value m (as *missing*) if a sighting of l_1 was predicted in this area ($y_1^4 \in \mathbf{C}$).

If $\bar{y}_1^{4\star} = m$, this signals the underlying solver (see 4.3) that the set of errors $\xi \triangleq \{\nu_1^0, \zeta_1^2, \zeta_2^3, \zeta_3^4, \zeta_4^5\}$ should be corrected with the constraint $y_1^4 \notin \mathbf{C}$ (i.e. the landmark does not find itself in the sighting area of $\hat{\mathbf{x}}_\mathbf{5}$):

$$\underset{\xi}{\operatorname{argmin}} \frac{1}{2}\xi^\top \xi \quad \text{s.t.} \quad y_1^4 \notin \mathbf{C}. \tag{A.2}$$

The constraint in equation (A.2) can be formulated, in the case of cone $\mathbf{C}$ as two inequalities, one on the extent of the field of view (angle) and another on the range.

Remark however that there are several roadblocks that would prevent to generalize this missing data specification in SLAM problems. Firstly, the proxy variable, as being an observation node in figure A.4, creates a 'closure' in the sense that it makes all relevant errors mutually dependent (according to the graphical rules of d-separation). In practice, if such proxy variables $\bar{y}^\star$ were to be added at every time step and for every landmark, any benefit from sparsity would be lost, making the optimization problem untractable. A naive way to address the issue is to conduct inference without concern for the missing data and only introduce the proxy variables in a second optimization process, for the variables y_i^j whose estimates fall under $y_i^j \in \mathbf{C}$. But there is no a priori guarantee that the end result would be similar to the dense untractable system. Furthermore, as the optimization process is inherently nonlinear, the set of variables y_i^j affected by a proxy variable might evolve following re-linearization. Secondly, the proposed specification is currently unable to mix with another

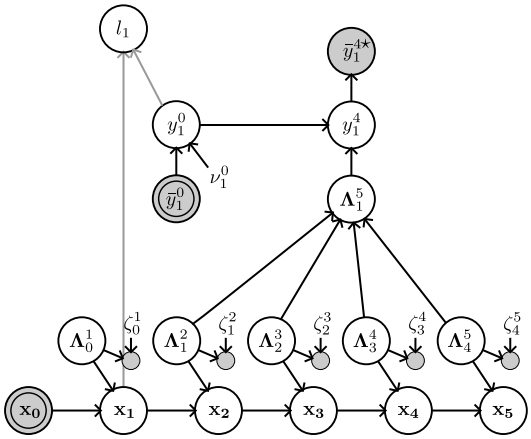


Figure A.4: Causal graph to represent scenario 12, including the exploitation of missing data $\bar{y}_1^{4\star}$.

mechanism behind missing data: the random failure of the sensor (or, equivalently, of the segmentation software), in situations when the landmark is in fact well within the sighting area **C**. Indeed, the proposed scheme assumes perfect occurrence detections and non-detections, not to be confused with sensor noise.

Appendix B

Algorithms for structural recomposition of Λ-motions

Additional algorithms for causal graphs manipulations, used in section 4.5 in support of algorithm 5.

Algorithm 7: Recompose a pair $\left\{\mathbf{\Lambda}_i^j, \mathbf{\Lambda}_\mathbf{j}^\mathbf{k}\right\}$ into $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$ in a directed graph G . If the new node $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$ does not exist, createe that part of the graph It also deals with the reconfiguration of the edges in the intersection I_{ijk} .

Input: Graph G, pair $\left\{\mathbf{\Lambda}_i^j, \mathbf{\Lambda}_\mathbf{j}^\mathbf{k}\right\}$

Output: Composed element $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$, Graph G recomposed

$G' \leftarrow G$; // make a copy

if $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k} \notin G'$ **then**

  /* if Λ_i^k does not exist, create it */

  AddVertices(G', $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}, \mathbf{Z_v}{}_i^k, \xi_i^k$);

  AddEdge(G', $\mathbf{\Lambda}_\mathbf{i}^\mathbf{j} \rightarrow \mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$);

  AddEdge(G', $\mathbf{\Lambda}_\mathbf{j}^\mathbf{k} \rightarrow \mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$);

  AddEdge(G', $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k} \rightarrow \mathbf{Z_v}{}_i^k$);

  AddEdge(G', $\xi_i^k \rightarrow \mathbf{Z_v}{}_i^k$);

end

for $\mathbf{\Lambda} \in \left\{\mathrm{I}_{ijk} \setminus \mathbf{\Lambda}_\mathbf{i}^\mathbf{k}\right\}$ **do**

  /* Outgoing Edges to the intersection set (excluding $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$) are removed and replaced by edges outgoing from $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$ to the intersection set $\left\{\mathrm{I}_{ijk} \setminus \mathbf{\Lambda}_\mathbf{i}^\mathbf{k}\right\}$. */

  RemEdge(G', $\mathbf{\Lambda}_\mathbf{i}^\mathbf{j} \rightarrow \mathbf{\Lambda}$) ; // remove old edges

  RemEdge(G', $\mathbf{\Lambda}_\mathbf{j}^\mathbf{k} \rightarrow \mathbf{\Lambda}$);

  AddEdge(G', $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k} \rightarrow \mathbf{\Lambda}$);

end

return $\mathbf{\Lambda}_\mathbf{i}^\mathbf{k}$, G' ;

Algorithm 8: ChildrenIntersection: returns the intersection of the outedges (children) of 2 vertices in a graph.

Input: Graph G, vertices $\mathbf{\Lambda_i^j}$ & $\mathbf{\Lambda_j^k}$
Output: I_{ijk} intersection set of children

1 $V_1 \leftarrow$ `GetOutEdges(`G, $\mathbf{\Lambda_i^j}$`)` ;
2 $V_2 \leftarrow$ `GetOutEdges(`G, $\mathbf{\Lambda_j^k}$`)` ;
3 $\mathrm{I}_{ijk} \leftarrow V_1 \cap V_2$;
4 **return** I_{ijk}

Algorithm 9: ChildrenUnion: returns the union of the outedges (children) of 2 vertices in a graph.

Input: Graph G, vertices $\mathbf{\Lambda_i^j}$ & $\mathbf{\Lambda_j^k}$
Output: U_{ijk} union set of children

1 $V_1 \leftarrow$ `GetOutEdges(`G, $\mathbf{\Lambda_i^j}$`)` ;
2 $V_2 \leftarrow$ `GetOutEdges(`G, $\mathbf{\Lambda_j^k}$`)` ;
3 $\mathrm{U}_{ijk} \leftarrow V_1 \cup V_2$;
4 **return** U_{ijk}

Bibliography

Agarwal, P. and Olson, E. (2012). Variable reordering strategies for SLAM. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3844–3850, Vilamoura-Algarve, Portugal. IEEE.

Agarwal, P., Tipaldi, G. D., Spinello, L., Stachniss, C., and Burgard, W. (2013). Robust map optimization using dynamic covariance scaling.

Aji, S. and McEliece, R. (2000). The generalized distributive law. *IEEE Transactions on Information Theory*, 46(2):325–343.

Albo, F. (1522). The first voyage round the world/log-book of francisco albo. https://en.wikisource.org/wiki/The_First_Voyage_Round_the_World/Log-Book_of_Francisco_Alvo_or_Alvaro.

Alder, K. (2003). *The measure of all things: the seven-year odyssey and hidden error that transformed the world*. Free Press, New York, 1. ed edition.

Amestoy, P. R., Davis, T. A., and Duff, I. S. (1996). An Approximate Minimum Degree Ordering Algorithm. *SIAM Journal on Matrix Analysis and Applications*, 17(4):886–905.

Arago, F. (1854). *Histoire de ma jeunesse*. C. Bourgois, Paris, France.

Arago, F. (1857). Astronomie populaire. https://fr.wikisource.org/wiki/Astronomie_populaire_(Arago).

Arnborg, S., Corneil, D. G., and Proskurowski, A. (1987). Complexity of finding embeddings in a ¡i¿k¡/i¿ -tree.

Bai, F., Vidal-Calleja, T., and Huang, S. (2018). Robust incremental slam under constrained optimization formulation.

Bareinboim, E., Correa, J. D., Ibeling, D., and Icard, T. (2022). On pearl's hierarchy and the foundations of causal inference. In *Probabilistic and causal inference: the works of judea pearl*, pages 507–556.

Barfoot, T. D. (2017). *State Estimation for Robotics*. Cambridge University Press.

Barrau, A. and Bonnabel, S. (2017). The invariant extended kalman filter as a stable observer.

Bayart, P. (2007). *La méridienne de France et l'aventure de sa prolongation jusqu'aux Baléares*. Acteurs de la science. Harmattan, Paris.

Berthaut, H. M. A. (1898a). La carte de France, 1750-1898 : étude historique. Tome 1. https://bibnum-patrimoniale.univ-grenoble-alpes.fr/items/show/72.

Berthaut, H. M. A. (1898b). La carte de France, 1750-1898 : étude historique. Tome 2. https://bibnum-patrimoniale.univ-grenoble-alpes.fr/items/show/72.

Bertrand (1855). *Gauss - Méthode des moindres carrés.* //fr.wikisource.org/w/index.php?title=Livre:Gauss_-_M%C3%A9thode_des_moindres_carr%C3%A9s,_trad._Bertrand,_1855.djvu&oldid=11551701.

Biot, J. B. (1803). Essai sur l'histoire générale des sciences pendant la révolution française. https://gallica.bnf.fr/ark:/12148/bpt6k9627974d/.

Blair, J. R. and Peyton, B. (1993). An introduction to chordal graphs and clique trees. In *Graph theory and sparse matrix computation*, pages 1–29. Springer.

Bodlaender, H., Gilbert, J., Hafsteinsson, H., and Kloks, T. (1995). Approximating treewidth, pathwidth, frontsize, and shortest elimination tree.

Bouasse, H. (1919). *Géographie mathématique.* Bibliothèque scientifique de l'ingénieur et du physicien. Delagrave. https://archive.org/details/gographiemath00boua/mode/2up.

Bourmaud, G., Mégret, R., Arnaudon, M., and Giremus, A. (2015). Continuous-Discrete Extended Kalman Filter on Matrix Lie Groups Using Concentrated Gaussian Distributions. *Journal of Mathematical Imaging and Vision*, 51(1):209–228.

Brezinski, C. (2018). La méthode de Cholesky. *Revue d'histoire des mathématiques.*

Brezinski, C. and Gross-Cholesky, M. (2005). La vie et les travaux d'André Louis Cholesky. *Bulletin de la Sabix*, (39):6–32. http://journals.openedition.org/sabix/521.

Brooks, R. (1985). Visual map making for a mobile robot. In *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, volume 2, pages 824–829.

Brossard, M., Barrau, A., Chauchat, P., and Bonnabel, S. (2022). Associating uncertainty to extended poses for on lie group imu preintegration with rotating earth.

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., and Leonard, J. J. (2016). Past, Present, and Future of Simultaneous Localization and Mapping: Toward the Robust-Perception Age. *IEEE Transactions on Robotics*, 32(6):1309–1332.

Carlevaris-Bianco, N., Kaess, M., and Eustice, R. M. (2014). Generic node removal for factor-graph slam.

Carlone, L. and Censi, A. (2014). From angular manifolds to the integer lattice: Guaranteed orientation estimation with application to pose graph optimization. *IEEE Transactions on Robotics*, 30(2):475–492.

Carlone, L., Kira, Z., Beall, C., Indelman, V., and Dellaert, F. (2014). Eliminating conditionally independent sets in factor graphs: A unifying perspective based on smart factors. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4290–4297, Hong Kong, China. IEEE.

Carlone, L., Tron, R., Daniilidis, K., and Dellaert, F. (2015). Initialization techniques for 3D SLAM: A survey on rotation estimation and its use in pose graph optimization. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4597–4604, Seattle, WA, USA. IEEE.

Cassini, Méchain, Pierre, and Legendre, André-Marie (1787). Exposé des opérations faites en France, en 1787, pour la jonction des observatoires de Paris et de Greenwich, par MM. Cassini, Méchain et Le Gendre. https://gallica.bnf.fr/ark:/12148/bpt6k6587528k/.

Castro, X. d. and Bernand, C. (2010). *Le voyage de Magellan, 1519-1522: la relation d'Antonio Pigafetta & autres témoignages*. Magellane. Chandeigne, 2e éd edition.

Celmiņš, A. (1998). The method of gauss in 1799. *Statistical Science*, 13(2):123–135.

Chandrasekaran, V., Srebro, N., and Harsha, P. (2008). Complexity of inference in graphical models.

Chatila, R. and Laumond, J.-P. (1985). Position referencing and consistent world modeling for mobile robots.

Chen, Y., Davis, T. A., Hager, W. W., and Rajamanickam, S. (2008). Algorithm 887: CHOLMOD, Supernodal Sparse Cholesky Factorization and Update/Downdate. *ACM Transactions on Mathematical Software*, 35(3):1–14.

Chirikjian, G. S. (2009). *Stochastic Models, Information Theory, and Lie Groups, Volume 1*. Birkhäuser Boston, Boston.

Chirikjian, G. S. (2012). *Stochastic Models, Information Theory, and Lie Groups, Volume 2*. Birkhäuser Boston, Boston.

Cholesky, A. L. (1910). Sur la resolution numerique des systèmes d'équations linéaires. https://www.sabix.org/bulletin/b39/resolution.html.

Choudhary, S., Carlone, L., Nieto, C., Rogers, J., Christensen, H. I., and Dellaert, F. (2016). Distributed trajectory estimation with privacy and communication constraints: A two-stage distributed gauss-seidel approach.

Cieslewski, T., Choudhary, S., and Scaramuzza, D. (2018). Data-efficient decentralized visual slam.

Condamine, L. (1745). *Journal du voyage fait par ordre du Roi à l'Equateur servant d'introduction historique à la mesure des trois premiers degrés du méridien*. https://gallica.bnf.fr/ark:/12148/bpt6k1051290z.

Condamine, L. (1748). *Nouveau projet de mesure invariable propre à servir de mesure commune à toutes les nations*. https://www.academie-sciences.fr/pdf/dossiers/Reaumur/Reaumur_pdf/p489_514_vol3545m.pdf.

Condamine, L. (1751). *Mesure des trois premiers degrés du méridien dans l'hémisphère austral , tirée des observations de MM. de l'Académie royale des sciences, envoyés par le roi sous l'Équateur, par M. de La Condamine*. https://gallica.bnf.fr/ark:/12148/bd6t57723433#.

Cooper, G. F. (1990). The computational complexity of probabilistic inference using bayesian belief networks.

Crepel, P. (1989). De condorcet à arago : l'enseignement des probabilités en france de 1786 à 1830. http://journals.openedition.org/sabix/562.

Cunningham, A., Indelman, V., and Dellaert, F. (2013). DDF-SAM 2.0: Consistent distributed smoothing and mapping. In *2013 IEEE International Conference on Robotics and Automation*, pages 5220–5227, Karlsruhe, Germany. IEEE.

Darwiche, A. (2009). *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, 1 edition.

Davis, C. (1857a). Theoria motus corporum coelestium in sectionibus conicis solem ambientium.

Davis, C. H. (1857b). Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections: A Translation of Gauss's "Theoria Motu.

Davis, T. A. (2004). A column pre-ordering strategy for the unsymmetric-pattern multifrontal method. *ACM Transactions on Mathematical Software*, 30(2):165–195.

Davis, T. A. (2011). Algorithm 915, SuiteSparseQR: Multifrontal multithreaded rank-revealing sparse QR factorization. *ACM Transactions on Mathematical Software*, 38(1):1–22.

Davis, T. A. and Hager, W. W. (2009). Dynamic Supernodes in Sparse Cholesky Update/Downdate and Triangular Solves. *ACM Transactions on Mathematical Software*, 35(4):1–23.

Davison, A. J. and Ortiz, J. (2022). FutureMapping 2: Gaussian Belief Propagation for Spatial AI. arXiv:1910.14139 [cs].

Dawid, A. P. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 41(1):1–15. Publisher: Oxford University Press.

Delambre, J.-B. (1806). Tome 1. base du système métrique décimal. https://gallica.bnf.fr/ark:/12148/bpt6k110604s.

Delambre, J.-B. (1807). Tome 2. base du système métrique décimal. https://gallica.bnf.fr/ark:/12148/bpt6k56994307.

Delambre, J.-B. (1810). Tome 3. base du système métrique décimal. https://gallica.bnf.fr/ark:/12148/bpt6k1106055.

Delambre, J.-B. (1817). Grandeur et figure de la terre; ouvrage augmenté de notes, de cartes et publié par les soins de G. Bigourdan.

Delambre, J.-B. and Legendre, A.-M. (1799). *Méthodes Analytiques pour la Détermination d'un Arc du Méridien*.

Dellaert, F. (2012). Factor Graphs and GTSAM: A Hands-on Introduction. page 26.

Dellaert, F. (2021). Factor Graphs: Exploiting Structure in Robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 4(1):141–166.

Dellaert, F. and Kaess, M. (2006). Square Root SAM: Simultaneous Localization and Mapping via Square Root Information Smoothing. *The International Journal of Robotics Research*, 25(12):1181–1203.

Dellaert, F. and Kaess, M. (2017). Factor Graphs for Robot Perception. *Foundations and Trends in Robotics*, 6(1-2):1–139.

Deray, J. and Solà, J. (2020). Manif: A micro lie theory library for state estimation in robotics applications.

Dhombres, J., editor (2012). *Pierre-Simon de Laplace, 1749-1827: le parcours d'un savant*. Histoire des sciences. Hermann : L'Observatoire de Paris, Paris.

Doherty, K., Baxter, D., Schneeweiss, E., and Leonard, J. (2019). Probabilistic data association via mixture models for robust semantic slam.

Durrant-Whyte, H. and Bailey, T. (2006). Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms. *IEEE Robotics & Automation Magazine*, 13(2):9.

Dutka, J. (1990). Robert Adrain and the method of least squares. *Archive for History of Exact Sciences*, 41(2):171–184. http://link.springer.com/10.1007/BF00411864.

Dutka, J. (1995). On Gauss' Priority in the Discovery of the Method of Least Squares. (Archive for History of Exact Sciences 49):355–70. http://www.jstor.org/stable/41134010.

Duviols, J.-P. and Castro, X. d. (2019). *Idées reçues sur les grandes découvertes: XVe-XVIe siècles*. Magellane poche. Chandeigne, nouvelle éd. remaniée et augmentée edition.

Falguerolles, A. D. and Pinchon, D. (2006). Une commémoration du bicentenaire de la publication (1805-1806) de la méthode des moindres quarrés par Adrien Marie Legendre.

Fan, T., Wang, H., Rubenstein, M., and Murphey, T. (2020). Cpl-slam: Efficient and certifiably correct planar graph-based slam using the complex number representation.

Fisher, R. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222(594-604):309–368. https://royalsocietypublishing.org/doi/10.1098/rsta.1922.0009.

Forster, C., Carlone, L., Dellaert, F., and Scaramuzza, D. (2017). On-Manifold Preintegration for Real-Time Visual-Inertial Odometry. *IEEE Transactions on Robotics*, 33(1):1–21. arXiv: 1512.02363.

Fourie, D., Leonard, J., and Kaess, M. (2016). A nonparametric belief solution to the Bayes tree. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2189–2196, Daejeon, South Korea. IEEE.

Fourier, J. B. J. (1822). *Théorie Analytique de la Chaleur*. Cambridge University Press, 1 edition. https://www.cambridge.org/core/product/identifier/9780511693229/type/book.

Gauss, C. F. .-. (1809). Theoria motus corporum coelestium in sectionibus conicis solem ambientium.

Gauss, C. F. (1826). *Supplementum theoriae combinationis observationum erroribus minimis obnoxiae*.

Golub, G. H. and Plemmons, R. J. (1980). Large-scale geodetic least-squares adjustment by dissection and orthogonal decomposition. *Linear Algebra and Its Applications*, 34:3–28. Publisher: Elsevier.

Grcar, J. F. (2011). How ordinary elimination became Gaussian elimination. *Historia Mathematica*, 38(2):163–218.

Grisetti, G., Guadagnino, T., Aloise, I., Colosi, M., Della Corte, B., and Schlegel, D. (2020). Least Squares Optimization: From Theory to Practice. *Robotics*, 9(3):51.

Grisetti, G., Kummerle, R., Stachniss, C., and Burgard, W. (2010a). A Tutorial on Graph-Based SLAM. *IEEE Intelligent Transportation Systems Magazine*, 2(4):31–43.

Grisetti, G., Kummerle, R., Stachniss, C., Frese, U., and Hertzberg, C. (2010b). Hierarchical optimization on manifolds for online 2d and 3d mapping.

Guedj, D. (2000). *Le mètre du monde*. Number 1059 in Points. Seuil, Paris.

Haggstrom, P. (2015). The central limit theorem - how laplace actually proved it.

Hald, A. (2007). *A history of parametric statistical inference from Bernoulli to Fischer: 1713-1935*. Sources and studies in the history of mathematics and physical sciences. Springer, New York.

Haüy, R. (1793). Instruction sur les mesures déduites de la grandeur de la terre, uniformes pour toute la république / , et sur les calculs relatifs à leur division décimale, par la commission temporaire des poids et mesures républicaines. https://archive.org/details/instructionsurl00fran/page/xiv/mode/2up.

Hess, W., Kohler, D., Rapp, H., and Andor, D. (2016). Real-time loop closure in 2d lidar slam.

Hofmann-Wellenhof, B. and Moritz, H. (2006). *Physical geodesy*. SpringerWienNewYork, Wien ; New York, 2nd, corrected ed edition.

Holmes, C. and Barfoot, T. D. (2022). An efficient global optimality certificate for landmark-based slam.

Hsiao, M. and Kaess, M. (2019). Mh-isam2: Multi-hypothesis isam using bayes tree and hypo-tree.

Hughes, N., Chang, Y., and Carlone, L. (2022). Hydra: A Real-time Spatial Perception System for 3D Scene Graph Construction and Optimization. In *Robotics: Science and Systems XVIII*. Robotics: Science and Systems Foundation.

Ila, V., Polok, L., Solony, M., Smrz, P., and Zemcik, P. (2015). Fast covariance recovery in incremental nonlinear least square solvers. In *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4636–4643, Seattle, WA, USA. IEEE.

Ila, V., Polok, L., Solony, M., and Svoboda, P. (2017). SLAM++ <sup/> -A highly efficient and temporally scalable incremental SLAM framework. *The International Journal of Robotics Research*, 36(2):210–230.

Jozeau, M.-F. (1997). *Géodésie au XIXème siècle. De l'hégémonie française à l'hégémonie allemande. Regards belges. Compensation et méthode des moindres carrés.* Theses, Univeristé Paris Diderot Paris 7. https://theses.hal.science/tel-01252747.

Kaess, M., Ila, V., Roberts, R., and Dellaert, F. (2010). The Bayes Tree: An Algorithmic Foundation for Probabilistic Robot Mapping. In Siciliano, B., Khatib, O., Groen, F., Hsu, D., Isler, V., Latombe, J.-C., and Lin, M. C., editors, *Algorithmic Foundations of Robotics IX*, volume 68, pages 157–173. Springer Berlin Heidelberg, Berlin, Heidelberg.

Kaess, M., Johannsson, H., Roberts, R., Ila, V., Leonard, J., and Dellaert, F. (2011). iSAM2: Incremental smoothing and mapping with fluid relinearization and incremental variable reordering. In *2011 IEEE International Conference on Robotics and Automation*, pages 3281–3288, Shanghai, China. IEEE.

Kaess, M., Ranganathan, A., and Dellaert, F. (2008). iSAM: Incremental Smoothing and Mapping. *IEEE Transactions on Robotics*, 24(6):1365–1378.

Kai Ni and Dellaert, F. (2010). Multi-level submap based SLAM using nested dissection. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2558–2565, Taipei. IEEE.

Kepner, J. V. and Gilbert, J. R. (2011). *Graph algorithms in the language of linear algebra*. Software, environments, and tools. Society for Industrial and Applied Mathematics, Philadelphia.

Koller, D. and Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. Adaptive computation and machine learning. MIT Press, Cambridge, MA.

Kožnjak, B. (2015). Who let the demon out? Laplace and Boscovich on determinism. *Studies in History and Philosophy of Science Part A*, 51:42–52.

Kožnjak, B. (2022). God and Boscovich's Demon. *The European Legacy*, 27(1):39–56.

Kummerle, R., Grisetti, G., Strasdat, H., Konolige, K., and Burgard, W. (2011). G2o: A general framework for graph optimization. In *2011 IEEE International Conference on Robotics and Automation*, pages 3607–3613, Shanghai, China. IEEE.

Laplace, P. S. (1796). *Exposition du système du monde*, volume 1. Courcier.

Laplace, P.-S. (1814a). Essai philosophique sur les probabilités. https://fr.wikisource.org/wiki/Essai_philosophique_sur_les_probabilitÃ¯s.

Laplace, P.-S. (1814b). A Philosophical Essay on Probabilities. https://en.wikisource.org/wiki/A_Philosophical_Essay_on_Probabilities.

Laplace, P.-S. (1820). *Théorie analytique des probabilités*, volume Tome 7 of *Oeuvres Complètes de Laplace*. Éd. Gabay, Paris, courcier edition. https://gallica.bnf.fr/ark:/12148/bpt6k775950/f706.item.

Lauritzen, S. L. and Spiegelhalter, D. J. (1988). Local computations with probabilities on graphical structures and their application to expert systems.

Legendre, A.-M. (1805). Nouvelles Méthodes pour la Détermination des Orbites Des Comètes. https://gallica.bnf.fr/ark:/12148/bpt6k15209372/f84.item.

Lepar, V. and Shenoy, P. P. (2003). A Comparison of Lauritzen-Spiegelhalter, Hugin, and Shenoy-Shafer Architectures for Computing Marginals of Probability Distributions. *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 228–337.

Levallois, J.-J. (1988). *Mesurer la terre: 300 ans de géodesie française; de la toise du Châtelet au satellite*. Presses de l'Ecole Nationale des Ponts et Chaussées [u.a.], Paris.

Liagre, J. B. J. (1852). *Calcul des probabilités et théorie des erreurs*. Société pour l'Émancipation Intellectuelle. https://gallica.bnf.fr/ark:/12148/bpt6k9646757f.

Lu, F. and Milios, E. (1997). Robot pose estimation in unknown environments by matching 2d range scans. *Journal of Intelligent and Robotic systems*, 18:249–275. Publisher: Springer.

Maire, C. and Boscovich, R. (1755). Voyage astronomique et géographique dans l'État de l'Église, entrepris par l'ordre et sous les auspices du pape benoît xiv, pour mesurer deux degrés du méridien et corriger la carte de l'État ecclésiastique, par les pp. maire et boscovich.

Mansuy, R. (2008). Analyse du manuscrit d'André-Louis Cholesky intitulé "Sur la résolution numérique des systèmes d'équations linéaires". *BibNum*. http://journals.openedition.org/bibnum/618.

Marat, J.-P. (1791). Les charlatans modernes, ou lettres sur le charlatanisme académique. https://fr.wikisource.org/wiki/Les_Pamphlets_de_Marat/Les_Charlatans_modernes.

Martinec, D. and Pajdla, T. (2007). Robust rotation and translation estimation in multiview reconstruction. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

Mohan, K., Geffner, H., Dechter, R., and Halpern, J. Y. (2022). Causal graphs for missing data: A gentle introduction.

Mohan, K., Pearl, J., and Tian, J. (2013). Graphical models for inference with missing data.

Montemerlo, M., Thrun, S., Koller, D., and Wegbreit, B. (2002). FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem. *Eighteenth national conference on Artificial intelligence*, page 6.

Mukadam, M., Dong, J., Yan, X., Dellaert, F., and Boots, B. (2018). Continuous-time gaussian process motion planning via probabilistic inference.

Noël, G. and Tilleuil, P. (2005). D'où sort la méthode des moindres carrés.

Olson, E., Leonard, J., and Teller, S. (2006). Fast iterative alignment of pose graphs with poor initial estimates. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006.*, pages 2262–2269, Orlando, FL, USA. IEEE.

Ortiz, J., Evans, T., and Davison, A. J. (2021). A visual introduction to Gaussian Belief Propagation. arXiv:2107.02308 [cs].

Ortiz, J., Evans, T., Sucar, E., and Davison, A. J. (2022). Incremental abstraction in distributed probabilistic slam graphs.

Paskin, M. A. (2002). Thin junction tree filters for simultaneous localization and mapping. *Computer Science Division, University of California.* Report No. UCB/CSD-02-1198. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=5e906275a2e61ef2bc7b2c8c35325a1540b751cf.

Paskin, M. A. and Lawrence, G. D. (2003). *Junction tree algorithms for solving sparse linear systems.* Computer Science Division, University of California. Report No. UCB/CSD-03-1271. https://ai.stanford.edu/~paskin/pubs/csd-03-1271.pdf.

Paull, L., Seto, M., and Leonard, J. J. (2014). Decentralized cooperative trajectory estimation for autonomous underwater vehicles.

Pearl, J. (1985). Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of the 7th conference of the Cognitive Science Society, University of California, Irvine, CA, USA*, pages 15–17.

Pearl, J. (2009). *Causality.* Cambridge University Press.

Pearl, J. (2012). The do-calculus revisited. *arXiv preprint arXiv:1210.4852.*

Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning.

Pearl, J. (2022). Causal inference: History, perspectives, adventures, and unification (an interview with judea pearl).

Pearl, J. and Dechter, R. (2013). Identifying independencies in causal graphs with feedback. *arXiv preprint arXiv:1302.3595.*

Pearl, J., Glymour, M., and Jewell, N. P. (2021). *Causal inference in statistics: a primer.* Wiley, Chichester, reprinted with revisions edition.

Pearl, J. and Mackenzie, D. (2020). *The book of why: the new science of cause and effect.* Basic Books, New York, first trade paperback edition edition.

Pearl, J. and Paz, A. (1985). Graphoids: A graph-based logic for reasoning about relevance relations.

Pigafetta, A. (1522). The first voyage round the world/pigafetta's account of magellan's voyage. https://en.wikisource.org/wiki/The_First_Voyage_Round_the_World/Pigafetta%27s_Account_of_Magellan%27s_Voyage#cite_ref-112.

Pinies, P., Paz, L. M., Haner, S., and Heyden, A. (2012). Decomposable Bundle Adjustment using a junction tree. In *2012 IEEE International Conference on Robotics and Automation*, pages 1246–1253, St Paul, MN, USA. IEEE.

Poincaré, H. (1900). Rapport sur le projet de révision de l'arc méridien de quito. http://rcin.org.pl/impan/Content/215307/PDF/WA35_240685_6642-8_Art40.pdf.

Polok, L., Ila, V., Solony, M., Smrz, P., and Zemcik, P. (2013). Incremental Block Cholesky Factorization for Nonlinear Least Squares in Robotics. In *Robotics: Science and Systems IX*. Robotics: Science and Systems Foundation.

Puissant, L. (1842). *Traité de géodésie: ou, Exposition des méthodes trigonométriques et astronomiques, applicables à la mesure de la terre, et à la construction du canevas des cartes topographiques*. https://g.co/kgs/Mt73pnu.

Ranganathan, A., Kaess, M., and Dellaert, F. (2007). Loopy sam. In *Proceedings of the 20th international joint conference on Artifical intelligence*, pages 2191–2196.

Reinel, J. (1519). World map of jorge reinel. From a copy of Progel, Otto (1815-1887). Accessible at https://gallica.bnf.fr/ark:/12148/btv1b52523067m#.

Ribeiro, D. (1529). Diego ribeiro's globe for the spanish crown. https://gallica.bnf.fr/ark:/12148/btv1b53023022k/f1.item.r=Diogo%20Ribeiro.

Rosen, D. M., Carlone, L., Bandeira, A. S., and Leonard, J. J. (2019). SE-Sync: A certifiably correct algorithm for synchronization over the special Euclidean group. *The International Journal of Robotics Research*, 38(2-3):95–125.

Rosen, D. M., Doherty, K. J., Espinoza, A. T., and Leonard, J. J. (2021). Advances in Inference and Representation for Simultaneous Localization and Mapping. *arXiv:2103.05041 [cs]*. arXiv: 2103.05041.

Rosinol, A., Violette, A., Abate, M., Hughes, N., Chang, Y., Shi, J., Gupta, A., and Carlone, L. (2021). Kimera: from SLAM to Spatial Perception with 3D Dynamic Scene Graphs. arXiv:2101.06894 [cs].

Schön, T. B. and Lindsten, F. (2011). Manipulating the multivariate gaussian density. *Div. Automat. Control, Linköping Univ., Linkping, Sweden, Tech. Rep*, 4(3.4):4.

Shenoy, P. P. and Shafer, G. (1990). Axioms for Probability and Belief Function Propagation.

Shimony, S. E. (1994). Finding maps for belief networks is np-hard.

Smith, R., Self, M., and Cheeseman, P. (1988). A Stochastic Map For Uncertain Spatial Relationships. *Proc. of the Intl. Symp. of Robotics Research (ISSR)*, pages 467–474.

Smith, R. C. and Cheeseman, P. (1986). On the Representation and Estimation of Spatial Uncertainty. *The International Journal of Robotics Research*, 5(4):56–68.

Sola, J. (2014). Simulataneous localization and mapping with the extended kalman filter.

Solà, J., Deray, J., and Atchuthan, D. (2021). A micro Lie theory for state estimation in robotics. arXiv:1812.01537 [cs].

Stachniss, C., Leonard, J. J., and Thrun, S. (2016). 46. simultaneous localization and mapping.

Stahl, S. (2006). The evolution of the normal distribution. *Mathematics magazine*, 79(2):96–113.

Stigler, S. M. (1981). Gauss and the Invention of Least Squares. *The Annals of Statistics*, 9(3).

Stigler, S. M. (1986). *The history of statistics: the measurement of uncertainty before 1900*. Belknap Pr. of Harvard Univ. Pr, Cambridge, Mass., 9. print edition.

Stigler, S. M. (2002). *Statistics on the table: the history of statistical concepts and methods*. Harvard Univ. Press, Cambridge, Mass., 3. printing edition.

Strasdat, H., Montiel, J., and Davison, A. J. (2010). Scale Drift-Aware Large Scale Monocular SLAM. In *Robotics: Science and Systems*, volume 2, page 5. Issue: 3.

Strasdat, H., Montiel, J., and Davison, A. J. (2012). Visual SLAM: Why filter? *Image and Vision Computing*, 30(2):65–77.

Sunderhauf, N. and Protzel, P. (2012). Switchable constraints for robust pose graph slam.

Sunderhauf, N. and Protzel, P. (2013). Switchable constraints vs. max-mixture models vs. rrr - a comparison of three approaches to robust pose graph slam.

Tardi, P. (1934). *Traité de géodésie*. LIBRAIRE DU BUREAU DES LONGITUDES, DE L'ÉCOLE POLYTECHNIQUE, Quai des Grands-Augustins, 55, gauthier-villars edition.

Tari, J. (2024). *Causal Representation of the Simultaneous Localization And Mapping Problem*. Phd thesis, Université de Toulouse.

Thrun, S., Burgard, W., and Fox, D. (2005). *Probabilistic robotics*. Intelligent robotics and autonomous agents. MIT Press, Cambridge, Mass. OCLC: ocm58451645.

Tian, Y., Khosoussi, K., Rosen, D. M., and How, J. P. (2021). Distributed certifiably correct pose-graph optimization.

Torge, W. (2023). *Geodesy*. De Gruyter Textbook. De Gruyter Oldenbourg, Boston, 5th edition.

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle Adjustment — A Modern Synthesis. In Goos, G., Hartmanis, J., van Leeuwen, J., Triggs, B., Zisserman, A., and Szeliski, R., editors, *Vision Algorithms: Theory and Practice*, volume 1883, pages 298–372. Springer Berlin Heidelberg, Berlin, Heidelberg.

Valette, P. (2007). La base de Perpignan. La mesure des bases de la méridienne de France (1792-1798). *Revue XYZ*, 110(1er trimestre):16–18.

Vallvé, J., Solà, J., and Andrade-Cetto, J. (2018). Graph slam sparsification with populated topologies using factor descent optimization.

Vallvé, J., Solà, J., and Andrade-Cetto, J. (2019). Pose-graph slam sparsification using factor descent.

Van Strien, M. (2014). On the origins and foundations of Laplacian determinism. *Studies in History and Philosophy of Science Part A*, 45:24–31.

Vaníček, P. and Foroughi, I. (2019). How gravity field shortened our metre. *Journal of Geodesy*, 93(9):1821–1827. http://link.springer.com/10.1007/s00190-019-01257-7.

Verne, J. (1872). *Aventures de trois Russes et de trois Anglais dans l'Afrique australe*. OCLC: 1309918893.

Vincent, R. (1998). Il Y A 200 ANS ... du 26 avril au 5 juin 1798, Delambre mesurait la base géodésique de Lieusaint à Melun. *REVUE XYZ*, 75(2e trimestre):95–98.

Wang, J. and Olson, E. (2016). Apriltag 2: Efficient and robust fiducial detection.

Wang, X., Marcotte, R., Ferrer, G., and Olson, E. (2018). AprilSAM: Real-Time Smoothing and Mapping. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2486–2493, Brisbane, QLD. IEEE.

Wikipedia Commons, W. (2023). Hanover: General overview of the triangular systems measured by Carl Friedrich Gauß 1821–1844, sheet from the Papen Atla. https://commons.wikimedia.org/w/index.php?title=File:Dreieckssystem-%C3%9Cbersicht_Gau%C3%9F-Ausstellung_Bomann-Museum_(1).JPG&oldid=734167634.

Yang, H., Antonante, P., Tzoumas, V., and Carlone, L. (2020). Graduated non-convexity for robust spatial perception: From non-minimal solvers to global outlier rejection.

Yang, S., Chen, G., Zhang, Y., Choset, H., and Dellaert, F. (2021). Equality constrained linear optimal control with factor graphs.

Ycart, B. (2001). Le procès des étoiles entre De Moivre et Laplace. *CUBO, A Mathematical Journal*, 3(1):1–11.

Zhang, Z., Gallego, G., and Scaramuzza, D. (2018). On the comparison of gauge freedom handling in optimization-based visual-inertial state estimation. *IEEE Robotics and Automation Letters*, 3(3):2710–2717. Publisher: IEEE.

Zweig, S. (1938). *Magellan*. Pushkin.

Ćesić, J., Marković, I., Bukal, M., and Petrović, I. (2017). Extended information filter on matrix Lie groups. *Automatica*, 82:226–234.