Certifiable algorithms for 3D computer vision problems

AUTOR: Mercedes García Salguero

DIRECTOR: Javier González Jiménez

Tesis doctoral por compendio de publicaciones
Programa de Doctorado en Ingeniería Mecatrónica
Dept. de Ingeniería de Sistemas y Automática
Universidad de Málaga

UNIVERSIDAD DE MÁLAGA

AUTORA: María de las Mercedes García Salguero

https://orcid.org/0000-0002-3382-5872

EDITA: Publicaciones y Divulgación Científica. Universidad de Málaga

DECLARACIÓN DE AUTORÍA Y ORIGINALIDAD DE LA TESIS PRESENTADA PARA OBTENER EL TÍTULO DE DOCTOR

D./Dña María de las Mercedes García Salguero
Estudiante del programa de doctorado de Ingeniería Mecatrónica de la Universidad de Málaga, autor/a de la tesis, presentada para la obtención del título de doctor por la Universidad de Málaga, titulada: Certifiable algorithms for 3D computer vision problems (Algoritmos certificadores para problemas de visión por ordenador)

Realizada bajo la tutorización de Antonio Javier González Jiménez y dirección de Antonio Javier González Jiménez (si tuviera varios directores deberá hacer constar el nombre de todos)

DECLARO QUE:

La tesis presentada es una obra original que no infringe los derechos de propiedad intelectual ni los derechos de propiedad industrial u otros, conforme al ordenamiento jurídico vigente (Real Decreto Legislativo 1/1996, de 12 de abril, por el que se aprueba el texto refundido de la Ley de Propiedad Intelectual, regularizando, aclarando y armonizando las disposiciones legales vigentes sobre la materia), modificado por la Ley 2/2019, de 1 de marzo.

Igualmente asumo, ante a la Universidad de Málaga y ante cualquier otra instancia, la responsabilidad que pudiera derivarse en caso de plagio de contenidos en la tesis presentada, conforme al ordenamiento jurídico vigente.

En Málaga, a 31 de Octubre de 2023

Fdo.: María de las Mercedes García Salguero Doctoranda	Fdo.: Antonio Javier González Jiménez Tutor y Director de tesis

UNIVERSIDAD DE MÁLAGA

DEPARTAMENTO DE
INGENIERÍA DE SISTEMAS Y AUTOMÁTICA

El Dr. D. Javier González Jiménez director de la tesis titulada Certifiable algorithms for 3D computer vision problems realizada por Mercedes García Salguero, certifica su idoneidad para la obtención del título de Doctor en Ingeniería Mecatrónica.

Málaga, 17 de Enero de 2024

_____

Dr. D. Javier González Jiménez

To all the people
who have made it
possible for me

to get here

Contents

Abstract

The integration of cameras into a wide variety of devices, from mobile phones to robots, has boosted the development of efficient and reliable algorithms to exploit the visual data they provide. Whereas different tasks require specific solutions, two of them are among the most relevant in geometric computer vision, namely the localization of the camera and the 3D reconstruction of the scene. In fact, they are the cornerstone of more generic pipelines such as Visual Odometry (VO), Structure-from-Motion (SfM) and Simultaneous Localization and Mapping (SLAM), which encounter applicability in paramount fields such as self-driving cars, intelligent robots, augmented reality, or autonomous aerial systems. Despite the vast literature dealing with these two tasks, most of the proposed approaches come with a critical weakness: the found solution may not be the global optimal one, but a local minimum. This is so because the problem is stated through non-convex formulation that is usually solved with iterative algorithms, which are prone to get trapped in local minima. This issue becomes even more crucial because this non-optimality is unnoticed and may have an unpredictable effect on the performance and reliability of the whole pipeline. This thesis contributes a set of efficient algorithms that certify the optimality of the solution for these two relevant tasks under different assumptions of the camera motion and scene configuration.

First, we address the relative pose problem between two calibrated cameras and propose a series of certifiable algorithms that estimate and certify the solutions with different certification ratios and computational times. These results motivate the second set of contributions where fast certifiable algorithms are proposed for the relative pose when the axis of rotation is known for general and planar configurations. Our third contribution addresses the absolute pose for central and noncentral cameras and we provide a fast certifiable al-

gorithm that certifies solutions even for random problem instances. Our last contribution tackles the triangulation problem for both two and N views for both planar and general configurations. For the minimal case with two views, we propose three different certifiable algorithms with different numbers of detected optimal solutions and computational times. For the nonminimal case with N views for both the general and planar scene, we propose a fast certifier empirically obtained in closed-form.

Resumen

La incorporación de cámaras en prácticamente todos los dispositivos, desde teléfonos móviles hasta robots, ha motivado el desarrollo de algoritmos eficientes y fiables. La versatilidad de las cámaras se traduce en la misma variablidad de las aplicciones y sistemas que las utilizan, aunque hay dos tareas principales que suelen requrerirse en todos estos sistemas: localizar la cámara y reconstruir la escena que se ve en las imágenes. De hecho, estos dos problemas son la base en la que se construyen sistemas más complejos aunque genéricos, como *visual odometry* (VO), *Simultaneous Localization and Mapping* (SLAM) o *Structure-from-Motion* (SfM). A pesar de la amplia literartura disponible para estos dos problemas, la mayoría de los trabajos actuales sufren de un problema común: la nonconvexidad del problemas hace que haya, en general, varios mínimos locales. Los algoritmos iterativos que comúmente se usan para resolver estos problemas pueden devolver cualquiera de estos mínimos, no necesariamente la solución global. A priori, no hay manera de saber cómo de lejos los mínimos locales están del global y esta diferencia en calidad de la solución puede afectar a otros bloques que las usen e inluso al resultado final del sistema. En esta tesis proponemos un conjunto de algoritmos que son capaces de certificar soluciones óptimas para estas dos líneas de problemas, y en particular consideramos la especialización de estos bajo distintas suposiciones en cuanto a movimiento de la cámara y distribución de los puntos en el entorno.

Nuestra primera contribución se centra en el problema de la *pose* (rotación y traslación) relativa entre dos cámaras calibradas en la que proponemos una serie de algoritmos certificadores con distintos ratios de certificación y coste computacional. Basándonos en nuestras observaciones en estos trabajos, nuestra siguiente contribución es un algoritmo certificador iterativo para el problema de la pose, pero en este caso bajo la suposición de que el eje de gravedad

es conocido, por ejemplo, por los datos de una *Inertial Measurement Unit* (IMU). Nuestra propuesta considera las configuraciones con un entorno general y aquellas en las que todos los puntos 3D pertenencen a un plano desconocido. Nuestra tercera contribución y la última en el tema de la *pose* considera el problema de la *pose* absoluta entre una cámara y un sistema de referencia. La propuesta se extiende a cámaras centrales y no-centrales, y es incluso capaz de certificar problemas con datos aleatorios. La última contribución de esta tesis aborda el problema de triangulación, contribuyendo a las configuraciones con dos vistas con puntos coplanares y N vistas (general) para puntos en configuración general y coplanares. Para el primer caso, proponemos tres algoritmos distintos que obtienen y/o certifican la solución a este problema, y cuyas principales diferencias radican en el número de soluciones óptimas detectadas y el coste computacional. Para el case no-minimo con N vistas/cámaras/imágenes, y dado el alto número de restricciones, nuestras propuestas optan por obtener y certificar de manera rápida la solución. Empíricamente observamos que el certificador trabaja con una formulación cerrada, reduciendo considerablemente el tiempo necesario para certificar soluciones.

Introducción

La capacidad de percibir el entorno es esencial para cualquier sistema autónomo, y aunque existan multitud de sensores capaces de proveer con esta información, debemos destacar las cámaras por su prevalencia en la mayoría de estos sistemas. Su bajo coste económico, variedad y disponibilidad junto con la cantidad de información relevante que se puede extraer de las imágenes, tanto por el robot como por el operador, han impulsado su incorporación en diferentes ámbitos. La necesidad de procesar y extraer esta información ha hecho que las técnicas de visión por ordenador se hayan desarrollado en múltiples áreas, además de la investigación académica e industrias 'high-tech'. Las especificaciones de estos sistemas han motivado distintas líneas de investigación y desarrollo, y la mejora de métodos ya existentes con un enfoque en precisión, eficiencia y/o restricciones computacionales (memoria, herramientas, librerías, ...). Por ejemplo, en ciertos contextos se exige alta precisión y diseños fiables, como en el robot mostrado en 1.1a que trabaja sobre pestañas, mientras que otros productos se diseñan para operar en entornos no controlados con substancias atípicas, como el aceite de freir en la figura 1.1b.

Las aplicaciones que ejecutan estos sistemas y robots generalmente requieren la inter-conexión de varias tareas, como mostramos en la figura 1.2, y el correcto funcionamiento de la aplicación completa se basa en que cada uno de estos bloques opere bien y obtenga buenas soluciones. Aunque estos bloques pueden ser particulares para cada aplicación, hay dos problemas que suelen aparecer en la mayoría de las aplicaciones basadas en información visual, incluyendo *Visual odometry (V), Structure-from-Motion (SfM)* y *Simultaneous Localization and Mapping (SLAM)*. Estos dos problemas tienen como objetivos respectivos: (a) estimar el movimiento de la(s) cámara(s); y (b) la reconstrucción 3D de la escena. Aunque existen diferentes métodos para resolver estas tareas, en esta tesis nos enfocamos en algoritmos geométricos que utilizan información sobre la imagen (observaciones aisladas) y no la imagen completa. Esta información es procesada por el algoritmo en cuestión para estimar la solución al problema, y decimos que esta estimación es un mínimo/máximo de la función de coste a optimizar. Por tanto, otros costes y en general, formulaciones distintas incluso para el mismo problema pueden tener distintas soluciones. En la mayoría de los casos, los problemas son además no-convexos, por tanto teniendo en general más de un mínimo local. La diferencia entre estos mínimos puede ser grande y los efectos de estas soluciones subóptimas en otros bloques e incluso la solución final es variable, llegando incluso al fallo del sistema completo. Por tanto, ser capaz de garantizar que la solución obtenida por estos bloques es el óptimo global (la mejor estimación para los datos dados) es crucial para el correcto funcionamiento de la aplicación.

Motivación

En general, los problemas que aparecen en sistemas de visión 3D complejos, como SLAM, son no-convexos, y sus estimaciones suelen usarse como entrada de otros bloques computacionales, que también suelen presentar problemas no-convexos. A pesar de esto, la optimalidad global de la solución encontrada raramente se comprueba y la solución final del sistema es la que suele determinar si ésta y los bloques que la componen han funcionado correctamente. Una solución subóptima en cualquiera de estos bloques puede afectar al resto del sistema, especialmente si éstos son sensibles a la inicialización o usan la solución previa para detectar datos correctos (inliers).

Otros trabajos han propuesto soluciones cerradas para estos problemas, aunque la mayoría se centra en algoritmos 'mínimos', esto es, que utilizan olo el mínimo número de datos para estimar una solución. Esto hace que la estimación no se ajuste al resto de observaciones en general, limitando la exactitud de estos algoritmos. Como alternativa encontramos los algoritmos iterativos, que partiendo de una solución inicial, la refinan con todos los datos disponibles. Este tipo de método es común y preferible, especialmente por la observación empírica de que buenas inicializaciones tienden a converger al óptimo global, esto es, el mejor mínimo para los datos disponibles. No obstante, la convergencia a esta solución no se puede garantizar *a priori*, dado que el problema puede presentar varios mínimos locales. Garantizar que la solución final o incluso la inicialización son adecuadas cobra especial importancia, a pesar de ser un problema complejo.

Los métodos 'globales' tratan de resolver las limitaciones de los algoritmos anteriores, garantizando que la solución devuelta es la óptima. Ejemplos de estos metodos son 'Branch and Bound algorithms', y también los algoritmos comentados anteriormente que calculan las soluciones de manera cerrada. Sin embargo, el alto tiempo computacional del primero y la inestabilidad de los segundos motivan el uso de otra familia de métodos globales, como los tratados en esta tesis. Nuestro enfoque se centra en derivar relajaciones convexas de los problemas originales y, específicamente, usamos relajaciones para problemas con coste y restricciones ('constraints') cuadráticas. Aunque no todos los problemas pueden reformularse de manera cuadrática, en esta tesis mostramos que las tareas más comunes en visión por ordenador pueden escribirse de esta manera. Las relajaciones convexas derivadas se pueden resolver con las herramientas disponibles con complejidad polinómica en el número de variables y restricciones. Sin embargo, estos métodos se basan en la suposición de que la relajación es *'tight'*, esto es, la relajación aproxima bien el problema original, lo que en la mayoría de los casos solo puede respaldarse de manera empírica. Cuandos estas relajaciones no son 'tight', es necesario modificar la formulación, y aunque hay ciertos procedimientos que se pueden seguir para asegurar esto, por ejemplo, añadiendo restricciones redundantes pero linealmente independientes, es necesario estudiar el problema y la razón detrás del fallo de la

relajación original. A pesar de admitir varios centenares de variables y restricciones, estas herramientas requieren varios milisegundos para resolver un único problema con menos de 20 variables y 30 restricciones, haciendo que este algoritmo sea demasiado lento para ser utilizado en sistemas de tiempo real. Trabajos previos, por ejemplo, [Burer and Monteiro, 2005] han propuesto métodos alternativos para reducir el coste computacional de estas relajaciones, aunque suponen ciertas condiciones en la formulación que no pueden extenderse a todos los problemas. Como alternativa encontramos los llamados algoritmos certificadores [Bandeira, 2016], que no estiman la solución y solo confirman que solución dada es óptima o no son concluyentes sobre su optimalidad. Sin embargo, la forma esándar de estos algoritmos también pone ciertas limitaciones en el tipo de problemas para los que se pueden derivar, manteniéndose eficientes. La idea fundamental detrás de estos métodos es la observación de que los algoritmos iterativos introducidos anteriormente funcionan bien, especialemente con buenas inicializaciones, y son rápidos. Por tanto, únicamente certificar la solución devuelta como óptima puede reducir el tiempo requerido para resolver el problema con garantías a la vez que mantiene la certificación de optimalidad. Los trabajos dentro de esta familia se centran generalmente en la certificación y asumen que la solución al problema, potencialmente un mínimo local, es dada por otro algoritmo. En esta tesis nos centramos en distintas técnicas de certificación para problems comunes de visión 3D.

Contribución

Esta sección resume primero los objetivos y contribuciones principales de la tesis, y finaliza listando las publicaciones que avalan este manuscrito.

Objetivos

A pesar del número de trabajos dedicados a los problemas de visióon 3D por ordendor, la complejidad de éstos y las limitaciones para obtener algoritmos eficientes y fiables siguen haciendo de estos problemas un tema relevante de investigación. La reciente introducción de cámaras en prácticamente todos los dispositivos ha motivado el desarrollo de algoritmos más eficientes, así como la consideración de configuraciones específicas y algoritmos especializados para estas. Uno de las mayores problemas, además de la detección de emparejamientos erróneos de datos, es la falta de garantías de calidad u optimalidad de las soluciones obtenidas por estos métodos, dado que, en general los problemas son no-convexos, esto es, el problema puede tener más de un mínimo local. Dado que estos problemas raramente aparecen individualmente, la falta de garantías sobre la calidad de la solución puede afectar a otros problemas que utilicen

estos resultados, incluso llegando a afectar a la solución final de la aplicación que los utiliza. Detectar soluciones subóptimas se convierte en una tarea vital para evitar errores en estos sistemas y hacerlos más robustos y fiables. Es importante señalar que, además de certificar la optimalidad, estos algoritmos deben ser eficientes para mantener las restricciones de tiempo real que se les demanda.

Esta tesis se centra en problemas de visión por ordenador que utilizan únicamente correspondencias 2D-2D y 2D-3D. Nuestro objetivo ha sido desarrollar algoritmos certificadores eficientes que estimen y certifiquen soluciones óptimas. Específicamente, hemos considerado dos temas generales: (1) la estimación de la *pose* (posición y orientación) tanto relativa como absoluta de la cámara; y (2) triangulación de los puntos de la escena. A continuación desarrollamos estos dos temas y resumimos las contribuciones de esta tesis. En la figura 1.2 mostramos un resúmen visual de estas propuestas.

Estimación de la *pose* relativa: Este problema consiste en estimar la *pose* relativa entre dos cámaras calibradas a partir de observaciones (2D) en correspondencia. Esta tarea aparece en la mayoría de aplicaciones que utilizan cámaras, por ejemplo, un sistema autónomo o robot que intenta localizarse. El problema es complejo y el número de trabajos previos que lo tratan es numeroso. Por otro lado, es posible includir información previa durante la estimación sobre el movimiento o la distribución de los puntos de la escena, lo que simplifica el problema pero requiere, en algunos casos, de un método específico para que la estimación tenga en cuenta esta información.

En su forma general, el problema es no-convexo y la posible existencia de varios mínimos locales hace que la calidad (optimalidad) de la solución estimada por algoritmos iterativos no pueda ser garanzatizada *a priori*. Además, en la mayoría de las aplicaciones este problema debe resolverse para cada par de imágenes, lo que requiere de algoritmos eficientes. Nuestra contribución en este tema se centra en utilizar nuevas formulaciones tanto para el caso general como para el caso en el que se disponga del vector de gravedad, desarrollando algoritmos certificadores más eficientes que los disponibles en la literatura.

Estimación de la *pose* absoluta: EL problema de la *pose* absoluta trata de estimar la *pose* entre la cámara con respecto a un sistema de referencia, generalmente fijo, a partir de correspondencias entre observaciones 2D y puntos en 3D. Esta tarea aparece, por ejemplo, tras la detección de un cierre de bucle o al calibrar los parámetros extrínsecos de una cámara. Aunque nuestro trabajo se centra en puntos 3D, la literatura actual también considera las correspondencias con líneas y planos, lo que amplía el número de situaciones en las que este problema es relevante.

En su forma general, el problema es no-convexo, incluso cuando se considera como error a minimizar la distancia punto-línea y varios mínimos pueden existir [Schweighofer and Pinz, 2006]. Esta tesis contribuye a este problema con un algoritmo certificador para cámaras centrales y no-centrales que es capaz

de certificar soluciones en microsegundos, incluso para problemas con datos aleatorios.

Triangulación: Una vez las *poses* de las cámaras han sido estimadas, es posible reconstruir la escena a partir de las observaciones que generaron los puntos en 3D (desconocidos). Esta tarea es denominada 'triangulación' y es el objetivo final de varias aplicaciones, por ejemplo, Structure-from-Motion (SfM). Cuando los datos contienen ruido, la reconstrucción no es tan simple dado que el punto 3D que genera la observaciones no se puede recupear con exactitud. En este caso, el método 'óptimo' corrige las observaciones para que este punto exista y sea único. En ciertos casos, es necesario añadir restricciones adicionales, por ejemplo, cuando las observaciones provienen de un plano, para que los puntos 3D finales cumplan con esta restricción de coplanaridad.

Esta tesis contribuye con algoritmos para los casos con dos y N vistas (general) considerando escenas generales y también cuando todos los puntos son coplanares. Las propuestas que hacemos son más eficientes que los trabajos previos, y son capaces de trabajar con un alto número de vistas. Todas nuestras contribuciones se pueden implementar con las herramientas básicas disponibles en cualquier librería de álgebra lineal, por ejemplo, EIGEN.

Publicaciones

Las publicaciones que avalan esta tesis son las siguientes:

1. **Certifiable relative pose estimation**, *Mercedes Garcia-Salguero, Jesus Briales* and *Javier Gonzalez-Jimenez*, Image and Vision Computing, 63. Vol. 109, May 2021. DOI: https://doi.org/10.1016/j.imavis.2021.104142. IF 2020: 2.818, position 66/140. Area: computer science, artificial intelligence. Q2, T2

2. **Fast and Robust Certifiable Estimation of the Relative Pose Between Two Calibrated Cameras**, *Mercedes Garcia-Salguero* and *Javier Gonzalez-Jimenez*, Journal of Mathematical Imaging and Vision,8. Vol. 63. pp. 1036-1056, 2021. https://doi.org/10.1007/s10851-021-01044-0. IF 2021: 1.629, Position: 107/267. Area: mathematics, applied. Q2, T2.

3. **A Sufficient Condition of Optimality for the Relative Pose Problem between Cameras**, *Mercedes Garcia-Salguero* and *Javier Gonzalez-Jimenez*, SIAM Journal on Imaging Sciences, 4. Vol 14, pp.1617-1634, 2021. DOI: 10.1137/21M1397970. IF 2020: 2.867. Position: 28/265. Area: Mathematics, applied. Q1, T1

4. **A Tighter Relaxation for the Relative Pose Problem Between Cameras**, *Mercedes Garcia-Salguero, Jesus Briales* and *Javier Gonzalez-Jimenez*, Journal of Mathematical Imaging and Vision, 5. Vol 64, pp.493-

505, 2022. DOI: https://doi.org/10.1007/s10851-022-01085-z. IF 2021: 1.629. Position: 107/267. Area: Mathematics, applied. Q2,T2.

5. **Certifiable algorithms for the two-view planar triangulation problem**, *Mercedes Garcia-Salguero* and *Javier Gonzalez-Jimenez*, Computer Vision and Image Understanding, Vol. 225, 2022.
DOI: 10.1016/j.cviu.2022.103570. IF 2022: 4.5. Position: 62/145. Area: computer science, artificial intelligence. Q2, T2

6. **Fast certifiable relative pose estimation with gravity prior**, *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, Artificial Intelligence, Vol. 317, 2023. DOI: https://doi.org/10.1016/j.artint.2023.103862. IF 2022: 14.4. Position: 5/145. Area: Computer science, artifical intelligence. Q1, T1.

7. **Certifiable solver for real-time N-view triangulation** , *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, IEEE Robotics and Automation Letters, 4. Vol 8, pp. 1999-2005, 2023.
DOI: 10.1109/LRA.2023.3245408. IF 2022: 5.2. Position: 10/30. Area: Robotics. Q2

8. **Certifiable Planar Relative Pose Estimation with Gravity Prior**, *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, Computer Vision and Image Understanding, 2023. Vol. 239, pp. 103887.
DOI: https://doi.org/10.1016/j.cviu.2023.103887 IF 2022: 4.5. Position: 62/145. Area: computer science, artificial intelligence. Q2, T2

9. **A fast certifiable algorithm for the N-view planar triangulation**, *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, (submitted)

10. **Fast certifiable algorithm for the absolute pose estimation of a camera** , *Mercedes Garcia-Salguero and Elijs Dima and André Mateus and Javier Gonzalez-Jimenez*, (submitted)

Los codigos asociados a estos trabajos pueden encontrarse en https://github.com/mergarsal.

Marco de la tesis

Esta tesis se ha desarrollado durante los años 2019 y 2023 en el grupo de investigación *Machine Perception and Intelligent Robotics* (MAPIR), parte del Departamento de Ingeniería de Sistemas y Automatización de la Universidad de Málaga (UMA). La principal fuente de financiación durante este periodo

fue la beca para la Formación de Profesorado Universitario (FPU) otorgada por el Ministerio Español de Ciencia, Innovación y Universidades, con código FPU18/01526. Además, los trabajos que avalan esta tesis formaron parte de los proyectos WISER, $DPI2017-84827-R$; HOUNDBOT, $P20\_01302$; y ARPEGGIO, PID2020 − $117057GB-I00$. Dado que la mayor parte de esta tesis se desarrolló durante la pandemia COVID, las limitaciones en general y de viajes internaciones en particular nos llevaron a enfocar y dirigir todos nuestros trabajos a editoriales internacionales y no a conferencias. Sin embargo, en el último tramo de la tesis pude acudir a *International Computer Vision Summer School* (ICVSS) en Sicilia, Italia y realizar una estancia internacional como interino de investigación en Ericsson AB, en Estocolmo, Suecia.

Adicionalemnte, en estos últimos cuatro años he contribuido como revisor en revistas internacionales e impartiendo clases en la Universidad de Málaga. Para finalizar y aunque no se encuentra dentro de las líneas generales de esta tesis, al inicio de este periodo publicamos el artículo

Human 3D Pose Estimation with a Tilting Camera for Social Mobile Robot Interaction, *Mercedes Garcia-Salguero*, *Javier Gonzalez-Jimenez* and *Francisco Ángel Moreno*. MDPI Sensors, vol. 19, no. 22: 2019. Position: 15/63 Area: Instruments & Instrumentation Q1, T1.

Organización de la tesis

El resto de esta tesis está dividida en los siguientes capítulos:

- **Capítulo** 2 incluye los conceptos fundamentales de visión por ordenador y optimización convexa, así como los métodos herramientas disponibles para resolver las relajaciones convexas. Este capítulo no es exhaustivo e incluye solo los conceptos necesarios para que esta memoria sea autocontenida.

- **Capítulo** 3 incluye las contribuciones para el problema de la *pose* relativa entre dos cámaras calibradas.

- **Capítulo** 4 extiende el capítulo anterior y se enfoca en los casos en los que el vector de gravedad es conocido. Esta información introduce restricciones adicionales en las formulaciones, lo que requiere nuevos métodos.

- **Capítulo** 5 finaliza la parte I sobre la *pose* de la cámara. En este capítulo nos centramos en el problema de la *pose* absoluta entre la cámara y un sistema de referencia fija, tanto para cámaras centrales como no-centrales.

- **Capítulo** 6 conforma la segunda parte de este manuscrito y contiene las contribuciones para el problema de triangulación. Estos trabajos consideran tanto 2 como N (general) vistas/imágenes, así como escenas con puntos 3D en distribución general y coplanar.

- **Capítulo** 7 finaliza este manuscrito con las conclusiones y posibles líneas futuras de investigación.

Conclusiones y líneas futuras de investigación

Esta tesis se enfoca en dos de los problemas más comunes en aplicaciones de visión 3D: la estimación de la *pose* de la cámara y la reconstrucción 3D de la escena, considerando además distintas condiciones en la configuración de estos problemas. A pesar de tener una amplia literatura, los trabajos previos formulan estas tareas como problemas no-convexos, esto es, el problema puede tener en general más de un mínimo local. Dado que la herramienta más común para resolver estas formulaciones son algoritmos iterativos, no es posible garantizar a priori que la solución devuelta sea el mínimo global. Empirícamente, estos algoritmos funcionan bien y tienden a obtener buenas soluciones, especialmente cuando la inicialización es 'buena'. Sin embargo, si la solución no es el mínimo global y esto pasa desapercibido esto puede afectar a algoritmos críticos que dependen de esta solución (por ejemplo, la localización de un coche autónomo). Los llamados 'algoritmos globales' son capaces de proveer una solución óptima globalmente, aunque dependendiendo de la formulacioń del problema pueden requerir herramientas para su reolución demasiado lentas para ser usadas en aplicaciones de tiempo real. Como alternativa, los llamados 'algoritmos certificadores', aunque solo son capaces de certificar soluciones, no de calcularlas, ofrecen la ventaja de ser mucho más eficientes, e idóneos para aplicaciones en timpo real. Este es el enfoque que se ha buscado en esta tesis.

Concretamente, nos hemos centrado en problemas que pueden reformularse como minimizaciones de un coste cuadrático en las variables cuyas restricciones se pueden formular también mediante expresiones cuadráticas. Esto nos permite utilizar dos relajaciones convexas muy conocidas: el problema dual y la relajación de Shor. Estos problemas convexos pueden resolverse con las herramientas disponibles con complejidad polinómica en el número de variables y restricciones, aunque para ciertos problemas existen métodos más eficientes, como la descomposición 'low-rank' propuesta en [Burer and Monteiro, 2005]. Nuestro interés en estas relajaciones radica en la posibilidad de derivar algoritmos certificadores que son más eficientes aunque presentan ciertas limitationes para formulaciones con más restricciones que variables y cuando la condición conocida como 'Linear Independence Constraint Qualification' (LICQ) falla. La mayoría de los problemas tratados en esta tesis no cumplen con estas dos

condiciones, motivando el desarrollo de algoritmos alternativos para la certificación.

Las contribuciones de estas tesis se agrupan en los dos problemas anteriormente comentados: el de la *pose* de la cámara y la reconstrucción 3D de la escena. A continuación listamos estas contribuciones así como sus limitaciones.

Estimación de la *pose*: La primera parte de la tesis se centra en problemas que estiman la *pose*, tanto de manera relativa entre dos cámaras y absoluta entre una cámara y un sistema de referencia. Aunque estos suelen ser los primeros problemas que necesitan ser resueltos, incluso las formulaciones iniciales y más simples pueden presentar varios mínimos locales. Nuestra primera contribución en [Garcia-Salguero et al., 2021, Garcia-Salguero and Gonzalez-Jimenez, 2021a] propone certificadores de optimalidad que se pueden estimar en forma cerrada. A pesar de ser eficientes, estos certificadores relajan la formulación inicial y después derivan el problema dual, que es otra relajación. Esto hace que un fallo a la hora de certificar la solución pueda deberse a (a) la solución es suboptimal; y/o (b) cualquiera de estas relajaciones no es/son 'tight'. Empíricamente observamos que esta propuesta es capaz de certificar la mayoría de las estimaciones, especialmente para problemas sin mucho ruido. Estos resultados motivan nuestro siguiente trabajo [Garcia-Salguero and Gonzalez-Jimenez, 2021b] en el que proponemos una condición de suficiencia que garantiza la optimalidad de la solución sin estimar el certificador explícitamente. Esta propuesta es por tanto más eficiente que las anteriores y empíricamente es capaz de detectar un gran número de soluciones óptimas, aunque dado que es una relajación del certificador original, el ratio de detección es menor que para los algoritmos originales. La siguiente propuesta en [Garcia-Salguero et al., 2022] se centra en derivar relajaciones más 'tight' para este mismo problema, aunque estas formulaciones y sus relajaciones asociadas requiren ser resueltas con las herramientas tradicionales que son más lentas pero tienen la ventaja de estimar *y* certificar la solución con el mismo algoritmo. Empíricamente observamos que las formulaciones con más restricciones se mantienen 'tight' incluso para problemas con alto ruido y pocos datos.

Los resultados y observaciones de estos trabajos nos permiten proponer el siguiente algoritmo en [Garcia-Salguero and Gonzalez-Jimenez, 2023b] que se centra en el problema de la *pose* entre dos cámaras pero suponiendo que el vector de gravedad es conocido, por ejemplo, a partir de los datos de una IMU. Esto reduce el número de grados de libertad de la rotacioń de tres a uno, modificando por tanto el problema a resolver. Las restricciones usadas en [Garcia-Salguero et al., 2022] para la configuración general pueden trasladarse a este caso, permitiendo obtener relajaciones que se mantienen 'tight' bajo las mismas condiciones de ruido y datos. Sin embargo, el objetivo de este trabajo es derivar un certificador que funcione con estas formulaciones que tienen más restricciones que variables y donde LICQ falla. Este certificador iterativo es capaz de certificar la mayoría de las soluciones y es más

rápido que resolver la relajación desde cero con las herramientas disponibles. La limitación de esta propuesta radica en la herramienta utilizada para resolver el certificador, que es específica y puede no ser adecuada para sistemas con memoria limitada. No obstante, este método puede extenderse a otros problemas y formulaciones, manteniendo su capacidad de certificación. Nuestro trabajo en [Garcia-Salguero and Gonzalez-Jimenez, a] usa este certificador para el problema de la *pose* con vector de gravedad conocido con la restricción adicional de que los puntos 3D (desconocidos) que generan las observaciones pertenecen a un plano, también desconocido. Este es un caso degenerado para el problema de la *pose* basado en la matriz esencial, siendo necesario utilizar la matriz de homografía. Este trabajo propone cuatro formulaciones distintas con varios números de variables y restricciones que, como en el caso general, tienen distintos niveles de 'tightness'. Desafortunadamente, las formulaciones no cumplen la condición LICQ y dos de ellas tienen más restricciones que variables, haciendo necesario el uso del certificador iterativo presentando anteriormente. Aunque el algoritmo es capaz de certificar soluciones, quizá la observación más relevante es la diferencia en el comportamiento del algoritmo al cambiar la forma de la matríz de coste. Sin embargo, la mayor limitación de la propuesta sigue siendo la herramienta necesaria para resolver el certificador.

Nuestro último trabajo en el tema de la *pose* se centra en estimar la *pose* absoluta de la cámara con referencia a un sistema de referencia fijo, problema conocido como PnP o 'resectioning'. La propuesta en [Garcia-Salguero et al.,] considera tanto cámaras centrales como nocentrales y minimiza el error punto-línea entre las observaciones (líneas) y los correspondientes puntos en 3D. El problema es re-formulado para depender solo de la rotación en 3D y como en los trabajos anteriores, trabajamos con varios sets de restricciones para obtener formulaciones mś 'tight', obteniendo al final formulaciones con más restricciones que variables y donde LICQ falla. El certificador en [Garcia-Salguero and Gonzalez-Jimenez, 2023b] muestra Hessianass con rango deficiente, aunque este algoritmo es demasiado lento debido a la mala inicialización. Este último comportamiento motiva el desarrollo de un nuevo certificador que, en este caso, reformula el certificador como una optimización sin restricciones que maximiza los menores autovalores de la Hessiana. Esta optimización es en general convexa pero 'nonsmooth', y puede implementarse con las herramientas básicas disponibles en cualquier librería de álgebra lineal. Empíricamente, la propuesta es capaz de estimar y certificar soluciones incluso para problemas con coste aleatorio, aunque a priori puede ser complicado extender este algoritmo para trabajar com más de una rotación.

Triangulación: La segunda parte de esta tesis abarca las propuestas en el tema de triangulación, específicamente los métodos denominados 'óptimos' en el que las observaciones se corrigen para que la triangulación del punto en 3D sea exacta. Nuestro primer trabajo en [Garcia-Salguero and Gonzalez-Jimenez, 2022] se centra en el caso con dos vistas/imágenes suponiendo que los

puntos que originan las observaciones pertenecen a un plano. En este caso, la matríz de homografía da la relación que tienen que cumplir las observaciones corregidas. Nuestra primera propuesta en este trabajo es un 'primal-dual' algoritmo que es capaz de estimar la solución y además garantiza la optimalidad de esta solución cuando el algoritmo converge. Basándonos en esta formulación del problema, nuestra segunda contribución es un certificador que no estima la solución, pero solo la certifica. Nuestra tercera contribucion extiende este concepto proponiendo una condición de suficiencia, que es más rápida que el certificador, pero falla al detectar soluciones óptimas en ciertos casos. Sin embargo, todas estas propuestas se centran en el caso con dos vistas, y no todos los métodos pueden ser extendidos al caso no-mínimo con N vistas de una manera directa. Nuestro segundo trabajo en [Garcia-Salguero and Gonzalez-Jimenez, 2023a] se enfoca en esta configuración no-mínima y proponemos un algoritmo para el caso general, donde los puntos 3D no tienen que pertenecer a un plano. Para ello, nos basamos en la formulación en [Aholt et al., 2012] en el que la matríz esencial/fundamental es usada para relacionar las observaciones en lugar de las matrices de proyección. Esta formulación hace que el problema tenga restricciones cuadráticas, con el inconveniente de requerir $\binom{N}{2}$ restricciones para N vistas, haciendo que el problema alcance las limitaciones actuales de las herramientas necesarias para resolver la relajación. Nuestra propuesta evita estas herramientas y contribuye con un algoritmo certificador, en el que la solución es estimada mediante el método de Newton y la optimalidad se garantiza con un certificador. Aunque para más de cuatro vistas la formulación tiene más restricciones que variables y falla LICQ, empíricamente observamos que el certificador puede estimarse en fórmula cerrada, con la solución con norma mínima. A pesar de que nuestros experimentos con problemas sintéticos y reales muestra que este es el caso, por ahora es solo una observación y no tenemos una prueba matemática que respalde este comportamiento. La última contribución de esta tesis expande estos dos trabajos previos y se centra en el caso con N vistas suponiendo la restricción coplanar. En este caso, es posible usar una formulación con el mínimo número de restricciones, aunque empíricamente observamos que no es 'tight' para algunos problemas. Siguiendo la propuesta anterior, también proponemos usar la formulación con todas las restricciones disponibles $2\binom{N}{2}$ que observamos se mantiene 'tight' para todos los problemas considerados. Sin embargo, el alto número de constraints require más tiempo para ser resuelto. En ambos casos seguimos el algoritmo certificador de nuestro trabajo anterior, donde el método de Newton estima la solución y el certificador intenta certificar esta solución como el óptimo global. Como en el caso general, la formulación con más restricciones falla LICQ, pero la solución con norma mínima es capaz de certificar positivamente las soluciones. Sin embargo, este comportamiento sigue siendo una observación, y la prueba matemática es una linea abierta de investigación, a la vez que motiva la búsqueda de problemas que muestren las mismas características.

Líneas futuras de investigación:

Concluimos esta tesis con dos posibles líneas de investigación que consideramos especialmente relevantes. Ambos temas son complementarios y pueden usarse en conjunto como alternativa a las herramientas disponibles actualmente para resolver relajaciones convexas con la forma tratada en esta tesis. A pesar de ser excepcionalemtne útiles, estas herramientas son poco eficientes mientras que los sistemas actuales requiren algoritmos más rápidos, aunque estos tengan que ser específicos para el problema que se quiere resolver.

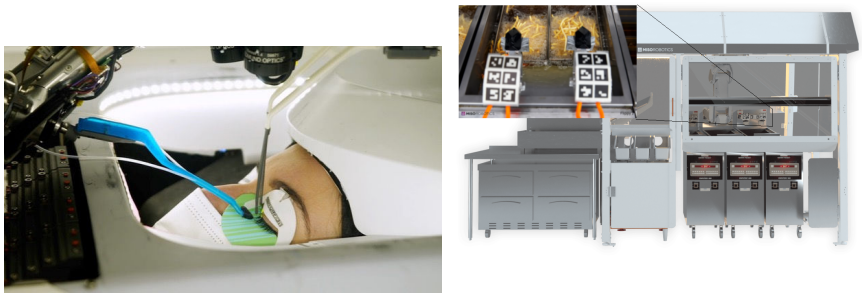
Certificadores eficientes: Como hemos vistos durante esta tesis, el número de problemas que requieren más restricciones que variables y donde la condición LICQ falla es elevado. Mientras las herramientas disponibles se pueden usar para resolver las relajaciones asociadas, la falta de certificadores es una limitación a la hora de implementar estos métodos en sistemas en tiempo real. Una futura línea de investigación podría centrarse en desarrollar algoritmos de esta clase, aunque sean específicos para ciertos problemas y que sean capaces de certificar soluciones incluso cuando LICQ falla y la formulación usada tenga más restricciones que variables. Adicionalemnte, limitar el número de librerias necesarias para resolver estos certificadores y hacerlos más eficientes puede expandir el número de aplicaciones que pueden usarlos. Una línea secundaria de investigación debería centrarse en aproximaciones e inicializaciones de estos certificadores, que sean más sencillas de calcular y obtenga soluciones próximas a la solución final. Derivar estas aproximaciones puede mejorar la convergencia de certificadores iterativos y explicar la razón por la que iertos algortimos funcionan mejor que otros.

Estimar soluciones a partir de certificadores negativos: Mientras la primera propuesta intenta certificar soluciones, nuestra segunda propuesta trabaja con los certificadores negativos (no-concluyentes) de estos algoritmos. Como comentábamos anteriormente, estos certificadores pueden fallar generalmente por dos razones: (1) la solución es una solución subóptima; y (2) las relajaciones no son 'tight'. Cuando el fallo del certificador se deba a la solución, esta línea se centra en obtener soluciones aproximadas al problema original a partir de los fallos del certificador. Este método puede usarse como inicialización para algoritmos iterativos que estimen la solución al problema, potencialmente siendo una alternativa mejor a las inicializaciones actuales. Usando estos algoritmos junto con los certificadores anteriores permitiría obtener primal-dual algoritmos más eficientes que las herramientas actuales y aplicables a todas las fomulaciones, incluso aquellas con fallo de LICQ y con restricciones redundantes.

1 Introduction

Being able to perceive the surroundings is a fundamental ability for any autonomous system, and whereas different sensors can provide this information with different ranges and operational requirements, cameras are a constant in most systems. Their low economic cost, variety and availability together with the amount of information we can extract and infer from images have boosted their introduction in most autonomous systems. Cameras offer a visual of the scene and a fast, reliable source of information for the system *and* the operator, and thus, computer vision techniques has seen a development in many areas, beyond academic research and high-tech industries. The specific requirements have motivated different lines of research and development, but also the improvement of the already-tackled tasks focusing on precision, efficiency and/or computational requirements (memory, tools, software libraries, ...). For example, some applications require not only high precision for small scales but also a reliable and bullet-proof designs, for example, the robot shown in Figure 1.1a which works lashes, a remarkably small part of the human body. Other products are devised to work in highly non-controlled environments with non-typical and arguably unadvised substances, as as the hot, frying oil in Figure 1.1b.

These complex applications often involve the interconnection of several of these tasks, as we shown in Figure 1.2, assuring that these blocks perform well and return good solutions is key for a safe operation. Two of the most common problems that appear in most visual-based pipelines including camera tracking (aka visual odometry (VO)), Structure-from-Motion (SfM) and Simultaneous Localization and Mapping (SLAM) aim to estimate the camera(s) motion and the 3D structure of the scene. Although different approaches have been proposed to solve these problems, in this thesis we focus on geometric approaches that work with low-level information in the form of feature points, for example, corners or blobs in the image. This information then feeds the algorithm that



(a) Luum lashes (b) Flippy

Figure 1.1: Collaborative robots in the beauty and fast food industry relying on computer vision techniques for their tasks: 1.1a automatically glues lashes (courtesy https://www.luumlash.com); and 1.1b fries different goods (courtesy https://misorobotics.com/flippy/).

estimates the relevant solution, which is said to be a minimum/maximum of a cost function. Therefore, different cost objectives and in general different formulations for the same problem may have a different solution and special care must be put into the correct formulation of the problem. In addition, in most cases, including those within the two above-mentioned topics, the problems we solve are nonconvex, hence having in general more than one local minima. The difference between these minima may be large and so may be their effect on other tasks that leverage these solutions and even the final solution of the whole pipeline. Thus, being able to guarantee that the returned solution is the global optimum, *i.e.*, the *best* estimation for the given data, is crucial for the correct performance of the blocks and the application in general.

1.1 Motivation

As introduced before, the problems that formed complex pipelines as the one in Figure 1.2 are often nonconvex, whereas their solutions are usually used as inputs for other, subsequent problems during the execution of the pipeline. However, the quality of the solutions for these tasks is rarely measured and the final estimation is often used to determine whether or not the behavior was correct, for example, the final point cloud. A "bad" (suboptimal) solution from one of the inter-blocks could potentially affect other blocks, a situation which is of special importance if the other blocks are sensitive to the input or the solution is used to determine inliers and outliers.

Whereas for some specific instances closed-form methods have been proposed for these nonconvex problems, the majority of these approaches are minimal solvers that rely on the minimal number of data, hence having limited accuracy for nonminimal problem instances. On the other hand, iterative approaches are preferred due to their accuracy as they consider all the available data and the empirical observation that good initializations. For example those returned by the minimal solvers, tend to return good solutions, although the convergence to the global optimum, *i.e.*, the *best* solution, for the provided data, is not guaranteed. However, assessing the quality of the final solution or even the initial guess is a difficult task and any optima, including the suboptimal ones, can be returned by the iterative algorithms.

Alternative approaches rely on "global methods" that solve the problem globally, thus coming with an optimality certification for the returned solution, for example, Branch-and-Bound algorithms. Another example is found in the above-mentioned closed-form algorithms that state the problem as a polynomial system whose roots are the solutions (potentially many) of the problem. These algorithms estimate and then check all the solutions, keeping the one with minimum cost. However, even estimating all the solutions may be not suitable for some problems and the tools required to obtain them may also become unstable. Similarly, the computational complexity of Branch-and-Bound algorithms are affected by the search nature of the approach.

In this thesis, we focus instead on convex relaxations of the original problems and specifically on relaxations for problems with quadratic cost and constraints. Although, at first, the restriction about the quadratic form of the formulation may appear too strict for some problems, we show that many problems in computer vision can be stated in this form. In this line, recent developments such as the Lassare's hierarchy [Lasserre, 2001] brings to alleviate this requirement and increase the range of problems that can benefit from these relaxations. The convex relaxations used in this thesis can be solved by off-the-shelf tools that have polynomial complexity in the number of variables and constraints. These tools admit several hundreds of variables and constraints, although they require milliseconds per problem instance even for less than 20 variables and 30 constraints. An additional drawback of these methods is the lack of tight relaxations, that is, the relaxation does not approximate well the original, nonconvex problem. Whereas certain actions can be taken when this happens, for example, adding redundant but linearly independent constraints, some extra work must be put into the formulation to alleviate this. The main drawback, however, is the computational requirement of these tools, and whereas recent works have tackled this problem, for example, the low-rank decomposition [Burer and Monteiro, 2005], even these tools come with some limitations and/or assumptions about the formulation that may hinder the application of the approach. Fortunately, more efficient approaches can be derived from these same convex relaxations, although they still require some conditions on the formulation of the problem to be used in practice and remain

efficient. Among them, we highlight the so-called *certifiable algorithms* [Bandeira, 2016] that only certify but do not estimate the solution. The motivation behind these works is that iterative solvers are efficient, can be derived for a wide variety of problems and work well in most instances provided the initial guess is good "enough". Hence, only certifying this solution alleviates computational costs from the pipeline while assuring the quality of said solution. These works focus rather on the certification stage and how to derive an efficient algorithm for it. This thesis tackles different certification techniques for common 3D computer vision problems, which appear in most visual-based pipelines, including SLAM and SfM.

1.2 Contributions

Section 1.2.1 previews the problems tackled in this thesis and the proposed contributions. Further details can be found in their respective chapters. Section 1.2.2 lists the published and submitted papers that support this thesis.

1.2.1 Goals

Despite the maturity of the field, 3D computer vision problems are still a hot research topic due to their inherent complexity and the difficulty of obtaining efficient yet reliable algorithms. Further, the ubiquitous use of cameras has motivated the development of more efficient approaches for these problems whereas the introduction of new system configurations has inspired new specific algorithms. One of the main complexities arises from the lack of quality guarantees of the returned solutions, as these problems are in general nonconvex and thus more than one local minimum may exist. Since these problems rarely appear on their own but rather as a sub-task within a more complex pipeline, for example, SLAM or SfM, this lack of guarantees on the quality of the solution may cascade into other blocks, hindering the result of the applications and even making them fail. Detecting these poor-quality solutions at an early stage reduces the probability of this outcome and enhance the reliability of the full application. In addition to certify the quality of the solution, these algorithms are often required to be efficient to maintain the "real-time" requirement of the applications.

In this thesis, we tackle 3D computer vision problems that rely on pairwise 2D-2D and 2D-3D point correspondences. We seek efficient, certifiable solvers that obtain and certify the quality of the solution, usually by either saying that the point is the global optimum or being inconclusive about its optimality. We consider two main problems: (1) pose estimation, either (1.a) relative or (1.b) absolute (with respect to a predefined world frame); and (2)

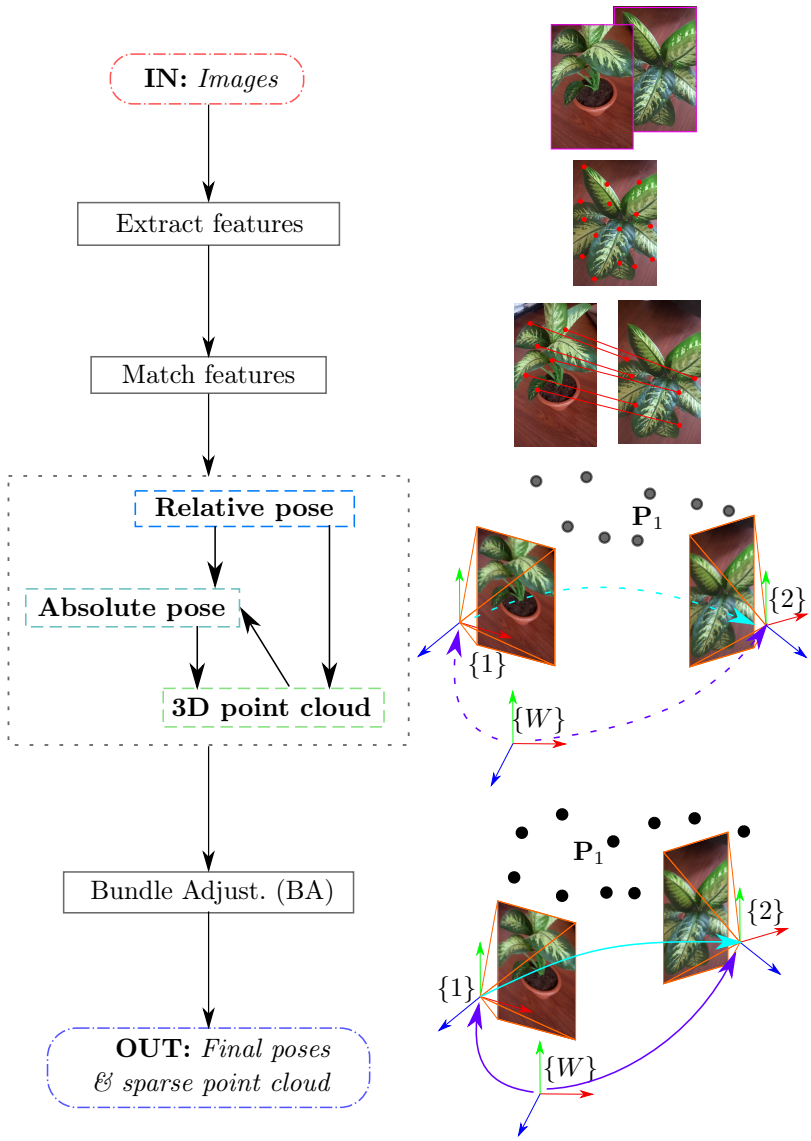


Figure 1.2: Simplified pipeline for image-based localization and mapping applications. This thesis focuses on the three tasks in colored rectangles: relative and absolute pose; and 3D point cloud extraction.

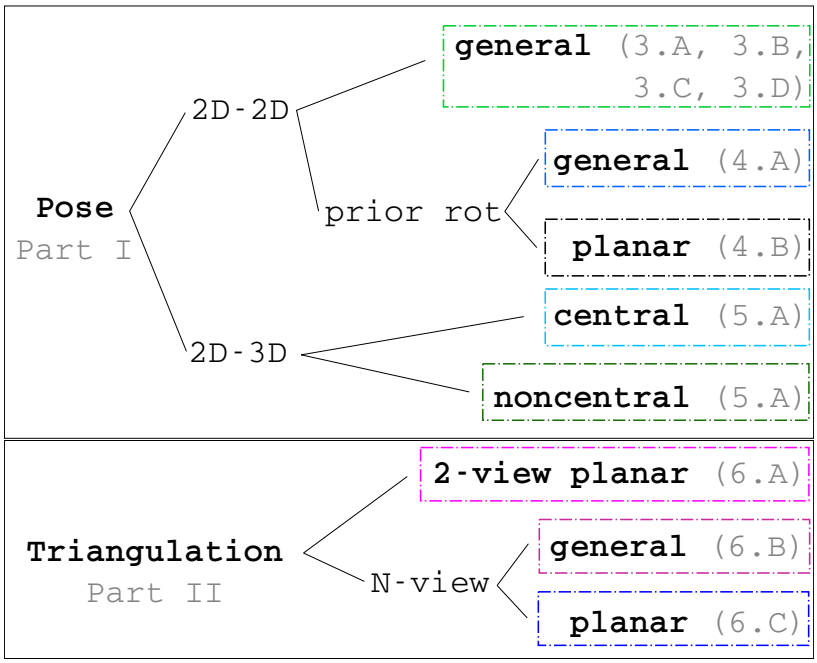


Figure 1.3: Contributions of this thesis are enclosed in rectangles and their associated general computer vision topics are shown in bold letters.

3D point estimation or triangulation. The rest of this section elaborates on these topics and the proposed approaches presented in this thesis. We provide a visual summary of the contributions in Figure 1.3.

Relative pose estimation: This task seeks the 3D pose (position and orientation, i.e. 6 degree-of-freedom) between two (calibrated) cameras given a set of corresponding 2D-2D observations. This problem is the cornerstone of many applications involving cameras, for instance, an autonomous system trying to localize itself in a scene. This problem turns out to be remarkably complex, and the associated literature is extensive. Further, the problem can be simplified via prior information or assumptions about the camera movement and/or the distribution of the 3D points, then requiring a different approach that accounts for this information.

In general, the problem is nonconvex, thus the quality of the solution cannot be guaranteed a priori when relying on iterative algorithms. Further, most applications solve this problem for each pair of images, thus requiring solvers that are both efficient and reliable. Our contributions, for both the general configuration and the one with a known axis of rotation, focus on applying new formulations for these problems and on providing efficient algorithms for

optimality certification. We show that the proposals are faster than state-of-the-art algorithms while removing assumptions about the formulations that preclude standard certification approaches

Absolute pose estimation: This problem seeks the pose of a camera *w.r.t.* a fixed frame, given correspondences between 2D observations and 3D points in the space. This task appears, for example, after a loop-closure has been detected or during the extrinsic calibration of a camera. The extensive literature for this problem also considers 3D lines and 3D planes, which hugely extends the applicability of the problem.

In general, the problem is nonconvex and under a point-line distance formulation it presents more than one local minimum [Schweighofer and Pinz, 2006]. We contribute an efficient certifiable solver for central and noncentral cameras that runs in microseconds in a standard setup and works even for random data, being faster than other approaches in the literature. The proposal poses the certification problem as an eigenvalue optimization that can be solved with a (sub) gradient descending scheme.

Triangulation problem: Once the camera poses are known, it is possible to retrieve the 3D points that originate the observations from two or more images. This is the so-called triangulation problem, and it is the purpose of reconstruction applications, for instance, 3D mapping. For noisy data, the reconstruction process is not that simple as the 3D unknown point may not be consistent for the given data. In this case, the so-called "optimal" approach seeks to correct the data so that the point can be retrieved trivially. Additional restrictions, as the unknown point belonging to a plane, must be taken into consideration while correcting the observations, as the final solution must be consistent with this prior knowledge.

We tackle the configuration with two views and the general case with N views, and consider 3D points in general and coplanar positions. We provide efficient and low-cost certifiable approaches that scale well with the number of views/cameras, even when this number reaches the hundreds, surpassing the memory limits of other works. Our contributions are also faster than state-of-the-art methods and are implemented with the library for linear algebra Eigen, hence not requiring any additional specific tool.

1.2.2 List of publications

This thesis is presented under the modality of "thesis supported by a compendium of relevant published papers". Following is the list of those publications.

1. **Certifiable relative pose estimation**, *Mercedes Garcia-Salguero, Jesus Briales* and *Javier Gonzalez-Jimenez*, Image and Vision Computing, 63. Vol. 109, May 2021. DOI: https://doi.org/10.1016/j.imavis.2021.104142.

IF 2020: 2.818, position 66/140. Area: computer science, artificial intelligence. Q2, T2

2. **Fast and Robust Certifiable Estimation of the Relative Pose Between Two Calibrated Cameras**, *Mercedes Garcia-Salguero* and *Javier Gonzalez-Jimenez*, Journal of Mathematical Imaging and Vision,8. Vol. 63. pp. 1036-1056, 2021. https://doi.org/10.1007/s10851-021-01044-0. IF 2021: 1.629, Position: 107/267. Area: mathematics, applied. Q2, T2.

3. **A Sufficient Condition of Optimality for the Relative Pose Problem between Cameras**, *Mercedes Garcia-Salguero* and *Javier Gonzalez-Jimenez*, SIAM Journal on Imaging Sciences, 4. Vol 14, pp.1617-1634, 2021. DOI: 10.1137/21M1397970. IF 2020: 2.867. Position: 28/265. Area: Mathematics, applied. Q1, T1

4. **A Tighter Relaxation for the Relative Pose Problem Between Cameras**, *Mercedes Garcia-Salguero, Jesus Briales* and *Javier Gonzalez-Jimenez*, Journal of Mathematical Imaging and Vision, 5. Vol 64, pp.493-505, 2022. DOI: https://doi.org/10.1007/s10851-022-01085-z. IF 2021: 1.629. Position: 107/267. Area: Mathematics, applied. Q2,T2.

5. **Certifiable algorithms for the two-view planar triangulation problem**, *Mercedes Garcia-Salguero* and *Javier Gonzalez-Jimenez*, Computer Vision and Image Understanding, Vol. 225, 2022.
DOI: 10.1016/j.cviu.2022.103570. IF 2022: 4.5. Position: 62/145. Area: computer science, artificial intelligence. Q2, T2

6. **Fast certifiable relative pose estimation with gravity prior**, *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, Artificial Intelligence, Vol. 317, 2023. DOI: https://doi.org/10.1016/j.artint.2023.103862. IF 2022: 14.4. Position: 5/145. Area: Computer science, artifical intelligence. Q1, T1.

7. **Certifiable solver for real-time N-view triangulation** , *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, IEEE Robotics and Automation Letters, 4. Vol 8, pp. 1999-2005, 2023.
DOI: 10.1109/LRA.2023.3245408. IF 2022: 5.2. Position: 10/30. Area: Robotics. Q2

8. **Certifiable Planar Relative Pose Estimation with Gravity Prior**, *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, Computer Vision and Image Understanding, 2024. Vol. 239, pp. 103887.
DOI: https://doi.org/10.1016/j.cviu.2023.103887. IF 2022: 4.5. Position: 62/145. Area: computer science, artificial intelligence. Q2, T2

9. **A fast certifiable algorithm for the N-view planar triangulation**, *Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez*, (submitted)

10. **Fast certifiable algorithm for the absolute pose estimation of a camera** ,
Mercedes Garcia-Salguero and Elijs Dima and André Mateus and Javier Gonzalez-Jimenez, (submitted)

Codes are available at https://github.com/mergarsal.

1.3 Thesis framework

This thesis has been developed during four years (2019-2023) in the Machine Perception and Intelligent Robotics (MAPIR) research group, which is part of the Department of System Engineering and Automation at the University of Málaga in Spain (UMA). The author received the grant *Formación de Profesorado Universitario* (FPU) (FPU1801526) by the Spanish Ministy of Science, Innovation and Universities, which was the main funding source of this research. However, the works presented in this thesis were framed in the projects WISER, DPI2017 − 84827-R; HOUNDBOT, *P*20_01302; y ARPEGGIO, PID2020 − 117057*GB* − *I*00. The vast majority of this thesis took place during COVID pandemic and the restrictions in general and particularly those affecting international and national travels led us to direct all our manuscripts to international journals, hence having no presence in conferences. Nonetheless, by the end of this period the author had the opportunity to (safely) attend the International Computer Vision Summer School (ICVSS) in Sicily, Italy and later, spend 4 months as a research intern at Ericcson AB, Stockholm, Sweden.

On the other hand, during the last four years the author reviewed manuscripts for different journals and taught courses at the University of Málaga. Last, and although outside of the scope of this thesis, we published the article

Human 3D Pose Estimation with a Tilting Camera for Social Mobile Robot Interaction, *Mercedes Garcia-Salguero*, *Javier Gonzalez-Jimenez* and *Francisco Ángel Moreno*. MDPI Sensors, vol. 19, no. 22: 2019. Position: 15/63 Area: Instruments & Instrumentation Q1, T1.

1.4 Outline

The rest of this manuscript is divided into the following parts:

- **Chapter 2** reviews the fundamental concepts from computer vision and convex optimization. The chapter is not comprehensive and we include

only those concepts required to follow this manuscript. Standard approaches and tools for convex optimization are also explained to provide an overview of the state-of-the-art.

- **Chapter 3** starts part I and introduces the problem of estimating the relative pose between two calibrated cameras for the general case.
- **Chapter 4** extends chapter 3 for those cases in which the gravity vector is known, for example, from the information provided by an IMU. This prior information introduces additional constraints on the relative pose problem.
- **Chapter 5** finishes part I and tackles the resectioning or PnP problem that estimates the absolute pose of the camera.
- **Chapter 6** forms part II and contains the contributions regarding the triangulation problem for both planar and generic point clouds.
- **Chapter 7** concludes this thesis, providing an overview of the contributions and future research lines.

2 Foundation

This chapter collects the main concepts about computer vision and convex optimization employed in this thesis. We refer interested readers to the books [Hartley and Zisserman, 2003, Ma et al., 2004] and [Boyd and Vandenberghe, 2004, Nocedal and Wright, 1999] for more detailed information. We start by reviewing some core mathematical concepts that are used throughout the thesis.

2.1 Mathematical concepts and terminology

2.1.1 Common notation and operations

Common notation: Through the thesis, bold, upper-case letters $\mathbf{R}, \mathbf{E}$ indicate matrices, while bold, lower-case letters $\mathbf{r}, \mathbf{t}$ indicate vectors. The space of real numbers is denoted by $\mathbb{R}^k$ and we reserve the notation $\boldsymbol{I}_k$ for identity matrices of size $k \times k$, and $\mathbf{0}_{n\times k}$ for matrices of size $n \times k$ whose entries are zeros. Unless otherwise stated, the element on the i-th row and j-th column of the matrix $\mathbf{A}$ is denoted by a_{ij}. In addition, if a matrix is heavily used we number its entries column-wise, for example, for the 3×3 matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$ we have

$$\mathbf{A} = \begin{pmatrix} a_1 & a_4 & a_7 \\ a_2 & a_5 & a_8 \\ a_3 & a_6 & a_9 \end{pmatrix}. \tag{2.1}$$

The transpose of the matrix $\mathbf{A}$ is denoted by $\mathbf{A}^T$. Additionally, and although we mainly focus on real spaces, we will occasionally treat complex matrices as a generalization and we will specifically indicate that the matrices

are complex with the notation $\mathbf{A} \in \mathbb{C}^{n\times n}$. In those cases, the complex conjugate transpose of $\mathbf{A}$ is denoted by $\mathbf{A}^*$. The inner-product for two real vectors $\mathbf{r}, \mathbf{t} \in \mathbb{R}^n$ is $\langle \mathbf{r}, \mathbf{t} \rangle = \mathbf{r}^T\mathbf{t} \in \mathbb{R}$ and the outer product is $\mathbf{r}\mathbf{t}^T \in \mathbb{R}^{n\times n}$. The set of indices from 1 to M is indicated as $[M] = \{1, 2, \dots, M-1, M\}$.

Common operations: The trace of a matrix is the sum of its diagonal entries, that is, $\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$, for instance, for the 3×3 matrix $\mathbf{A}$ in Equation (2.1) is $\text{tr}(\mathbf{A}) = a_1 + a_5 + a_9$. Further, the trace of any real is the real itself, *i.e.*, $a = \text{tr}(a)$ for $a \in \mathbb{R}$ and we will use the cyclic property of the trace as $\text{tr}(abc) = \text{tr}(bca) = \text{tr}(cab)$. The column-wise vectorization of a matrix $\mathbf{A}$ is denoted by $\text{vec}(\mathbf{A}) = \mathbf{a}$ and we can write the expression $\text{tr}(\mathbf{A}^T\mathbf{B}) = \text{vec}(\mathbf{A})^T \text{vec}(\mathbf{B})$. The Kronecker product is denoted by $\otimes$, and is defined for two matrices $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{B} \in \mathbb{R}^{r\times s}$ as

$$\mathbf{A} \otimes \mathbf{B} \doteq \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2} & \cdots & a_{mn}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{mr\times ns}, \tag{2.2}$$

where the element on the i-th row and j-th column of the matrix $\mathbf{A}$ is a_{ij}. The properties $(\mathbf{A} \otimes \mathbf{B})^T = \mathbf{A}^T \otimes \mathbf{B}^T$ and $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$ for appropriate dimensions hold, see more in [Golub and Van Loan, 2013, Sec. 1.3.6]. Different identities relate the vectorization of a matrix product and the kronecker product, among them

$$\text{vec}(\mathbf{AXB}) = (\mathbf{B}^T \otimes \mathbf{A}) \text{vec}(\mathbf{X}) \tag{2.3}$$

$$\text{tr}(\mathbf{ABCD}) = \text{vec}(\mathbf{A}^T)^T (\mathbf{D}^T \otimes \mathbf{B}) \text{vec}(\mathbf{C}) \tag{2.4}$$

The diagonal operator $\text{diag}(a_1, \dots, a_m)$ makes the $m \times m$ diagonal matrix whose diagonal entries are $a_1, \dots, a_m$. The block-diagonal operator works in the same manner but with matrices $\mathbf{A}_1, \dots, \mathbf{A}_m$ of any size as $\text{blkdiag}(\mathbf{A}_1, \dots, \mathbf{A}_m) = \mathbf{A}_1 \oplus \cdots \oplus \mathbf{A}_m$.

The cross-product is denoted by $\times$ and the equivalent skew-symmetric matrix to the expression $\mathbf{t}\times = [\mathbf{t}]_\times$ for the 3D vector $\mathbf{t} \doteq [t_1, t_2, t_3]^T$ is

$$[\mathbf{t}]_\times \doteq \begin{pmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{pmatrix} \tag{2.5}$$

and, since it is skew-symmetric, we know that $[\mathbf{t}]_\times^T = -[\mathbf{t}]_\times$. Further, the expression $[\mathbf{t}]_\times\mathbf{t} = \mathbf{0}_{3\times 1}$ holds by construction, and also, for all 3D vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ we have that $\mathbf{a}^T[\mathbf{b}]_\times\mathbf{a} = 0$.

The weighted sum for a set of functions $f_i : \mathbb{R}^N \to \mathbb{R}$ is defined as $f(x) = a_1 f_1(x) + \cdots + a_M f_M(x)$ for the weights $a_1, \dots, a_M$, whereas the composition of a function $f(x)$ with the affine mapping $Ax + b$ is $f(Ax + b)$.

Norms: For vectors $\mathbf{x} = [x_1, \dots, x_n]^T \in \mathbb{R}^n$, the general ℓ_{p}-norm with $p \geq 1$ is

$$\|\mathbf{x}\|_{\mathrm{p}} \doteq \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}, \tag{2.6}$$

where $|x_i|$ takes the absolute value of x_i. Three well-known norms are the ℓ_1, ℓ_2 and ℓ_∞ which follow the previous definition. Concretely, ℓ_1 is defined as

$$\|\mathbf{x}\|_1 \doteq |x_1| + \cdots + |x_n|, \tag{2.7}$$

the ℓ_2 or *Euclidean norm* as

$$\|\mathbf{x}\|_2 \doteq \left(|x_1|^2 + \cdots + |x_n|^2\right)^{1/2} \tag{2.8}$$

and the ℓ_∞ as

$$\|\mathbf{x}\|_\infty \doteq \max\left\{|x_1|, \dots, |x_n|\right\}. \tag{2.9}$$

For square, $n \times n$ real matrices we can define similar norms. The Frobenius norm is

$$\|\mathbf{A}\|_{\mathrm{F}} \doteq \mathrm{tr}(\mathbf{A}\mathbf{A}^T)^{1/2}. \tag{2.10}$$

Note that this norm is invariant to the application of orthogonal matrices and so $\|\mathbf{A}\|_{\mathrm{F}} = \|\mathbf{O}\mathbf{A}\|_{\mathrm{F}} = \|\mathbf{A}\mathbf{O}\|_{\mathrm{F}}$ for $\mathbf{O} \in \mathbb{O}(n)$ where $\mathbb{O}(n)$ is the space of $x \times n$ orthogonal matrices.

The ℓ_1 is defined as

$$\|\mathbf{A}\|_1 = \max_j \sum_{1=1}^{n} |a_{ij}|, \tag{2.11}$$

and the ℓ_∞ is

$$\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|. \tag{2.12}$$

Last, the ℓ_2 or *spectral norm*

$$\|\mathbf{A}\|_2 = \rho(\mathbf{A}\mathbf{A}^T)^{1/2} \tag{2.13}$$

where $\rho(\mathbf{B})$ is the *spectral radius* of $\mathbf{B}$ defined as

$$\rho(\mathbf{B}) = \max|\lambda_i| \tag{2.14}$$

with $\lambda_1, \dots, \lambda_n$ the eigenvalues of $\mathbf{B}$.

Notice that these norms, both for vectors and matrices, are defined for unconstrained variables in $\mathbb{R}^n$ and $\mathbb{R}^{n\times n}$, respectively. Certain sets of matrices that fulfill some conditions, for example vectors with unit norm, have specific metrics, see [Absil et al., 2009, Boumal, 2023a].

2.1.2 Sets

The non-negative orthant is formed by all real, non-negative elements in $\mathbb{R}^k$ and it's denoted by $\mathbb{R}^k_+$. As an extension, the set $\mathbb{R}^k_{++}$ contains all the strictly positive real numbers. The set of orthonormal matrices is defined as

$$\mathbb{O}(n) \doteq \left\{\mathbf{R} \in \mathbb{R}^{n\times n} \mid \mathbf{R}^T\mathbf{R} = \boldsymbol{I}_n\right\} \tag{2.15}$$

We define the set of 3D of rotations as

$$\mathbb{SO}(3) \doteq \left\{\mathbf{R} \in \mathbb{R}^3 \mid \mathbf{R}^T\mathbf{R} = \boldsymbol{I}_3, \det(\mathbf{R}) = +1\right\} \subset \mathbb{O}(3). \tag{2.16}$$

Different metrics can be defined for rotation matrices, see [Hartley et al., 2013, Table 2].

In a similar manner, the sphere is defined as

$$\mathbb{S}^n \doteq \left\{\mathbf{t} \in \mathbb{R}^n \mid \mathbf{t}^T\mathbf{t} = 1\right\} \tag{2.17}$$

and we mainly use the 2-sphere formed by 3D vectors with unit norm. Notice that this set does not include the points with norm below one, which is typically called *ball*. Last, we say that a matrix $\mathbf{A}$ is positive definite (PD), denoted by $\mathbf{A} \succ 0$, *iff* all its eigenvalues are strictly positive. A positive semidefinite matrix (PSD) has non-negative eigenvalues and it's denoted by $\mathbf{A} \succeq 0$. We elaborate further on these matrices in the next sections, as they are key for this thesis.

2.1.3 Schur's complement

Schur's complement will be useful in the next sections to marginalize variables and to check for (semi)definiteness of matrices, that is, all their eigenvalues are non-negative (positive semidefinite) or strictly positive (positive definite). For the block-partitioned matrix

$$\mathbf{M} \doteq \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \tag{2.18}$$

for $\mathbf{D}$ invertible, the Schur complement of the block $\mathbf{D}$ of the matrix $\mathbf{M}$ is $\mathbf{M}/\mathbf{D} \doteq \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$ and if $\mathbf{A}$ is invertible, we can also write it as $\mathbf{M}/\mathbf{A} \doteq \mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}$. If the matrices are singular, the generalized inverse denoted as $\mathbf{A}^g$ is used instead. Necessary and sufficient conditions for (semi)definiteness of matrices can be derived from the (generalized) Schur complement. In this case, we work with the symmetric matrix

$$\mathbf{X} \doteq \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix}. \tag{2.19}$$

If $\mathbf{A}$ is invertible, the matrix $\mathbf{X}$ is positive definite *iff* $\mathbf{A} \succ 0$ and $\mathbf{X}/\mathbf{A} \doteq \mathbf{C} - \mathbf{B}^T\mathbf{A}^{-1}\mathbf{B} \succ 0$. Alternatively, if $\mathbf{C}$ is invertible, then $\mathbf{X} \succ 0$ *iff* $\mathbf{C} \succ 0$ and

$\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T \succ 0$. For $\mathbf{X}$ to be semidefinite positive we require $\mathbf{A} \succ 0$ and the Schur complement $\mathbf{X}/\mathbf{A} \succeq 0$ and equivalently if $\mathbf{C} \succ 0$ we use $\mathbf{X}/\mathbf{C} \succeq 0$. We can derive alternative conditions for $\mathbf{X} \succeq 0$ using the generalized Schur complement for the generalized inverse $\mathbf{A}^g$ as (1) $\mathbf{A} \succeq 0, \mathbf{C} - \mathbf{B}^T\mathbf{A}^g\mathbf{B} \succeq 0, \big(\boldsymbol{I} - \mathbf{A}\mathbf{A}^g\big)\mathbf{B} = \mathbf{0}$; and (2) $\mathbf{C} \succeq 0, \mathbf{A} - \mathbf{B}\mathbf{C}^g\mathbf{B}^T \succeq 0, (\boldsymbol{I} - \mathbf{C}\mathbf{C}^g)\mathbf{B}^T = \mathbf{0}$.

2.1.4 Eigendecomposition

The eigendecomposition or matrix diagonalization of a squared matrix $\mathbf{A} \in \mathbb{C}^{n\times n}$ aims to estimate a set of linearly independent eigenvectors $\mathbf{x}_i \in \mathbb{C}^n$ and their corresponding eigenvalues $\lambda_i \in \mathbb{C}$, called the *spectrum* of the matrix, for $i = [n]$, and formally $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$ for all i. Note that it is not always possible to find a set of linearly independent eigenvectors and eigenvalues even if the matrix is also real, and so the eigendecomposition does not exist for that particular matrix.

If possible, the information about the eigenvalues and eigenvectors is obtained by solving the polynomial system $\det(\mathbf{A} - \lambda\boldsymbol{I}_n) = 0$, known as the *characteristic equation*. The matrices $\mathbf{U} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathbb{C}^{n\times n}$, which is full-rank, and $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n) \in \mathbb{C}^{n\times n}$ collect this information and allows to write the matrix $\mathbf{A}$ as a similarity transformation with $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^{-1}$. Whereas it's a common occurrence, the eigenvectors do not need to be normalized, *i.e.*, have norm one, and thus any scalar multiple of $\mathbf{x}_i$ is also an eigenvector of the matrix $\mathbf{A}$.

Some matrices are guaranteed to have this decomposition, among them, symmetric and *normal* matrices. We say that a complex square matrix $\mathbf{A}$ is *normal* if, for its conjugate transpose $\mathbf{A}^*$, we have that $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$. For these matrices, the eigenvectors $\mathbf{U}$ form an unitary matrix with $\mathbf{U}^* = \mathbf{U}^{-1}$, that is, the eigenvectors form an orthonormal basis. If the matrix is also Hermitian and so $\mathbf{A} = \mathbf{A}^*$, then the eigenvalues are all real. We can also obtain a set of real, orthonormal eigenvectors for real symmetric matrices for which $\mathbf{A} = \mathbf{A}^T$, and for this case, the eigendecompositon has the form $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{U}^T$. Further, for real symmetric matrices we use different labels depending on the signs of the eigenvalues and the presence of zeros. We elaborate further about these in Section 2.3.1.1. When two or more eigenvalues are repeated, the associated eigenvectors are not unique, and can be defined by any orthonormal basis of the space spanned by them, that is, if $\mathbf{U} \in \mathbb{R}^{n\times k}$ is a basis for the space, then so is $\mathbf{U}\mathbf{O}$ with $\mathbf{O} \in \mathbb{O}(k \times k)$. Last, the extreme (maximum and minimum) eigenvalues of the matrix can be defined as the optimization problems $\mu_{\min} = \arg\min \mathbf{x}^T\mathbf{A}\mathbf{x}$, subject to $\mathbf{x}^T\mathbf{x} = 1$, and $\mu_{\max} = \arg\max \mathbf{x}^T\mathbf{A}\mathbf{x}$, subject to $\mathbf{x}^T\mathbf{x} = 1$ for the maximum.

2.1.5 Solution to linear systems

An important concept heavily used in this thesis is related to the resolution of linear systems on the variable $\mathbf{x} \in \mathbb{R}^N$ with the form $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{b} \in \mathbb{R}^M$ and $\mathbf{A} \in \mathbb{R}^{M \times N}$. This system appears mainly when estimating the certifiers in Section 2.3.4.

We say the system is underdetermined when the number of variables N is larger than the number of equations M, overdetermined when $N < M$ and square when $N = M$. First we assume the rank of the matrix $\mathbf{A}$ is full and that the system is square or overdetermined, *i.e.*, $N \leq M$, conditions that we will drop later in this section. Even when these conditions hold, a solution to the system may not exist, but it does, as long as the vector $\mathbf{b}$ lays on the range space of the coefficient matrix $\mathbf{A}$, *i.e.* $\mathbf{b} \in \mathcal{R}(\mathbf{A})$. In practice though, we allow this solution $\mathbf{x}$ to distant itself from the range and we seek the one that minimizes the residual error $\epsilon_{\text{i}} = \mathbf{Ax} - \mathbf{b}$ under some norm for ϵ_{i}. The choice of norm will return different solutions, see [Golub and Van Loan, 2013, Sec. 5.3], and we will focus on the norm-2 (least squares) as it is differentiable. Under this norm, we can rely on the so-called *normal equations* $\mathbf{A}^T\mathbf{Ax} = \mathbf{A}^T\mathbf{b}$ that come from the gradient of $\|\mathbf{Ax} - \mathbf{b}\|_2$ being zero at the optimal solution. Different approaches can be leveraged to solve the system, [Golub and Van Loan, 2013, Sec. 5.3,5.4], although we will adopt the SVD-based approach for simplicity. Notice that when the normal equations are used, then the approach is applied on $\mathbf{A}^T\mathbf{A}$ instead of $\mathbf{A}$.

For underdetermined systems with $N > M$, the system may not have a solution or have a family of solutions. Considering that the rank of $\mathbf{A}$ is $r = M$, we denote the nullspace of $\mathbf{A}$ as $\mathcal{N}(\mathbf{A})$. Then, any point on the nullspace, say $\mathbf{y}$ fulfills $\mathbf{Ay} = \mathbf{0}$ and we can define this point as $\mathbf{y} = \mathbf{N}\boldsymbol{\phi}$ where $\mathbf{N}$ is a base for the nullspace $\mathcal{N}(\mathbf{A})$ and $\boldsymbol{\phi} \in \mathbb{R}$ is an unconstrained vector. Further, for *a* solution $\mathbf{x}$ to $\mathbf{Ax} = \mathbf{b}$ we have that any point with form $\mathbf{x} + \mathbf{y} = \mathbf{x} + \mathbf{N}\boldsymbol{\phi}$ is *also* a solution to the linear system as $\mathbf{Ay} = \mathbf{A}(\mathbf{x} + \mathbf{N}\boldsymbol{\phi}) = \mathbf{Ax} = \mathbf{b}$. Further, all these solutions have the same residual error ϵ_{i}, and so they are all optimal. In this case we are interested in the solution $\mathbf{x}$ with minimum 2-norm $\|\mathbf{x}\|_2$. Letting the SVD decomposition of the coefficient matrix be $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ with $\mathbf{U} \in \mathbb{O}(M \times M)$, $\mathbf{V} \in \mathbb{O}(M \times N)$ and $\Sigma \in \mathbb{R}^{M \times M}$ a diagonal matrix whose diagonal entries are the M non-null singular values, we have that the minimum 2-norm solution is given by $\mathbf{x} = \mathbf{V}\Sigma^{-1}\mathbf{U}^T\mathbf{b}$.

For rank-deficient systems where the rank of $\mathbf{A}$ is r, the linear system has a family of solution even if $N \geq M$, as any point on the nullspace of the matrix $\mathbf{A}$ is also a solution with the same associated residual error ϵ_{i}. As with the underdetermined case, here we seek the minimum norm solution. Further, if the the system has $N > M$, then the nullspace also reflects the difference between the number of equations and unknowns. In this case, we work with the thin SVD decomposition of the rank-r matrix $\mathbf{A}$ as $\mathbf{A} = \mathbf{U}\Sigma\mathbf{V}^T$ and so $\mathbf{U} \in \mathbb{O}(N \times r)$, $\mathbf{V} \in \mathbb{O}(r \times M)$ and $\Sigma \in \mathbb{R}^{r \times r}$, which again is a diagonal matrix

with all the non-zero singular values of $\mathbf{A}$. The minimum norm solution $\mathbf{x}$ that also minimizes the residual error is given by the above-mentioned formula. As a last note, we must indicate that the SVD approach allows to solve the system through a rank-k approximation, as a threshold ϵ may be applied to the singular values in Σ, which is of special importance when the matrix $\mathbf{A}$ is almost rank deficient.

2.2 Computer vision foundations

EVERYTHING SOUNDS LOUDER IN THE DARK

Lisa Schneidau

This section summarizes the main concepts about computer vision used in this thesis. We refer the interested readers to the books [Hartley and Zisserman, 2003] and [Ma et al., 2004] for further information. We only treat isolated observations on the image, for example, corners, rather than the full image, and we assume the coordinates of these keypoints have been already extracted when formulating the problems.

2.2.1 Image formation

An image $\mathcal{I}$ is formed as the projection of 3D points $\mathbf{Q}_\mathrm{i}$ in space to observations $\boldsymbol{f}_\mathrm{i}$ onto a compact region Ω on the imaging surface, which we will consider as either $\mathbb{R}^3$ or $\mathbb{S}^2$, and thus either $\Omega \subset \mathbb{R}^3$ or $\Omega \subset \mathbb{S}^2$. According to a central camera model, that projection is defined as a function $\Pi(\bullet) : \mathbb{R}^3 \to \Omega$, where the point $\mathbf{Q}_\mathrm{i} = [q_1, q_2, q_3]^T \in \mathbb{R}^3$ referred to the camera frame takes the form $\Pi(\mathbf{Q}_\mathrm{i}) = [q_1/q_3, q_2/q_3, 1]^T \in \mathbb{R}^3$ for planar projection or $\Pi(\mathbf{Q}_\mathrm{i}) = \mathbf{Q}_\mathrm{i}/\|\mathbf{Q}_\mathrm{i}\|_2 \in \mathbb{S}^2$ for a spherical projection. Notice that in both cases, the observation $\boldsymbol{f}_\mathrm{i}$ and its associated 3D point $\mathbf{Q}_\mathrm{i}$ are related through an (unknown) scale λ as $\lambda \boldsymbol{f}_\mathrm{i} = \mathbf{Q}_\mathrm{i}$, and the specific value of λ depends on the considered projection. In addition, observations in the planar projection, which have the form $\tilde{\mathbf{q}} = [a, b, 1]^T \in \mathbb{R}^3$ are also called *homogeneous*, and will be useful to write affine operations of vectors as linear transformations.

Point transformation: The projection function Π requires the 3D point to be referred to the camera coordinate frame. Thus the point $\mathbf{Q}_\mathrm{i} = [q_1, q_2, q_3]^T \in \mathbb{R}^3$ with coordinates in a different frame, *e.g.*, the world frame or the initial camera pose, is transformed to $\mathbf{Q}'_\mathrm{i}$ in the second frame, by considering the relative pose $(\mathbf{R}, \mathbf{t}) \in \mathbb{SO}(3) \times \mathbb{R}^3$ between these two poses as $\mathbf{Q}'_\mathrm{i} = \mathbf{R}\mathbf{Q}_\mathrm{i} + \mathbf{t}$. The transformation can be done more conveniently with homogeneous vectors $\tilde{\mathbf{Q}}'_\mathrm{i} = [\mathbf{Q}'^T_\mathrm{i}, 1]^T \in \mathbb{R}^4$, $\tilde{\mathbf{Q}}_\mathrm{i} = [\mathbf{Q}^T_\mathrm{i}, 1]^T \in \mathbb{R}^4$ using the 4×4 transformation matrix $\tilde{\mathbf{T}}$ as

$$\tilde{\mathbf{Q}}'_\mathrm{i} = \tilde{\mathbf{T}}\tilde{\mathbf{Q}}_\mathrm{i} \quad \text{with } \tilde{\mathbf{T}} = \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}. \tag{2.20}$$

This model assumes the image frame is at the second frame, which holds for central cameras. In a general configuration, though, we may have more than one camera per system, that is, we have available a set or rig of M cameras, see Figure 2.1. In this case, the pose associated with each image or camera is different from the system center with a known relative pose $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j}) \in$

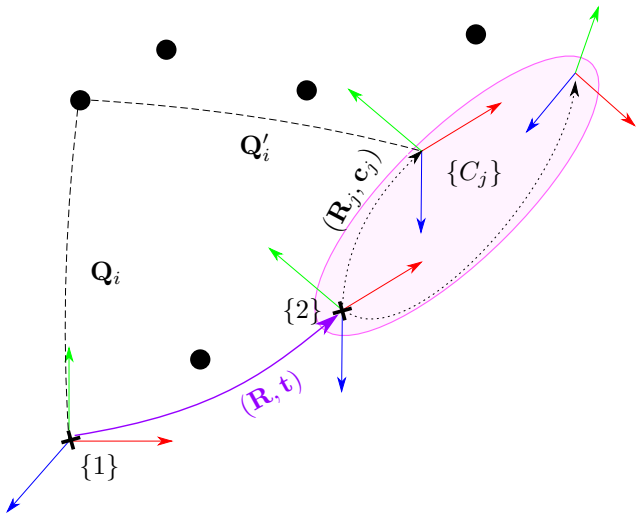


Figure 2.1: Scheme of a generalized camera $\{2\}$ with inter-cameras $\{C_j\}$ and relative poses $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ *w.r.t.* the main frame. A point $\mathbf{Q}_\mathrm{i}$ *w.r.t.* the frame $\{1\}$ is transformed to $\mathbf{Q}'_\mathrm{i}$ with the relative pose $(\mathbf{R}, \mathbf{t})$ between $\{1\}$ and $\{2\}$ and the inter-camera pose $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ as in Equation (2.21).

$\mathbb{SO}(3) \times \mathbb{R}^3$ *w.r.t.* the common frame. This configuration of cameras and the associated image formation model are defined by the *generalized camera model*, which also explains central cameras. The generalized model transforms the point $\mathbf{Q}_\mathrm{i}$ in the first frame onto the j-th camera with pose $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ *w.r.t.* the second camera frame as:

$$\mathbf{Q}'_\mathrm{i} = \mathbf{R}_\mathrm{j}^T(\mathbf{R}\mathbf{Q}_\mathrm{i} + \mathbf{t} - \mathbf{c}_\mathrm{j}) \Leftrightarrow \tilde{\mathbf{Q}}'_\mathrm{i} = \tilde{\mathbf{T}}_j \tilde{\mathbf{Q}}_\mathrm{i}, \quad \text{with:} \tilde{\mathbf{T}}_j \doteq \begin{pmatrix} \mathbf{R}_\mathrm{j}^T\mathbf{R} & \mathbf{R}_\mathrm{j}^T(\mathbf{t} - \mathbf{c}_\mathrm{j}) \\ \mathbf{0}_{1\times 3} & 1 \end{pmatrix}. \tag{2.21}$$

Notice that a central camera has $M = 1$ and can be seen as a special case of the generalized configuration in Equation (2.21) with $\mathbf{R}_\mathrm{j} = \boldsymbol{I}_3$ and $\mathbf{c}_\mathrm{j} = \mathbf{0}_{3\times 1}$.

2.2.2 Relative pose between two cameras

As introduced above, the observations of the same point on different images capture the information about the (relative) pose of said cameras. Thus, the observation of $\mathbf{Q}_\mathrm{i}$ on the first camera following a perspective camera model with projection function $\Pi(\bullet)$ is $\boldsymbol{f}_\mathrm{i} = \Pi(\mathbf{Q}_\mathrm{i})$, and the corresponding observation $\boldsymbol{f}'_\mathrm{i}$ in the second image has also the form $\boldsymbol{f}'_\mathrm{i} = \Pi(\mathbf{Q}'_\mathrm{i})$ for the transformed point $\mathbf{Q}'_\mathrm{i}$. We can then estimate the relative pose given these observations $\boldsymbol{f}_\mathrm{i}$ (or $\boldsymbol{f}'_\mathrm{i}$)

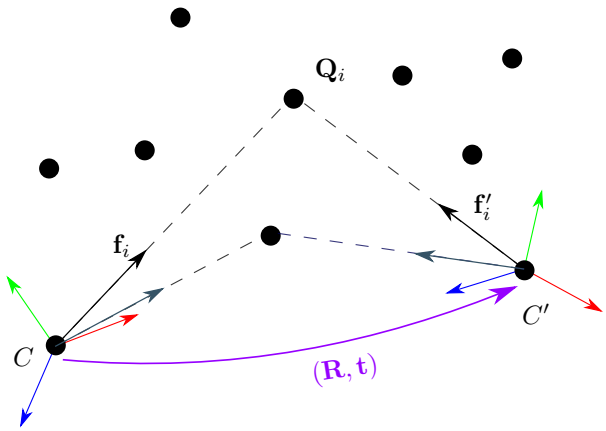


Figure 2.2: Given a set of N pair-wise correspondences $\{\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i}\}$, the goal of the relative pose problem between two cameras is to estimate the rotation $\mathbf{R}$ and translation $\mathbf{t}$.

and the 3D points $\mathbf{Q}'_\mathrm{i}$ (or $\mathbf{Q}_\mathrm{i}$) that originated them, which is usually stated as the minimization of the reprojection error for all the observations

$$h^\star = \min_{\mathbf{R}\in\mathbb{SO}(3),\mathbf{t}\in\mathbb{R}^3} \sum_{i=1}^{N} \mathrm{dist}(\boldsymbol{f}'_\mathrm{i}, \Pi(\mathbf{Q}'_\mathrm{i}))^2 = \sum_{i=1}^{N} \mathrm{dist}(\boldsymbol{f}'_\mathrm{i}, \Pi(\mathbf{R}\mathbf{Q}_\mathrm{i} + \mathbf{t})) \tag{2.22}$$

where the function $\mathrm{dist}(A, B)$ measures the distance between A and B according to some appropriate metric.

However, in most cases the only available information to solve this problem is the set of N observations, since the 3D points $\mathbf{Q}_\mathrm{i}$ are unknown. Solving for these points *and* the pose is the gold-standard approach, known as Bundle-Adjustment (BA). Introducing the 3D points into the formulation increases the number of variables and the overall complexity of the problem. Further, the internal restrictions on the rotation make the problem nonconvex and the large number of variables require a good initialization to converge fast. The common approach for initialization is to get and estimate of the relative pose without computing explicitly the coordinates of the 3D points. If the 3D points are in general position, the essential matrix is used to estimate this pose, while if the points belong to a plane π, the homography matrix is used instead. The rest of this section focuses on estimating the relative pose $(\mathbf{R}, \mathbf{t})$ between the two camera frames C, C' given *only* the N pair-wise observations $\{\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i}\}$ for $i = [N]$ without the 3D points $\mathbf{Q}_\mathrm{i}$ as we show in Figure 2.2.

2.2.2.1 Essential matrix estimation

As mentioned above, the point $\mathbf{Q}_\mathrm{i}$ *w.r.t.* the first camera frame is transformed through the relative pose between cameras $(\mathbf{R}, \mathbf{t})$ as the point $\mathbf{Q}'_\mathrm{i} = \mathbf{R}\mathbf{Q}_\mathrm{i} + \mathbf{t}$. The associated observations to each point are $\boldsymbol{f}_\mathrm{i}$ and $\boldsymbol{f}'_\mathrm{i}$, respectively, and are modeled by the projection function Π. Considering explicitly the non-zero scales $\lambda_1, \lambda_2 \neq 0$ for these projections we have the relation

$$\lambda_2 \boldsymbol{f}'_\mathrm{i} = \mathbf{R}\lambda_1 \boldsymbol{f}_\mathrm{i} + \mathbf{t}. \tag{2.23}$$

To eliminate the scales λ_1, λ_2 we first take the cross-product with $\mathbf{t}$, denoted by $[\mathbf{t}]_\times = \mathbf{t}\times$, and thus

$$\lambda_2 [\mathbf{t}]_\times \boldsymbol{f}'_\mathrm{i} = \lambda_1 [\mathbf{t}]_\times \mathbf{R} \boldsymbol{f}_\mathrm{i}, \tag{2.24}$$

where $[\mathbf{t}]_\times \mathbf{t} = \mathbf{0}_{3\times 1}$ by construction. Further, since $[\mathbf{t}]_\times \boldsymbol{f}'_\mathrm{i}$ is perpendicular to $\boldsymbol{f}'_\mathrm{i}$, then the inner product $\langle \boldsymbol{f}'_\mathrm{i}, [\mathbf{t}]_\times \boldsymbol{f}'_\mathrm{i} \rangle = \boldsymbol{f}'^T_\mathrm{i} [\mathbf{t}]_\times \boldsymbol{f}'_\mathrm{i} = 0$ and from Equation (2.23) we have that

$$\langle \boldsymbol{f}'_\mathrm{i}, [\mathbf{t}]_\times \boldsymbol{f}'_\mathrm{i} \rangle = \boldsymbol{f}'^T_\mathrm{i} ([\mathbf{t}]_\times \mathbf{R} \boldsymbol{f}_\mathrm{i}) = 0, \tag{2.25}$$

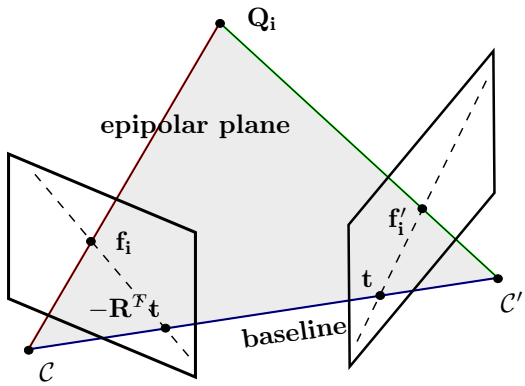
where the scales $\lambda_1, \lambda_2 \neq 0$ are not required since they do not affect the product.

This relation, known as the *epipolar constraint*, is the cornerstone of relative pose estimation algorithms. Geometrically, the epipolar constraint indicates that the translation vector $\mathbf{t}$, the observation $\boldsymbol{f}'_\mathrm{i}$ and $\boldsymbol{f}_\mathrm{i}$ seeing from the second camera frame are coplanar, see Figure 2.3a. The plane formed by these vectors is known as the *epipolar plane*, and the lines of intersection with the image planes are known as *epipolar lines*. For spherical models, the epipolar lines are the great circles as shown in Figure 2.3b, and in both cases, the points where the image planes/spheres intersect the line joining both optical centers, that is, the translation vector, are known as the *epipoles*. Further, the matrix $\mathbf{E}$ with the form $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$ is known as the *essential matrix* and we say it is the *normalized* essential matrix if $\mathbf{t} \in \mathbb{S}^2$ and so $\|\mathbf{E}\|_\mathrm{F}^2 = 2$. Notice from Equation (2.25) that any scale, including a sign, does not alter the relation and so $\mathbf{E}$ and $\pm\alpha\mathbf{E}$ for $\alpha \in \mathbb{R}$ are solutions. Therefore, we will work with the normalized essential matrix, and denote the space formed by all of them by $\mathbb{E}$, defined as

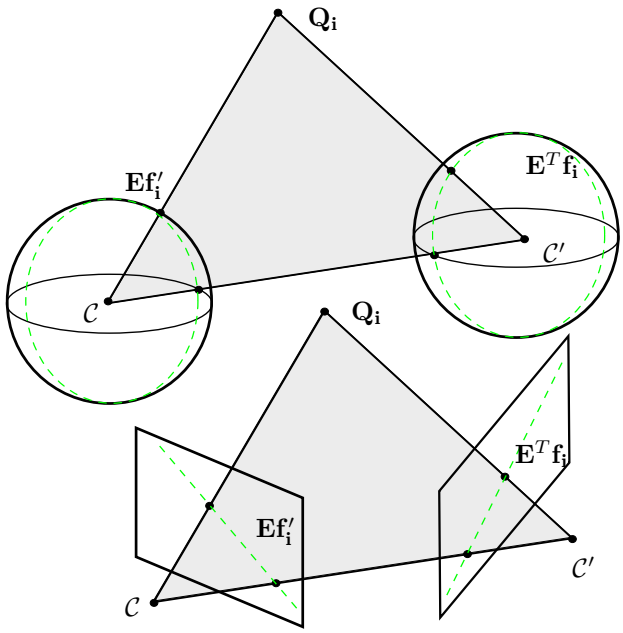
$$\mathbb{E} \equiv \left\{ \mathbf{E} = [\mathbf{t}]_\times \mathbf{R} \;|\mathbf{R} \in \mathbb{SO}(3), \mathbf{t} \in \mathbb{S}^2 \right\}. \tag{2.26}$$

An alternative form for the essential matrix is given by $\mathbf{E} = \mathbf{R}[\mathbf{R}^T\mathbf{t}]_\times$ from the relation $\lambda_1 \boldsymbol{f}_\mathrm{i} = \mathbf{R}^T \lambda_2 \boldsymbol{f}'_\mathrm{i} - \mathbf{R}^T\mathbf{t}$, taking the cross-product with $-\mathbf{R}^T\mathbf{t}$ instead of $\mathbf{t}$ and transposing the result as $(-\boldsymbol{f}'^T_\mathrm{i}[\mathbf{R}^T\mathbf{t}]_\times \mathbf{R}^T \boldsymbol{f}'_\mathrm{i})^T = \boldsymbol{f}'^T_\mathrm{i} \left(\mathbf{R}[\mathbf{R}^T\mathbf{t}]_\times\right) \boldsymbol{f}_\mathrm{i} = 0$. Equivalently, we can algebraically manipulate the product $[\mathbf{t}]_\times \mathbf{R}$ with the relation in [Hartley and Zisserman, 2003, Res. A4.3] as, up to scale, $[\mathbf{t}]_\times \mathbf{M} = \mathbf{M}^{-T}[\mathbf{M}^{-1}\mathbf{t}]_\times$ where for a rotation matrix $\mathbf{R}$ we have that $\mathbf{R}^{-1} = \mathbf{R}^T$ and so $\mathbf{R}^{-T} = \mathbf{R}$.

The terms $\mathbf{E}\boldsymbol{f}_\mathrm{i}$ and $\mathbf{E}^T\boldsymbol{f}'_\mathrm{i}$ are associated with the epipolar lines, as shown in Figure 2.3b. These lines pass through the *epipole* of the images, that is, the



(a) Elements associated with the epipolar geometry between two views $(\mathcal{C}, \mathcal{C}')$.



(b) Epipolar lines as green, dashed lines for a spherical (top) and planar (bottom) projection

Figure 2.3: Figure 2.3a shows the epipolar plane cuts the image planes on the observations $(\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i})$ and the epipoles $(\mathbf{R}^T\mathbf{t}, \mathbf{t})$. Figure 2.3b shows the epipolar great circles or lines pass through the epipole and the observation.

points $\mathbf{t}$ and $-\mathbf{R}^T\mathbf{t}$, and the corresponding observation $\boldsymbol{f}_\mathrm{i}$ and $\boldsymbol{f}'_\mathrm{i}$, provided both these features and the essential matrix are noiseless. The metrics $\boldsymbol{f}_\mathrm{i}'^T\mathbf{E}\boldsymbol{f}_\mathrm{i}$ and $\boldsymbol{f}_\mathrm{i}^T\mathbf{E}^T\boldsymbol{f}'_\mathrm{i}$ measure the distance of the given observation to the epipolar line, and is known as the *epipolar error* $\epsilon_\mathrm{i} = \boldsymbol{f}_\mathrm{i}'^T\mathbf{E}\boldsymbol{f}_\mathrm{i} = \boldsymbol{f}_\mathrm{i}^T\mathbf{E}^T\boldsymbol{f}'_\mathrm{i}$. The term ϵ_i is zero when the observations $\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i}$ and the essential matrix $\mathbf{E}$ are noiseless and/or when at least one of the observations is to the corresponding epipole. Otherwise, the term ϵ_i can be used as optimization error, and a common approach to tackle the relative pose problem is to minimize the sum of the squared error seeking the essential matrix instead of the pose, formally

$$f^\star = \min_{\mathbf{E}\in\mathbb{E}} \sum_{i=1}^{N} \epsilon_\mathrm{i}^2 = \min_{\mathbf{E}\in\mathbb{E}} \sum_{i=1}^{N} (\boldsymbol{f}_\mathrm{i}'^T\mathbf{E}\boldsymbol{f}_\mathrm{i})^2 \tag{2.27}$$

Although the epipolar error is widely used, other expressions also capture the relationship between the corresponding observations, see [Hartley and Zisserman, 2003, Sec. 12.3]. One of the main advantages of the epipolar error in comparison with other metrics is its linearity in the unknown essential matrix $\mathbf{E}$, also making the expression ϵ_i^2 quadratic. The linear form is leveraged by minimal solvers to develop a polynomial system to retrieve a solution for this problem, *e.g.*, [Nistér, 2004].

On the other hand, we can re-write the cost function of problem Equation (2.27) as a quadratic form. For that, we consider the column-wise vectorization of $\mathbf{E}$ as $\mathbf{e} = \mathrm{vec}(\mathbf{E}) \in \mathbb{R}^9$ and the identity $\mathrm{tr}(\mathbf{A}^T\mathbf{B}\mathbf{C}) = (\mathbf{C}^T \otimes \mathbf{A}^T)\,\mathrm{vec}(\mathbf{B}) = (\mathbf{C} \otimes \mathbf{A})^T\,\mathrm{vec}(\mathbf{B})$, making the epipolar error $\epsilon_\mathrm{i} = (\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})^T\mathbf{e}$. Therefore, the squared term is $\epsilon_\mathrm{i}^2 = \epsilon_\mathrm{i}^T\epsilon_\mathrm{i} = \mathbf{e}^T\left((\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})(\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})^T\right)\mathbf{e}$, and we define the contribution of each pair-wise correspondence as $\mathbf{C}_\mathrm{i} \doteq (\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})(\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})^T \in \mathbb{S}_+^9$. The total cost, which is the sum of each contribution, has the final form

$$f(\mathbf{E}) = \sum_{i=1}^{N} \epsilon_\mathrm{i}^2 = \sum_{i=1}^{N} \mathbf{e}^T\left((\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})(\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})^T\right)\mathbf{e} = \tag{2.28}$$

$$= \mathbf{e}^T\left(\sum_{i=1}^{N} (\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})(\boldsymbol{f}_\mathrm{i} \otimes \boldsymbol{f}'_\mathrm{i})^T\right)\mathbf{e} = \mathbf{e}^T\mathbf{C}\mathbf{e}, \tag{2.29}$$

where $\mathbf{C} = \sum_{i=1}^{N} \mathbf{C}_\mathrm{i} \in \mathbb{S}_+^9$ is the coefficient, cost or data matrix. This same matrix appears in the Direct Linear Transformation (DLT), and thus it can be used to approximate the solution to the problem. For that, recall that the epipolar error is zero for noiseless correspondences, that is, for these cases, the total error is also zero $\sum_{i=1}^{N} \epsilon_\mathrm{i}^2 = 0$, and the solution $\mathbf{E}$ makes the cost zero $f(\mathbf{E}) = 0$. Since the matrix $\mathbf{C}$ is Positive Semidefinite (PSD) by construction, this solution must lay on the nullspace of the matrix, *i.e.*, $\mathbf{C}\mathbf{e} = \mathbf{0}_{9\times1}$. If the nullspace is one-dimensional, then we can retrieve this solution through the

eigendecomposition of $\mathbf{C}$ as the eigenvector associated to the zeroed eigenvalue. However, this method only returns an essential matrix for well-defined, noiseless problem instances, as the eigenvector does not fulfill a priori any of the requirements of the set of essential matrices. Nevertheless, we can still leverage DLT for noisy data, expecting the smallest eigenvalue to be nonzero when the number of correspondences N is larger than 8, and thus this approach only returns an approximated solution. Even for data with low noise, it is always recommended to find the closest essential matrix to the solution output by DLT, which can be used to initialize an iterative algorithm. Nonetheless, this approach cannot be used directly when the cost matrix has more than one eigenvalue zero, which occurs, for instance, when all the observed points lay on a plane. In these cases, we say the problem is *degenerate* since multiple solutions achieve the same cost. Another *ill-conditioned* example appears when the number of correspondences is not enough to obtain isolated estimations. For both cases, though, we have a family of solutions for the given data, and the problem is not fully determined.

Singular value decomposition for an essential matrix: The Singular Value Decomposition (SVD) of an essential matrix $\mathbf{E}$ is given by $\mathbf{E} = \mathbf{U}\mathbf{D}\mathbf{V}^T$, where $\mathbf{U}, \mathbf{V} \in \mathbb{SO}(3)$ and $\mathbf{D}$ is a diagonal matrix with the associated singular values, which are required to be $(\alpha, \alpha, 0)$, see *e.g.*, [Ma et al., 2004, Th. 5.1]. For normalized essential matrices we have that $\|\mathbf{E}\|_{\mathrm{F}}^2 = \mathrm{tr}(\mathbf{E}\mathbf{E}^T) = \mathrm{tr}(\mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{D}\mathbf{U}^T) = \mathrm{tr}(\mathbf{D}^2) = 2\alpha^2 = 2$ and so $\alpha = 1$. Note, however, that the SVD is not unique due to having a repeated singular value, and so there exists a one-parameter family of solutions, as for any 2×2 rotation matrix $\mathbf{X} \in \mathbb{SO}(2)$ we also have that $\mathbf{E} = (\mathbf{U}\,\mathrm{diag}(\mathbf{X}, 1))\,\mathbf{D}\,(\mathbf{V}\,\mathrm{diag}(\mathbf{X}, 1))^T$ is also a valid decomposition. This parameterization will be useful when (1) seeking the closest essential matrix $\mathbf{E}$ to another matrix $\hat{\mathbf{E}}$; and (2) obtaining the rotation and translation from a valid $\mathbf{E}$.

Projection onto the set of essential matrices: As we commented above, in some cases it may be necessary to obtain the closest essential matrix $\mathbf{E}$, that is, a point in the set of essential matrices $\mathbb{E}$, to a matrix $\hat{\mathbf{E}} \in \mathbb{R}^{3\times3}$. Formally, we aim to solve the problem for $\mathbf{E}$

$$g^\star = \min_{\mathbf{E}\in\mathbb{E}} \mathrm{dist}(\mathbf{E}, \hat{\mathbf{E}}). \tag{2.30}$$

We consider usually the distance in the Frobenius norm sense as $\mathrm{dist}(\mathbf{E}, \hat{\mathbf{E}}) = \left\|\mathbf{E} - \hat{\mathbf{E}}\right\|_{\mathrm{F}}^2$, see [Ma et al., 2004, Th. 5.3]. For normalized essential matrices with $\|\mathbf{E}\|_{\mathrm{F}}^2 = 2$, the problem, up to a constant factor, requires to maximize $\mathrm{tr}(\mathbf{E}^T\hat{\mathbf{E}})$ over $\mathbf{E}$. The solution is estimated in closed form and takes the form $\mathbf{E} = \mathbf{U}\Sigma\mathbf{V}^T$, with $\hat{\mathbf{E}} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ the SVD decomposition of $\hat{\mathbf{E}}$ and $\mathbf{D}$ the diagonal matrix whose entries are (a, b, c) and $a \geq b \geq c \geq 0$. For the solution

$\mathbf{E}$ we make $\Sigma = \text{diag}(\alpha, \alpha, 0)$ with $\alpha = (a+b)/2$. Last, the error of this approximation, $g^\star$, is given by $\left\|\mathbf{E} - \hat{\mathbf{E}}\right\|_{\text{F}}^2 = \|\Sigma - \mathbf{D}\|_{\text{F}} = (\alpha-a)^2+(\alpha-b)^2+c^2$.

Recovering a valid pose from an essential matrix: Approaches that rely on the essential matrix ultimately require to retrieve the associated pose from $\mathbf{E}$. Such a recovery is not direct, as there are four combinations of poses that give rise to the same essential matrix $\mathbf{E}$: (1) the translation $\mathbf{t}$ and its negative $-\mathbf{t}$; and (2) the rotations $\mathbf{R}$ and $\mathbf{PR}$, where $\mathbf{P} \doteq \boldsymbol{I}_3 - \mathbf{t}\mathbf{t}^T$ is the reflection *w.r.t.* the vector $\mathbf{t}$. These four combinations are the valid poses associated to the same $\mathbf{E}$, [Hartley and Zisserman, 2003, Sec. 9.6.2]. The first pair $\pm\mathbf{t}$ comes from the scale ambiguity of the essential matrix, that includes also the sign. The pair of rotations, known as *twisted pair*, comes from the identity $[\mathbf{t}]_{\text{x}} = [\mathbf{t}]_{\text{x}}\mathbf{P} = [\mathbf{t}]_{\text{x}}\left(\boldsymbol{I}_3 - \mathbf{t}\mathbf{t}^T\right)$ as by definition $[\mathbf{a}]_{\text{x}}\mathbf{a} = \mathbf{0}_{3\times1}$ for any 3D vector $\mathbf{a}$ and so $[\mathbf{t}]_{\text{x}}\mathbf{PR} = [\mathbf{t}]_{\text{x}}\mathbf{R}$. The "true" pose is selected as the one that triangulates most points in front of the camera, if there exists a "front" of the camera, or which agrees with most points in general, for instance, for spherical cameras. This approach, however, requires to triangulate the points, which can be avoided using alternative algorithms *e.g.*, [Cai et al., 2019].

2.2.2.2 Homography matrix

When the 3D points belong to a given plane, these points and their respective observations are related through a perspective transformation with the form of a 3×3 full-rank matrix known as the *homography matrix* $\mathbf{H}$. This relation also appears when the translation between the cameras is zero, and so $\mathbf{E} = \mathbf{0}_{3\times3}$ since $[\mathbf{t}]_{\text{x}} = \mathbf{0}_{3\times3}$, for example, in panoramic images.

The homography matrix states a point-to-point relationship between the i-th 3D point $\mathbf{Q}_{\text{i}}$ and its associated observation $\mathbf{q}_{\text{i}}$ as $\mathbf{q}_{\text{i}} \sim \mathbf{H}_1\mathbf{Q}_{\text{i}}$ where $\sim$ indicates equality up-to-scale. As all the points belonging to the plane fulfill the homography constraint, we can retrieve the 3×3 matrix from these pairwise correspondences, as we did with the essential matrix. Furthermore, the same plane induces homographies in other images (also planes), as shown in Figure 2.4 and for the observation $\boldsymbol{f}'_{\text{i}}$ on the second image the relation $\boldsymbol{f}'_{\text{i}} \sim \mathbf{H}_2\mathbf{Q}_{\text{i}}$ also holds. As $\mathbf{H}_1, \mathbf{H}_2$ are full-rank, we can combine them into the image-image homography $\mathbf{H} \doteq \mathbf{H}_2\mathbf{H}_1^{-1}$ that relates the observations $\boldsymbol{f}_{\text{i}}, \boldsymbol{f}'_{\text{i}}$ as $\boldsymbol{f}'_{\text{i}} \sim \mathbf{H}\boldsymbol{f}_{\text{i}}$.

From the relation between the 3D point and the observations, we can derive the equation for each homography $\mathbf{H}_1, \mathbf{H}_2$ and thus also $\mathbf{H}$. For that, we consider the homography induced by the plane $\pi_S = [\mathbf{n}^T, d]^T \in \mathbb{R}^4$ where $\mathbf{n} \in \mathbb{S}^2$ is the normal to the plane and $d \in \mathbb{R}$ the minimum distance from the plane, all with respect to the original frame. The pose between the original and the first frame is $(\mathbf{R}_1, \mathbf{t}_1)$ and we know that points $\mathbf{x} \in \mathbb{R}^3$ belonging to the plane must fulfill the equation $\mathbf{n}^T\mathbf{x} + d = 0$, which can be expressed as $\pi_S^T\mathbf{X} = 0$,

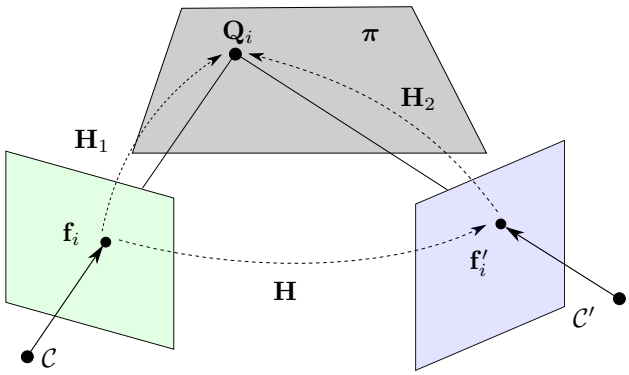


Figure 2.4: Relations induced by the plane π in the space given by the matrices $\mathbf{H}, \mathbf{H}_1, \mathbf{H}_2$ induced by the plane π between a point $\mathbf{Q}_\mathrm{i}$ on the plane and the associated observations in the two cameras $\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i}/$

with $\mathbf{X} = [\mathbf{x}^T, -\mathbf{n}^T\mathbf{x}/d]^T \in \mathbb{R}^4$. Transforming this point with the relative pose $(\mathbf{R}_1, \mathbf{t}_1)$ or its equivalent transformation matrix $\mathbf{T}_1$ leads to

$$\tilde{\mathbf{Q}}_\mathrm{i} = \mathbf{T}_1\tilde{\mathbf{X}} \quad \text{and} \quad \mathbf{Q}_\mathrm{i} = \mathbf{R}_1\mathbf{x} - (\mathbf{t}_1\mathbf{n}^T\mathbf{x})/d = \left(\mathbf{R}_1 - \mathbf{t}_1\mathbf{n}^T/d\right)\mathbf{x}, \tag{2.31}$$

and thus the observation $\boldsymbol{f}_\mathrm{i}$ associated with $\mathbf{Q}_\mathrm{i}$ is $\boldsymbol{f}_\mathrm{i} = \Pi(\mathbf{Q}_\mathrm{i}) \sim \left(\mathbf{R}_1 - \mathbf{t}_1\mathbf{n}^T/d\right)\mathbf{x}$. The homography matrix is defined by $\mathbf{H}_1 \doteq \mathbf{R}_1 - \mathbf{t}_1\mathbf{n}^T/d$, and whereas $\mathbf{n}$ is an unit-norm vector, $\mathbf{t}_1$ is actually unconstrained, so the distance d is absorb by it without loss of generality, and we usually write the expression for the homography as $\mathbf{H}_1 = \mathbf{R}_1 - \mathbf{t}_1\mathbf{n}^T$. A similar development allows us to obtain the homography $\mathbf{H}_2$ between the 3D plane π_S and the second frame with pose $(\mathbf{R}_2, \mathbf{t}_2)$, relating the observation on the second image $\boldsymbol{f}'_\mathrm{i}$ with the 3D point $\mathbf{x}$ as $\boldsymbol{f}'_\mathrm{i} \sim \mathbf{H}_2\mathbf{x}$. The third and last homography $\mathbf{H}$ between the image planes can be computed either by combining the previous two as $\mathbf{H} = \mathbf{H}_2\mathbf{H}_1^{-1}$ or with the above-mentioned procedure. In this case, the pose $(\mathbf{R}, \mathbf{t})$ collected by the homography is the relative pose between the images (or cameras) while the normal $\mathbf{n}$ is referred to the first frame, and the relation is $\boldsymbol{f}'_\mathrm{i} \sim \mathbf{H}\boldsymbol{f}_\mathrm{i}$. Notice that for all the cases the rotation has three degrees of freedom, so does the translation and the normal $\mathbf{n}$ has only two, hence endowing the general homography matrix with 8 degrees of freedom for 9 entries. Since all the relations are only known up-to-scale, there is an additional degree of freedom in the overall scale of the matrix $\mathbf{H}$.

As with the essential matrix, we leverage the relation between the homography and the observations as an error for the optimization, and we focus exclusively on the observation-observation relation given by $\mathbf{H}$. In this case, we work with the expression $\boldsymbol{f}'_\mathrm{i} \sim \mathbf{H}\boldsymbol{f}_\mathrm{i}$ and to remove the scale ambiguity, we take the cross-product with $\boldsymbol{f}'_\mathrm{i}$ as $[\boldsymbol{f}'_\mathrm{i}]_\times\mathbf{H}\boldsymbol{f}_\mathrm{i} = \mathbf{0}_{3\times1}$, which provides three

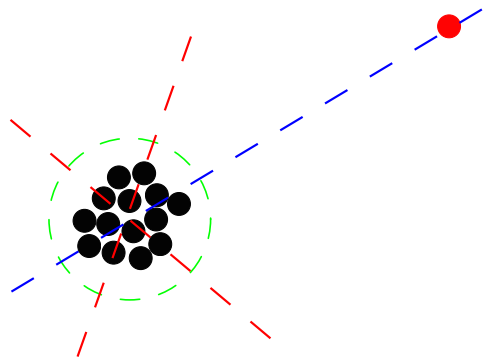


Figure 2.5: Set of inliers as black dots can be defined by the green circle (1D) but also the green lines (2D). The blue line fits the inliers *and* the outlier in red.

expressions, the *homography constraints*, although only two are linearly independent. We can re-write each column of the matrix $[\boldsymbol{f}'_{\mathrm{i}}]_{\mathrm{x}}$, which are linear in $\boldsymbol{f}'_{\mathrm{i}}$, with a sparse matrix $\mathbf{B}_i$ such that the i-th column is $\mathbf{B}_i\boldsymbol{f}'_{\mathrm{i}}$. The matrices $\mathbf{B}_i$ for $i = 1, 2, 3$ have the form

$$\mathbf{B}_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}, \; \mathbf{B}_2 = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{B}_3 = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (2.32)$$

Since the matrix $[\boldsymbol{f}'_{\mathrm{i}}]_{\mathrm{x}}$ is skew-symmetric, then $[\boldsymbol{f}'_{\mathrm{i}}]_{\mathrm{x}}^T = -[\boldsymbol{f}'_{\mathrm{i}}]_{\mathrm{x}}$ and the i-th row is the i-th column with a minus sign. Therefore, each of the entries for $[\boldsymbol{f}'_{\mathrm{i}}]_{\mathrm{x}}\mathbf{H}\boldsymbol{f}_{\mathrm{i}}$ can be formulated as $-\boldsymbol{f}_{\mathrm{i}}'^T\mathbf{B}_i^T\mathbf{H}\boldsymbol{f}_{\mathrm{i}}$, where the unknown is the homography $\mathbf{H}$. We define the two errors as $\epsilon_{\mathrm{i}1} = -\boldsymbol{f}_{\mathrm{i}}'^T\mathbf{B}_1^T\mathbf{H}\boldsymbol{f}_{\mathrm{i}}$ and $\epsilon_{\mathrm{i}2} = -\boldsymbol{f}_{\mathrm{i}}'^T\mathbf{B}_2^T\mathbf{H}\boldsymbol{f}_{\mathrm{i}}$ and the cost to be minimized is the sum of the squared of these errors. As with the essential matrix, the relation between the trace $\epsilon_{\mathrm{i}1} = \mathrm{tr}(\epsilon_{\mathrm{i}1})$ and the kronecker product allows us to write this expression as $\mathrm{tr}(\epsilon_{\mathrm{i}1}) = (\boldsymbol{f}_{\mathrm{i}} \otimes (\mathbf{B}_i\boldsymbol{f}'_{\mathrm{i}}))^T \mathrm{vec}(\mathbf{H})$ and thus each correspondence contributes

$$f_i(\mathbf{H}) = \epsilon_{\mathrm{i}1}^2 + \epsilon_{\mathrm{i}2}^2 = \mathrm{vec}(\mathbf{H})^T \underbrace{(\boldsymbol{f}_{\mathrm{i}} \otimes (\mathbf{B}_1\boldsymbol{f}'_{\mathrm{i}}))\,(\boldsymbol{f}_{\mathrm{i}} \otimes (\mathbf{B}_1\boldsymbol{f}'_{\mathrm{i}}))^T}_{\mathbf{C}_{1i}} \mathrm{vec}(\mathbf{H}) +$$

$$+ \mathrm{vec}(\mathbf{H})^T \underbrace{(\boldsymbol{f}_{\mathrm{i}} \otimes (\mathbf{B}_2\boldsymbol{f}'_{\mathrm{i}}))\,(\boldsymbol{f}_{\mathrm{i}} \otimes (\mathbf{B}_2\boldsymbol{f}'_{\mathrm{i}}))^T}_{\mathbf{C}_{2i}} \mathrm{vec}(\mathbf{H}) = \quad (2.33)$$

$$= \mathrm{vec}(\mathbf{H})^T (\mathbf{C}_{1i} + \mathbf{C}_{2i}) \mathrm{vec}(\mathbf{H}) = \mathrm{vec}(\mathbf{H})^T \mathbf{C}_{\mathrm{i}} \mathrm{vec}(\mathbf{H}) \quad (2.34)$$

and the minimization of the total cost, which considers all the correspondences, leads to the problem

$$f^\star = \min_{\mathbf{H}\in\mathbb{R}^{3\times3}} \operatorname{vec}(\mathbf{H})^T \left(\sum_{i=1}^{N} (\mathbf{C}_{1_i} + \mathbf{C}_{2i}) \right) \operatorname{vec}(\mathbf{H}) = \min_{\mathbf{H}\in\mathbb{R}^{3\times3}} \operatorname{vec}(\mathbf{H})^T \mathbf{C} \operatorname{vec}(\mathbf{H}) \tag{2.35}$$

where $\sum_{i=1}^{N} (\mathbf{C}_{1_i} + \mathbf{C}_{2i}) = \mathbf{C} \in \mathbb{S}_+^9$ is the coefficient or cost matrix. As the general homography matrix has 8 degrees of freedom plus the scale ambiguity, there are no internal restrictions on its entries. Since each correspondence provides two independent constraints, four pair-wise correspondences are the minimum required to obtain a unique solution, which is found trivially through the matrix $\mathbf{C}$ by computing its eigendecomposition as the solution is the eigenvector associated to the null eigenvalue. The approach is applied to the non-minimal case by taking instead the eigenvector associated with the smallest eigenvalue. Notice that in that case, no projection is needed as the retrieved solution is already a homography matrix. However, this only applies when the homography matrix has the 8 degrees of freedom, that is, no prior information or assumptions about the pose and/or plane exist, see Section 2.2.2.3. Once the homography has been estimated, it is necessary to recover the pose and plane normal from the 3×3 matrix. This process, although tractable, turns to be unstable, and previous works have approached it with different performances see *e.g.* [Malis and Vargas, 2007].

Last, and as above-mentioned, applying the essential matrix to problems with degenerate configuration results in a family of solutions. However, the homography matrix and the essential matrix associated to the pose are related, as they both collect information about the motion. This can be seen from their respective definitions with $\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{n}^T$ and $\mathbf{E} = [\mathbf{t}]_\times \mathbf{R}$ and thus we have that $\mathbf{H}^T\mathbf{E} = (\mathbf{R}^T + \mathbf{n}\mathbf{t}^T)[\mathbf{t}]_\times\mathbf{R} = \mathbf{R}^T[\mathbf{t}]_\times\mathbf{R} = [\mathbf{R}^T\mathbf{t}]_\times$, that is, $\mathbf{H}^T\mathbf{E}$ is skew-symmetric for any vector $\mathbf{n}$ and scale of $\mathbf{t}$. Further, since the product is skew-symmetric we have that $\mathbf{H}^T\mathbf{E} + (\mathbf{H}^T\mathbf{E})^T = \mathbf{H}^T\mathbf{E} + \mathbf{E}^T\mathbf{H} = \mathbf{0}_{3\times3}$ and by definition $\mathbf{a}^T(\mathbf{H}^T\mathbf{E})\mathbf{a} = 0$ for any 3D vector $\mathbf{a}$. Recall that for a pair-wise corresponding features $(\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i})$ we have that $\boldsymbol{f}'_\mathrm{i} \sim \mathbf{H}\boldsymbol{f}_\mathrm{i}$ and $\boldsymbol{f}'^T_\mathrm{i}\mathbf{E}\boldsymbol{f}_\mathrm{i}$. Thus, introducing the expression for $\boldsymbol{f}'_\mathrm{i}$ into the epipolar constraints gives us $(\mathbf{H}\boldsymbol{f}_\mathrm{i})^T\mathbf{E}\boldsymbol{f}_\mathrm{i} = \boldsymbol{f}_\mathrm{i}^T\mathbf{H}^T\mathbf{E}\boldsymbol{f}_\mathrm{i}$ up-to-scale, and because $\mathbf{H}^T\mathbf{E}$ is skew-symmetric, the epipolar constraint is always zero [Hartley and Zisserman, 2003, Sec. 13.1.1].

2.2.2.3 Reducing the degrees of freedom

Additional information can be derived from other sensors and/or assumptions about the motion and scene spatial structure. In this section we focus on those cases in which the gravity vector or a constant direction of motion can be safely assumed before estimating the relative pose of the cameras. This information reduces the number of degrees of freedom (DoF) of the pose, and

for the above-mentioned examples, the gravity vector reduces the DoF of the rotation from three to one and the constant direction reduces the DoF of the translation from three (or two if it's planar) to one as well. Notice that, in some cases, these assumptions may not be exact, that is, we known the gravity vector or direction with a certain error. While describing the implications of these errors when these sub-problems are used instead of the full relative pose problem is an interesting topic, it is unfortunately out-of-scope of this thesis. Empirically, however, we observe that if the error on the assumption is low, these simplifications return good estimations. Next, we describe how this prior information is applied to the relative pose problem.

Depending on which information is introduced, the associated essential and/or homography matrix may inherit some special structure, *i.e.*, some entries are zero or the same (up to sign). As an example, the essential matrix whose associated axis of rotation is the vertical axis, *e.g.*, the gravity vector, is

$$\mathbf{E}_{\mathrm{y}} \doteq \begin{pmatrix} e_1 & e_2 & e_3 \\ e_4 & 0 & e_5 \\ -e_3 & e_6 & e_1 \end{pmatrix}, \tag{2.36}$$

for the rotation with form

$$\mathbf{R}_{\mathrm{y}} \doteq \begin{pmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{pmatrix}, \tag{2.37}$$

and the angle of rotation θ. If this structure is clear, as in the previous case with $\mathbf{E}_{\mathrm{y}}$, it is possible to modify the solvers that estimate these solutions to return a matrix fulfilling the structure. For example, if some entries of the matrix are zero, the associated rows (or columns) in the DLT approach can be removed as they don't contribute. For the above-mentioned example, the 9×9 coefficient matrix for generic DLT approach can be reduced to the 6×6 matrix

$$\mathbf{C}_{\mathrm{R}} \doteq [\mathbf{C}_{0,:} + \mathbf{C}_{8,:},\ |\mathbf{C}_{1,:},\ |\mathbf{C}_{3,:},\ |\mathbf{C}_{5,:},\ |\mathbf{C}_{6,:} - \mathbf{C}_{2,:},\ |\mathbf{C}_{7,:}] \in \mathbb{S}^6 \tag{2.38}$$

Note that, as the original DLT algorithm, the solution from the reduced matrix $\mathbf{C}_{\mathrm{R}}$ is not necessary associated to an essential matrix and the projection onto the appropriated space has to be performed, taking into account the prior information.

On the other hand, fewer variables alleviates the complexity of RANSAC algorithms [Fischler and Bolles, 1981] which are heavily used in both computer vision and robotics applications, whereas the prior information may also help to remove outliers that don't fulfill the known structure but fit into the general model, see Figure 2.5. Further, the prior information can be applied to the observations directly. For example, the essential matrix with general form $\mathbf{E} = [\mathbf{t}]_{\mathrm{x}}\mathbf{R}$ can be reduced by writing the rotation matrix as $\mathbf{R} = \mathbf{R}_{\mathrm{y}}\hat{\mathbf{R}}$, and thus,

we can write as well $\mathbf{E} = \mathbf{E}_{\mathrm{y}}\hat{\mathbf{R}}$. Then, the epipolar constraint for the i-th pair-wise observation with this form of the essential matrix has the form

$$\boldsymbol{f}_{\mathrm{i}}'^{T}\mathbf{E}\boldsymbol{f}_{\mathrm{i}} = \boldsymbol{f}_{\mathrm{i}}'^{T}([\mathbf{t}]_{\mathrm{x}}\mathbf{R}_{\mathrm{y}}\hat{\mathbf{R}})\boldsymbol{f}_{\mathrm{i}} = \boldsymbol{f}_{\mathrm{i}}'^{T}\mathbf{E}_{\mathrm{y}}(\hat{\mathbf{R}}\boldsymbol{f}_{\mathrm{i}}), \tag{2.39}$$

and so we can rotate the observations as $\hat{\mathbf{R}}\boldsymbol{f}_{\mathrm{i}}$ for $i = 1, \ldots, N$ so that the estimated essential matrix has the form $\mathbf{E}_{\mathrm{y}}$ in Equation (2.36). Similarly for the homography matrix as we can decompose the rotation so $\mathbf{H} = \mathbf{R} + \mathbf{t}\mathbf{n}^T = \left(\mathbf{R}_{\mathrm{y}} + \mathbf{t}\left(\hat{\mathbf{R}}\mathbf{n}\right)^T\right)\hat{\mathbf{R}} = \mathbf{H}_{\mathrm{y}}\hat{\mathbf{R}}$ and the homography constraint has the form $\boldsymbol{f}_{\mathrm{i}}' \sim \mathbf{H}_{\mathrm{y}}\hat{\mathbf{R}}\boldsymbol{f}_{\mathrm{i}}$. As with the epipolar constraint, the observations can be rotated by the known rotation matrix $\hat{\mathbf{R}}$ and the solver estimates the unknown homography $\mathbf{H}_{\mathrm{y}}$.

2.2.3 Triangulation problem

The triangulation problem estimates the 3D coordinates of the point $\mathbf{Q}$ that originated the N observations on N different images/cameras.

Considering a perspective model, we seek the point $\mathbf{Q} \in \mathbb{R}^3$ such that all the observations $\boldsymbol{f}_{\mathrm{i}} \in \mathbb{R}^3$ with associated projection matrix $\mathbf{P}_i \in \mathbb{R}^{3\times4}$ fulfill $\boldsymbol{f}_{\mathrm{i}} \sim \mathbf{P}_i\tilde{\mathbf{Q}}$ where $\sim$ indicates equality up-to-scale and $\tilde{\mathbf{Q}} \doteq [\mathbf{Q}^T, 1]^T \in \mathbb{R}^4$ is the homogeneous vector for $\mathbf{Q}$. Alternatively, we can express this relation as $\rho_i\boldsymbol{f}_{\mathrm{i}} = \mathbf{P}_i\tilde{\mathbf{Q}}$ with $=$ strict equality and $\rho_i \in \mathbb{R}_+$ the depth of the point if $\boldsymbol{f}_{\mathrm{i}}$ has last entry 1. For general cameras where the image plane does not exist, the observation has norm one instead and ρ_i is only related to the depth. The relation and the next development hold in both cases.

If the above-mentioned equality holds for all N, then the system of equations with unknowns $\mathbf{Q}$ and ρ_i for $i = 1, \ldots, N$ is formed as

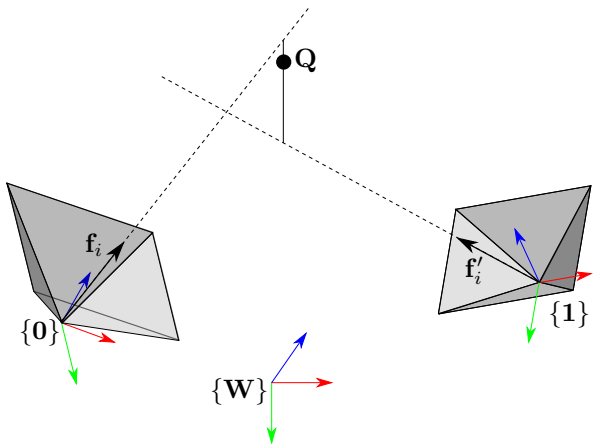
$$\rho_1\mathbf{f}_1 = \mathbf{P}_1\tilde{\mathbf{Q}} \tag{2.40}$$

$$\vdots$$

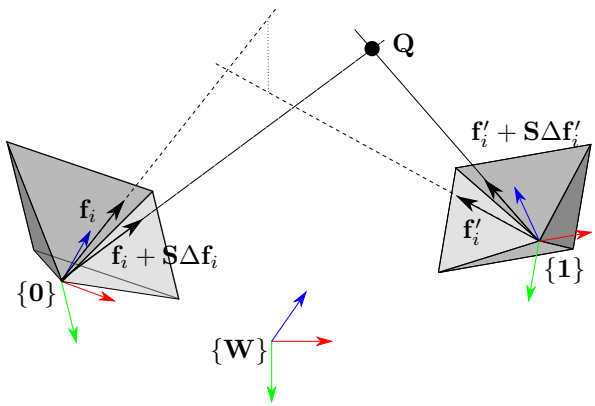
$$\rho_N\mathbf{f}_N = \mathbf{P}_N\tilde{\mathbf{Q}} \tag{2.41}$$

and has a unique solution provided the observations $\boldsymbol{f}_{\mathrm{i}}$ and projection matrices $\mathbf{P}_i$ are noiseless and the 3D point $\mathbf{Q}$ exists. Since the system is linear in the unknowns, we collect all the variables and form the linear system $\mathbf{A}\mathbf{X} = \mathbf{0}_{3N}$ as

$$\mathbf{A} = \begin{pmatrix} \mathbf{P}_1 & \mathbf{f}_1 & \mathbf{0}_{3\times1} & \cdots & \mathbf{0}_{3\times1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}_N & \mathbf{0}_{3\times1} & \mathbf{0}_{3\times1} & \cdots & \mathbf{f}_N \end{pmatrix} \in \mathbb{R}^{3N\times(4+N)}, \mathbf{X} = \begin{pmatrix} \tilde{\mathbf{Q}} \\ \rho_1 \\ \vdots \\ \rho_N \end{pmatrix} \in \mathbb{R}^{4+N} \tag{2.42}$$



(a) For noisy data the rays towards the observations do not meet in a single point in space. Midpoint methods find the middle point **Q** on the perpendicular to both rays.



(b) "Optimal" methods correct the observations to $\boldsymbol{f}_\mathrm{i} + \mathbf{S}\Delta\boldsymbol{f}_\mathrm{i}$ and $\boldsymbol{f}'_\mathrm{i} + \mathbf{S}\Delta\boldsymbol{f}'_\mathrm{i}$ so the rays do intersect in space at **Q**.

Figure 2.6: Different approaches to the triangulation problem: a shows the solution from the midpoint or line method for rays that do not intersect, whereas b shows the "optimal" method where the observations are modified.

For noiseless data $\boldsymbol{f}_\mathrm{i}$ and $\mathbf{P}_i$, the matrix $\mathbf{A}$ has a zero singularvalue and the associated singularvector is the solution $\mathbf{X}$. This procedure is known as the linear algorithm for triangulation see [Hartley and Zisserman, 2003, Sec. 12.2], and it is the one commonly employed. However, for noisy data the matrix $\mathbf{A}$ does not have a zero singular value. Still, we can estimate an approximate solution by considering the singularvector associated with the smallest singular value. Geometrically, this nonzero singular value implies that the rays that emanate from the camera centers towards the observations, see Figure 2.6a, do not intersect in space. However, this solution does not minimize any geometric error and an alternative approach for these cases is to take the middle point on the common perpendicular to all the rays [Yang et al., 2019]. This method has been shown to perform well when the projection matrices $\mathbf{P}_i$ are also noisy and when some weights are introduced into the observations, hence calculating a weighted, mean point [Lee and Civera, 2019b]. However, the "optimal" approach to triangulation as introduced in [Hartley and Zisserman, 2003, Sec. 12.3] is to correct the observations $\boldsymbol{f}_\mathrm{i}$ so that the 3D point is recovered exactly, that is, the rays do intersect in the space and the matrix $\mathbf{A}$ has a zero singularvalue, see Figure 2.6b. Among all the solutions, provided they exist, we seek those with minimum correction $\Delta\boldsymbol{f}_\mathrm{i} \in \mathbb{R}^2$ that still guarantee the existence of the 3D point. Depending on which norm the correction is measured, *e.g.*, $\ell_2, \ell_1, \ell_\infty$, we may obtain different solutions. Formally, the problem is stated as

$$\min_{\Delta\boldsymbol{f}_\mathrm{i},\dots,\mathbf{Q}} \sum_{i=1}^{N} \|\Delta\boldsymbol{f}_\mathrm{i}\|^2 \,, \text{ subject to } \mathbf{S}^T\boldsymbol{f}_\mathrm{i} + \Delta\boldsymbol{f}_\mathrm{i} = \Pi(\mathbf{P}_i\tilde{\mathbf{Q}}) \tag{2.43}$$

where $\Pi(a,b,c) = [a/c, b/c]^T \in \mathbb{R}^2$ is the perspective function, $\|\bullet\|$ is some norm for the argument and $\mathbf{S} = [\boldsymbol{I}_2 \mid \mathbf{0}_2] \in \mathbb{R}^{3\times 2}$. While this is the general approach (for generic scenes) we must highlight that if the points are known to belong to a plane, this general approach does not suffice as the triangulated points don't necessary lay on the plane after the correction of the observations. While the coplanarity of the points can be achieved after the triangulation of the points, *e.g.*, [Poullis and You, 2011, Micusik and Wildenauer, 2017], it's recommended to enforce it while triangulating. In this case, though, the relation between the 3D point $\mathbf{Q}$ on the world plane π and the observation $\boldsymbol{f}_\mathrm{i}$ on the i-th image plane can be expressed through the (planar) homography matrix $\mathbf{H}_i \in \mathbb{R}^3$ [Hartley and Zisserman, 2003, Sec. 13] as $\boldsymbol{f}_\mathrm{i} = \mathbf{H}_i\mathbf{Q}$. This holds for all the image planes, so we have the set of N homography matrices $\mathbf{H}_1, \dots, \mathbf{H}_N$. In this case, the triangulation problem has the specific form

$$\min_{\Delta\boldsymbol{f}_\mathrm{i},\dots,\mathbf{Q}} \sum_{i=1}^{N} \|\Delta\boldsymbol{f}_\mathrm{i}\|^2 \,, \text{ subject to } \mathbf{S}^T\boldsymbol{f}_\mathrm{i} + \Delta\boldsymbol{f}_\mathrm{i} = \Pi(\mathbf{H}_i\mathbf{Q}). \tag{2.44}$$

Notice that for both cases, the 3D point $\mathbf{Q}$ is involved in the optimization. In practice, we would like to avoid these additional variables and rather

work only with the observations and the information about the pose through the projection matrix $\mathbf{P}_i$ and/or homography $\mathbf{H}$. For both cases, the relations that only involve the observations have been already introduced in previous sections, through the essential and homography matrices and their respective constraints. These constraints are bilinear in the observations, which will simplify the resolution of the problem in practice. In Chapter 6 we elaborate further on these approaches.

2.2.4 Resectioning problem

The resectioning or Perspective-n-Point (PnP) problem aims to estimate the absolute pose $(\mathbf{R}, \mathbf{t})$ between the world frame $\{\mathbf{W}\}$ and the camera frame $\{\mathbf{C}\}$ given a set of N observation-point pairs $(\boldsymbol{f}_\mathrm{i}, \mathbf{P}_\mathrm{i})$. For noncentral cameras the problem is stated similarly, but considering the known poses $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ between the M cameras. Since the central configuration can be seen as a specific case of the noncentral with $M = 1$, we provide the explanation for the former. Notice that for the central case, we have that $\mathbf{R}_\mathrm{j} = \boldsymbol{I}_3$ and $\mathbf{c}_\mathrm{j} = \mathbf{0}_{3\times 1}$ which in turn simplifies the formulation as we will indicate later on.

We consider the case where the i-th observation $\boldsymbol{f}_\mathrm{i}$ is seen on the j-th camera with known pose $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ *w.r.t.* the camera center $\{\mathbf{C}\}$, see Figure 2.7. We transform the coordinates of the associated 3D point $\mathbf{P}_\mathrm{i}$ on the world frame to the camera frame as $\mathbf{Q}_\mathrm{i} \doteq \mathbf{R}_\mathrm{j}^T(\mathbf{R}\mathbf{P}_\mathrm{i} + \mathbf{t} - \mathbf{c}_\mathrm{j}) \in \mathbb{R}^3$ where the unknown pose is $(\mathbf{R}, \mathbf{t})$. For a perspective camera model, the associated observation on the image plane to this point is $\mathbf{q}_\mathrm{i} \doteq \mathbf{Q}_\mathrm{i}/\mathbf{Q}_{\mathrm{i}3}$ and $\mathbf{q}_\mathrm{i} \doteq \mathbf{Q}_\mathrm{i}/\|\mathbf{Q}_\mathrm{i}\|_2 \in \mathbb{S}^2$ on the image sphere. For noiseless data, $\mathbf{q}_\mathrm{i}$ equals the (known) observation $\boldsymbol{f}_\mathrm{i}$ and the error $\epsilon_\mathrm{i} \doteq \mathrm{dist}(\boldsymbol{f}_\mathrm{i}, \mathbf{q}_\mathrm{i})$ is zero independently of the function used to measure distance. Otherwise, the error is not zero and the squared term ϵ_i^2 is typically used as cost to be minimized for the resectioning problem. Note that if the camera has an actual image plane, we can measure this distance by the ℓ_1, ℓ_2 or ℓ_∞ norms, including the weighted versions. If we need to work with the image sphere instead (observations with unit norm), the norms have to be adapted as we indicate in Section 2.1. In both cases, the problem is formally defined in its general form as

$$f^\star = \min_{\mathbf{R}\in\mathbb{R}^{3\times 3}, \mathbf{t}\in\mathbb{R}^3} \sum_{i=1}^{N} \mathrm{dist}(\boldsymbol{f}_\mathrm{i}, \mathbf{q}_\mathrm{i})^2 \text{ subject to } \mathbf{R} \in \mathbb{SO}(3) \tag{2.45}$$

However, the reprojection error $\epsilon_\mathrm{i} \doteq \mathrm{dist}(\boldsymbol{f}_\mathrm{i}, \mathbf{q}_\mathrm{i})$ is rational in the unknowns $(\mathbf{R}, \mathbf{t})$. An alternative error can be devised for this problem by considering the observation $\boldsymbol{f}_\mathrm{i}$ as the ray (line) emanating from the camera center towards the observation, see Figure 2.8. In this case, we can measure the error as the minimum distance between this ray and the transformed point $\mathbf{Q}_\mathrm{i}$. The minimum distance is obtained on the perpendicular to the observation, formally $\epsilon_\mathrm{i} \doteq (\boldsymbol{I}_3 - \boldsymbol{f}_\mathrm{i}\boldsymbol{f}_\mathrm{i}^T)\mathbf{Q}_\mathrm{i}$ which is linear in the unknowns. This metric is commonly

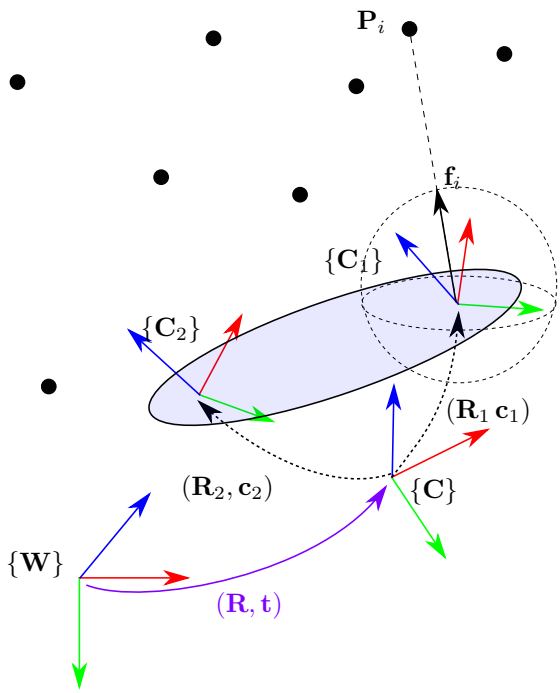


Figure 2.7: The resectioning problem estimates the pose $(\mathbf{R}, \mathbf{t})$ between the world frame $\{\mathbf{W}\}$ and the center of the generalized camera $\{\mathbf{C}\}$ given N point-observation pairs $\mathbf{Q}_\mathrm{i}, \boldsymbol{f}_\mathrm{i}$. For noncentral cameras, the absolute pose $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ between the M cameras and the center $\{\mathbf{C}\}$ are also known.

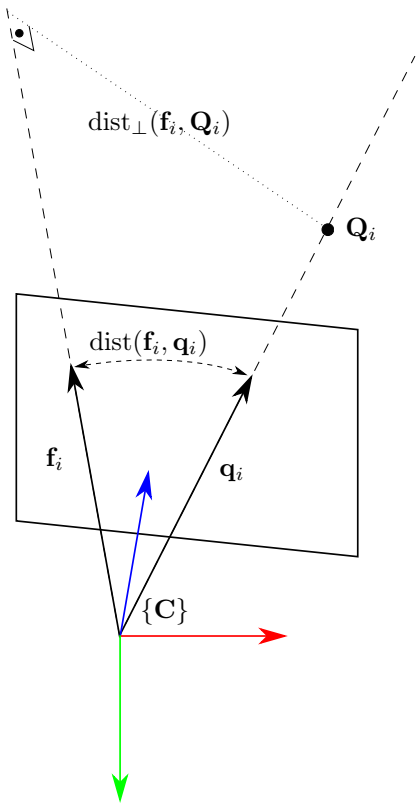


Figure 2.8: The reprojection error $\text{dist}(\boldsymbol{f}_\text{i}, \mathbf{q}_\text{i})$ measures distance between observations and the point-line error $\text{dist}_\perp(\boldsymbol{f}_\text{i}, \mathbf{Q}_\text{i})$ measures the perpendicular distance to the transformed 3D point $\mathbf{Q}_\text{i}$.

known as the point-line distance and is used as cost function in the registration of points and lines. Formally we define the problem as

$$f^\star = \min_{\mathbf{R}\in\mathbb{SO}(3),\mathbf{t}\in\mathbb{R}^3} \sum_{i=1}^{N} \left\| (\boldsymbol{I}_3 - \boldsymbol{f}_\mathrm{i}\boldsymbol{f}_\mathrm{i}^T)\mathbf{Q}_\mathrm{i} \right\|_2^2 \tag{2.46}$$

Notice that this approach states the PnP problem as a specific instance of registration, a well-known field with a vast literature.

2.2.5 Solution estimation for nonlinear problems: on-manifold approach

Iterative algorithms that estimate solutions to nonconvex problems can be used jointly with certifiable algorithms. Whereas it is possible to employ any iterative, nonminimal solver for this task, in computer vision and robotics we often find problems defined on rotations and spheres, and so the variables of the problem must be valid rotations and unit-norm vectors. In order to use standard algorithms, the variables should lie on the Euclidean space, and a possible approach to this is to reformulate the problem in terms of minimal parameterization of the variables to leverage these solvers. However, these transformations often suffer from degenerate configurations [Cayley, 1846], for example, Gimbal lock appears when 3D rotations are parameterized with Euler angles, and special care must be paid to avoid these situations.

On the other hand, it is possible to maintain the original formulation while fulfilling the internal constraints of the variables, although some extra steps are required. This approach optimizes the cost function directly on the associated non-Euclidean spaces, and if those spaces are smooth and are endowed with a (Riemannian) metric, we say the optimization takes place on a (Riemannian) manifold. This is the approach followed by all our work, as it allows us to reduce the computational effort required to derive the first and second-order information about the problem while avoiding the degenerate configurations. Further information about this can be found in the books [Absil et al., 2009, Nocedal and Wright, 1999, Boumal, 2023a], and we give here only the basic details to understand the method.

The core idea is to separate the problem resolution into three parts: (1) the solver; (2) the problem model; and (3) the domain. For the **solver**, as with Euclidean spaces, we may opt for first or second-order optimization algorithms, and the amount of information about the problem, that is provided by the second block (the problem model), depends on which option we choose. It's important to remind here that the solver does not depend on the problem formulation nor the variables.

The second block requires the definition of the **problem model** as, at least, the cost function and gradient. If the solver leverages second-order information, the Hessian or the Hessian-vector product should be also provided. These

expressions also depend on the form of the variable we use, for instance, for a problem seeking a 3D rotation, the definition of the cost will be different if we use the full 3×3 matrix $\mathbf{R}$ or its 4-D quaternion version $\mathbf{q}$. For example, for a problem with quadratic cost on the 3×3 matrix $\mathbf{R}$ defined as

$$f^\star = \min_{\mathbf{R} \in \mathbb{SO}(3)} \operatorname{vec}(\mathbf{R})^T \mathbf{C} \operatorname{vec}(\mathbf{R}) + 2\mathbf{c}^T \operatorname{vec}(\mathbf{R}) + c \tag{2.47}$$

with $\operatorname{vec}(\mathbf{R}) \in \mathbb{R}^9$ is the column-wise vectorization of $\mathbf{R}$, the cost provided to the tool is directly the expression $f(\mathbf{R}) = \operatorname{vec}(\mathbf{R})^T \mathbf{C} \operatorname{vec}(\mathbf{R}) + 2\mathbf{c}^T \operatorname{vec}(\mathbf{R}) + c \in \mathbb{R}$. The gradient (or Jacobian) on $\mathbf{R}$ is $\nabla f(\mathbf{R}) = 2\mathbf{C} \operatorname{vec}(\mathbf{R}) + 2\mathbf{c} \in \mathbb{R}^{9\times 1}$. In addition, the Hessian is $\nabla^2 f(\mathbf{R}) = 2\mathbf{C} \in \mathbb{R}^{9\times 9}$ and the Hessian-vector product with $\mathbf{U} \in \mathbb{R}^{3\times 3}$ is $\nabla^2 f(\mathbf{R})[\mathbf{U}] = 2\mathbf{C} \operatorname{vec}(\mathbf{U}) \in \mathbb{R}^{9\times 1}$. Note that a reshape of the 9D vectors $\nabla f(\mathbf{R})$ and $\nabla^2 f(\mathbf{R})^2[\mathbf{U}]$ into 3×3 matrices may be required.

To maintain the variables on their respective spaces, this approach also requires information about the **domain**, following the example above, about the 3D rotation matrices. As with the cost, this depends on which form of the variable we are working with and for the example with the rotation, we know that the internal restrictions of 3×3 rotation matrices and quaternions (4D vectors) are different. The information about the domain is provided through the *operators* and we mostly require four of them: the retraction; the tangent space projector, Euclidean gradient to Riemannian gradient; and Euclidean Hessian to Riemannian Hessian. The full description of these operators for the most common manifolds can be found in the above-mentioned references. It's necessary to add that different variables of the same problem can belong to different domains yet the approach is the same. For example, we can tackle the relative pose problem based on the essential matrix through its simpler form where we seek a 3D rotation and the translation is a 3D vector with unit norm. In this case, the domain is the direct product of 3D rotations $\mathbb{SO}(3)$ and the 2-sphere $\mathbb{S}^2$, which we write as $\mathbb{SO}(3) \times \mathbb{S}^2$. This allows us to express independently the contributions of the variables and so the gradient of the full problem has a component for the translation (that may involve the rotation) and a component for the rotation, and similarly for the Hessian. Further, we just need to include the associated operators for each domain, even though the problem involves all the variables. Powers of these domains can be also considered, for example, when solving the so-called rotation averaging problem when N rotations are involved, which makes $\mathbb{SO}(3)^N$ the domain.

As a last note, we must highlight two tools that have certainly helped us during this thesis: MANOPT `manopt.org` at `https://github.com/NicolasBoumal/manopt`; and the *Optimization* library `https://github.com/david-m-rosen/Optimization`. Both tools implement the above-mentioned process of separating each component of the optimization and the associated codes provide a great opportunity for learning. MANOPT provides first and second-order solvers, with the most common domains (among them, rotation, sphere and Euclidean

space) and all the associated operators for the manifolds, hence removing the need to implement those from scratch. Products and powers of these domains can be also used without any further implementation and gradient and Hessian verification procedures are available to check the numerical correctness of the problem model. We have tested that all these additional tools simplify the modeling and verification stages of the problem formulation. The implementation, however, is only available for MATLAB and PYTHON. On the other hand, the *Optimization* library provides the framework for this approach suitable for C++, although the domain operators need to be implemented. First and second-order solvers are also available, and several examples show how to implement the target problem with the library.

2.3 Convex optimization foundation

> In fact the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity
>
> *Rockafellar, 1993 [Rockafellar, 1993]*

In this section, we explain the main concepts of convex optimization that are exploited during this thesis. We refer interested readers to the books [Boyd and Vandenberghe, 2004, Nocedal and Wright, 1999, Luenberger et al., 1984] for further information.

There are several reasons behind our interest in this family of problems. First, we can use the solution from a convexification of the problem as initial guess for iterative algorithm, *e.g.*, Section 2.2.5. We can also leverage convex heuristic, for instance, randomized algorithms [Motwani and Raghavan, 1996] to tackle the nonconvex problems. Last, and our main focus, is the estimation of lower bounds on the optimal cost of the nonconvex problems from these convex relaxations. The remaining of this section focuses on these convex relaxations and their relations with the original, nonconvex problem.

2.3.1 Basic concepts

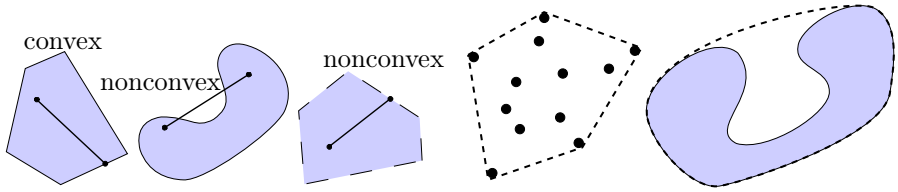
Before tackling these (convex) relaxations, we must introduce some necessary concepts about convex optimization. We follow the distribution in [Boyd and Vandenberghe, 2004] and adapt the content to the problems solved in this thesis.

2.3.1.1 Convex sets

We say that the set C is convex *iff* the line segment between two points $x_1, x_2 \in C$ lies also in C, that is, for the parameter θ with $0 \leq \theta \leq 1$ we have that $\theta x_1 + (1-\theta)x_2 \in C$. The convex hull is the intersection of all convex sets that contain the original set, that is, the smallest convex set containing all the points in the original set. For a set of N points p_i for $i = [N]$, the convex hull is defined as

$$\mathbf{Co}(p_1, \ldots, p_N) \doteq \left\{ \sum_{i=1}^{N} a_i p_i,\ a_i \in \mathbb{R}_+,\ \sum_{i=1}^{N} a_i = 1 \right\}. \tag{2.48}$$

Examples of a convex and a nonconvex set in 2D are given in Figure 2.9a, whereas the convex hulls for two nonconvex sets are shown in Figure 2.9b.



(a) Examples of convex and nonconvex sets in 2D.

(b) Dashed-line delimits the convex hull for a set of isolated points (left) and a generic, nonconvex set (right).

Figure 2.9: Examples of convex and nonconvex sets where shaded areas indicated points belonging to the set.

The set C is a cone if for every $x \in C$ and non-negative scalar $\theta \geq 0$, we have that $\theta x \in C$, [Boyd and Vandenberghe, 2004, Sec. 2.1.5]. A convex cone is a set that is convex and a cone, and so formally for $x_1, x_2 \in C$ an $\theta_1, \theta_2 \geq 0$ we have $\theta_1 x_1 + \theta_2 x_2 \in C$. Equivalently, the set C is a convex cone *iff* it contains all conic combinations $\theta_1 x_1 + \cdots + \theta_k x_k \in C$ for $x_i \in C$ and $\theta_i \geq 0$ for $i = [k]$. The conic hull of a set C is the set of all conic combinations $\{\theta_1 x_1 + \cdots + \theta_k x_k \mid x_i \in C, \theta_i \in \mathbb{R}_+, i = [k]\}$ and it is the smallest convex cone that contains C.

An important space related to cones is the so-called *dual cone*, and is formally defined as

$$K^* = \{y \mid \langle x, y \rangle \geq 0 \text{ for all } x \in K\} \tag{2.49}$$

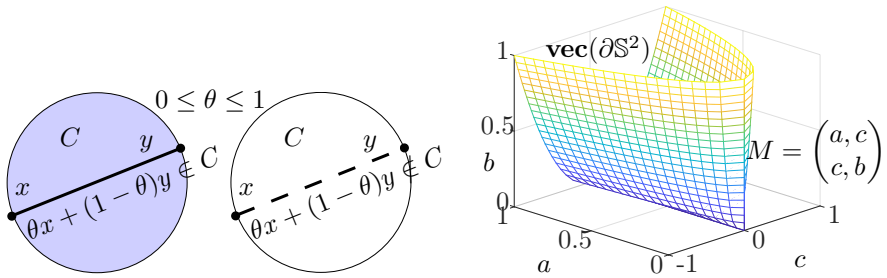
being K the associated cone and $\langle \bullet, \bullet \rangle$ the inner product associated to the cone K. The dual cone K^* is always a convex cone even with K is neither convex nor a cone. Further, we say that the cone C is proper *iff* is: (1) convex; (2) closed; (3) solid, that is, it has nonempty interior; and (4) pointed, that is, it does not contain any line. The proper cone C defines a *generalized inequality* that is a partial ordering on $\mathbb{R}^n$, and $x \succeq_C y \Leftrightarrow x - y \in C$.

Examples of convex sets are: (1) the empty set $\emptyset$; (2) a single point $\{a\}$; (3) all the reals $\mathbb{R}^n$; the half-spaces given by points $\mathbf{x} \in \mathbb{R}^n$ fulfilling $\mathbf{a}^T\mathbf{x} \geq (>) b$ for $\mathbf{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, which includes lines and in general subspaces (affine) as $\mathbf{a}^T\mathbf{x} = b$. Further, the dual cone for the set formed by $\{0\}$ are all the reals, while the dual cone for all the reals is the zero element and the dual cone of a subspace V is its orthogonal complement with $V^\perp = \left\{y \mid v^T y = 0, \text{for all } v \in V\right\}$.

Ball and sphere:

Other examples of convex cones include the Euclidean ball in $\mathbb{R}^n$ with radius $r \geq 0$ and centered at the point $\mathbf{x}_c \in \mathbb{R}^n$, formally defined as

$$\mathrm{B}(\mathbf{x}_c, r) = \left\{\mathbf{x} \mid \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r\right\} = \left\{\mathbf{x} \mid (\mathbf{x} - \mathbf{x}_c)^T(\mathbf{x} - \mathbf{x}_c) \leq r^2\right\}. \tag{2.50}$$



(a) The ball (left) is convex whereas the sphere (right) that only considers the points on the boundary is not

(b) The boundary of the PSD cone $\partial\mathbb{S}^2_+$ for 2×2 matrices have rank one and zero, and the interior is formed by rank-two matrices.

Figure 2.10: Common sets encountered in this thesis: the SDP cone 2.10b, and the ball and sphere 2.10a.

The ball is convex, so is the family of (convex) sets called ellipsoids for $\mathbf{P} = \mathbf{P}^T \succ 0$ defined by

$$\mathcal{E}(\mathbf{x}_c, \mathbf{P}) = \left\{\mathbf{x} \mid (\mathbf{x}-\mathbf{x}_c)^T\mathbf{P}^{-1}(\mathbf{x}-\mathbf{x}_c) \leq 1\right\}. \tag{2.51}$$

If $\mathbf{P} \succeq 0$, the ellipsoid is called degenerate yet it remains convex. The concept of ball is extended to any norm while remaining convex, and it gives rise to the concept of *norm cone* with the form $C = \{(x,t) \mid |x| \leq t\}$ for the norm $|x|$. Notice, though, that while the ball is convex, the sphere, *i.e.*, the boundary of the set defined by $\{\mathbf{x} \mid (\mathbf{x}-\mathbf{x}_c)^T(\mathbf{x}-\mathbf{x}_c) = r^2\}$ is not, see Figure 2.10a. The unit sphere $\mathbb{S}^n$ is the sphere of radius one and specifically, the 2-sphere (3D vectors with unit-norm) is intensively used in computer vision and robotics, see *e.g.* Section 2.2.2.

PSD and PD cone:
Symmetric matrices of size $n \times n$ belong to the set $\mathbb{S}^n$ formally defined as

$$\mathbb{S}^n \doteq \left\{\mathbf{X} \in \mathbb{R}^{n\times n} \mid \mathbf{X} = \mathbf{X}^T\right\} \tag{2.52}$$

which is a vector space with dimension $n(n+1)/2$. All symmetric matrices whose eigenvalues are non-negative formed the set of positive semidefinite (PSD) matrices $\mathbb{S}^n_+$,

$$\mathbb{S}^n_+ \doteq \left\{\mathbf{X} \in \mathbb{S}^{n\times n} \mid \mathbf{X} \succeq 0\right\} \tag{2.53}$$

and for any vector $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{S}^n_+$ we know that $\mathbf{x}^T\mathbf{X}\mathbf{x} \geq 0$. Further, $\mathbb{S}^n_+$ is a convex cone and thus $\alpha_1, \alpha_2 \in \mathbb{R}_+$ and $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{S}^n_+$ we know that $\alpha_1\mathbf{X}_1 + \alpha_2\mathbf{X}_2 \succeq 0$. Symmetric matrices whose eigenvalues are strictly positive formed the set of positive definite (PD) matrices $\mathbb{S}^n_{++}$

$$\mathbb{S}^n_{++} \doteq \left\{\mathbf{X} \in \mathbb{S}^{n\times n} \mid \mathbf{X} \succ 0\right\} \tag{2.54}$$

and for any non-null vector $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{X} \in \mathbb{S}^n_{++}$ we know that $\mathbf{x}^T\mathbf{X}\mathbf{x} > 0$. Matrices with negative, zero and/or positive eigenvalues are called *indefinite*. Notice that we use the notation $\succeq$ ($\succ$) to define PSD (PD) matrices, which is a simplification of the notation for proper cones. Through the thesis we will keep this notation as PSD/PD matrices are prevalent without specifying the PSD/PD space.

Operations preserving convexity and convexification:

Some operations preserve convexity, among them: (1) intersection; (2) affine functions; and (3) perspective and linear-fractional (composition of perspective and affine functions). This means that the intersection of affine equations and inequalities remain convex, and therefore the intersection of M linear expressions on $\mathbf{X} \succeq 0$ for symmetric matrices $\mathbf{A}_i$ and scalars b_i defined as $\text{tr}(\mathbf{A}_1\mathbf{X}) = b_i$ for $i = [M]$ is convex.

Nonetheless, nonconvex sets appear in this thesis, for instance, problems defined on the sphere $\mathbb{S}^2$ appear during the relative pose problem between two calibrated cameras. In this case, one way to convexify the domain is to extend domain from sphere to the ball, that is, we allow points with norm below one. Other problems seek variables with rank one, which is the boundary of the PSD cone, see Figure 2.10b. To convexify the domain in this particular instance we can discard the rank constraint, hence allowing the matrix $\mathbf{X}$ to have other ranks, that is, the interior of the PSD cone. In both examples the final domain is larger than the original one as we are considering more points and so the final domain is a *relaxation* of the original one, so it is the derived problem with respect to the original one.

2.3.1.2 Convex functions

We say that the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex *iff* its domain $\mathbf{dom}\, f$ is a convex set and for all $x_1, x_2 \in \mathbf{dom}\, f$ and scalar $0 \leq \theta \leq 1$

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta) f(x_2). \tag{2.55}$$

Figure 2.11 shows a simple representation of convex and nonconvex functions, where the latter may present, in general, many local minima. The function f is *strictly convex* if strict inequality holds in Equation (2.55) for $x_1 \neq x_2$ and $0 < \theta < 1$. Similarly, f is (strictly) concave if $-f$ is (strictly) convex. In addition, the function f is convex *iff* it remains convex when restricted to *any* line intersecting its domain and the convex function is continuous of the relative interior of the domain, but may have discontinuities on its relative boundary. For affine functions, equality holds in Equation (2.55) and are therefore convex and concave at the same time, while the converse holds.

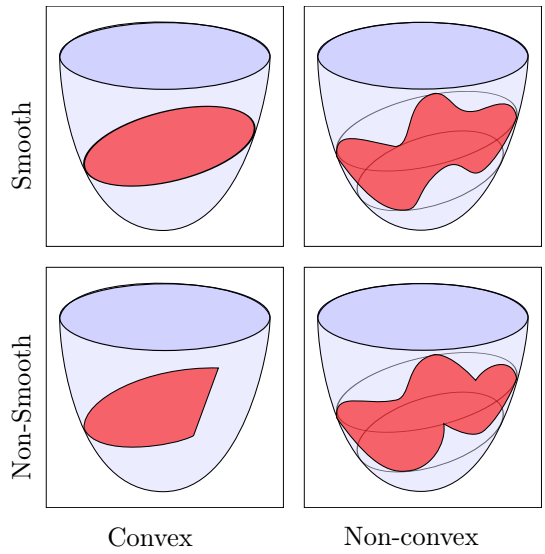


Figure 2.11: Examples of convex, nonconvex, smooth and nonsmooth functions in 3D.

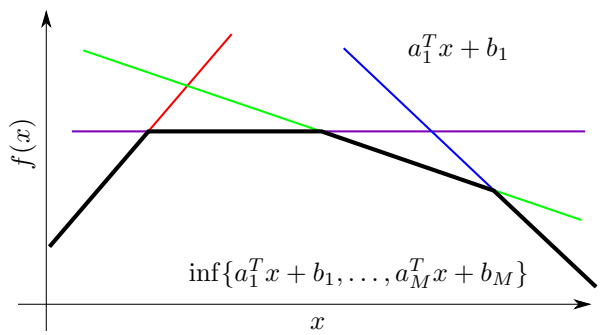
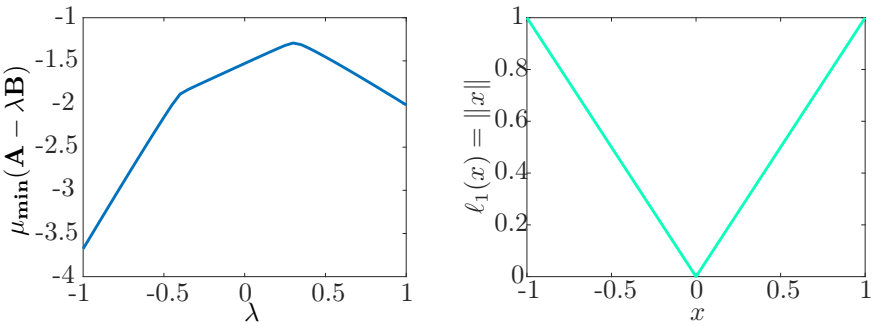


Figure 2.12: The pointwise infimum $\inf\{f_1(x), \ldots, f_M(x)\}$ of affine functions $f_j(x) \doteq \mathbf{a}_j^T \mathbf{x} + b_j$ is convex.



(a) Minimum eigenvalue of the linear matrix pencil $\mathbf{A} - \lambda\mathbf{B}$ for $\lambda \in [-1, 1]$.

(b) The ℓ_1 has a singularity at $x = 0$, where the minimum is.

Figure 2.13: Examples of nonsmooth but convex functions.

Convexity can also be defined by the first-order condition that states that if f is (first-order) differentiable with gradient $\nabla f(x)$, then f is convex *iff* its domain is convex and

$$f(x_2) \geq f(x_1) + \nabla f(x_1)^T (x_2 - x_1) \tag{2.56}$$

for any $x_1, x_2 \in \mathbf{dom}\, f$. Similarly, strict convexity implies strict inequality for $x_2 \neq x_1$. The right-hand side of the Equation (2.56) is the first-order Taylor approximation of x_2 at the point x_1, which is a global underestimator for a convex function f, and the converse holds as if the Taylor approximation is a global underestimator of f, then the function is convex. Notice that if $\nabla f(x_1) = 0$ then x_1 is a global minimizer of f, as for *any* $x_2 \in \mathbf{dom}\, f$ we have that $f(x_2) \geq f(x_1)$. The proof of this statement can be found in [Boyd and Vandenberghe, 2004, Sec. 3.1.3].

Notice, though, that a function can be nonsmooth while being convex or nonconvex, see Figure 2.11. If the function is nonsmooth at some point $x_1 \in \mathbf{dom}\, f$, then the gradient $\nabla f(x_1)$ is not uniquely defined. The *subgradient* generalizes the derivative to non-differentiable functions and the *sudifferential* $\partial f(x)$ is the closed, convex set of all subgradients at the point $x \in \mathbf{dom}\, f$, even when f is nonconvex [Boyd and Vandenberghe, 2006]. We say f is *subdifferentiable* at a point x if there exists are least one subgradient at the point, and if this holds for all $x \in \mathbf{dom}\, f$, then f is subdifferentiable without specification. Further, if the function f is convex and differentiable, then its gradient coincides with its subgradient. As with differentiable function, we say that x_1 is a minimizer of the convex function f *iff* f is subdifferentiable an $\mathbf{0} \in \partial f(x_1)$. A well-known example for this case is the ℓ_1 function $f = |x|$, and more important for us, the (sum of the) smallest eigenvalues of the matrix $\mathbf{A}_0 + \sum_{i=1}^{M} \lambda_i \mathbf{A}_i$ with respect to λ_i, as shown in Figure 2.13a.

If the function f is twice differentiable, then the function is convex *iff* the domain $\mathbf{dom}\, f$ is convex and its Hessian is PSD for *all* $x \in \mathbf{dom}\, f$, that is, $\nabla^2 f(x) \succeq 0$, and it is concave for $\nabla^2 f(x) \preceq 0$. While $\nabla^2 f(x) \succ 0$ for all $x \in \mathbf{dom}\, f$ implies f is strictly convex, the opposite is not true. For quadratic functions with the form $f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{a}^T\mathbf{x} + a$, for $\mathbf{dom}\, f = \mathbb{R}^n$ we have that $\nabla^2 f(\mathbf{x}) = \mathbf{A}$, and so f is (strictly) convex *iff* $\mathbf{A} \succeq 0$ ($\mathbf{A} \succ 0$). Notice that the convexity of the set is *required* for the convexity of the function, see [Boyd and Vandenberghe, 2004, Remark 3.1]. This is of special importance for this thesis, as we usually have functions with PSD coefficient matrices, but nonconvex domains, for instance, rotation matrices.

Convexity for functions over the PSD cone The above-mentioned conditions are extended to proper cones as introduced in Section 2.3.1.1. As one of the most prevalent domains of this thesis are PSD matrices, we use this set in the following statements. Thus, for the proper cone $\mathbb{S}^n_+$ with associated generalized inequality $\succeq$ we say that the function $f : \mathbb{S}^n_+ \to \mathbb{R}$ is *PSD-nondecreasing iff* $\mathbf{X} \preceq \mathbf{Y} \Rightarrow f(\mathbf{X}) \leq f(\mathbf{Y})$, and it is *PSD-increasing* if strict inequality holds for $\mathbf{X} \neq \mathbf{Y}$. For the PSD cone, these functions are also called *matrix monotone increasing/decreasing*, and among them, we highlight the trace of the product with another symmetric matrix $\mathrm{tr}(\mathbf{XY})$, the trace of the inverse $\mathrm{tr}(\mathbf{X}^{-1})$ and the determinant $\det(\mathbf{X})$, see [Boyd and Vandenberghe, 2004, Ex. 3.46]. The monotonicity can be also expressed through the gradient of the function as this function is PSD-nondecreasing if its domain is convex and $\nabla f(x) \succeq 0$ for all $x \in \mathbf{dom}\, f$, but the converse is not true. Similar notions are derived for convexity, as the function is PSD-convex if for all $x, y \in C$ and $0 \leq \theta \leq 1$ we have that

$$f(\theta x + (1-\theta)y) \preceq \theta f(x) + (1-\theta) f(y), \tag{2.57}$$

and strict convexity holds for strict inequality for $x \neq y$ and $0 < \theta < 1$. First-order condition for convexity also holds, as f is convex if its domain is convex and for all $x, y \in \mathbf{dom}\, f$ we have

$$f(y) \succeq f(x) + Df(x)(y - x) \tag{2.58}$$

where $Df(x)$ is the Jacobian of f, and we say f is strictly convex if inequality holds in Equation (2.58) for $x \neq y$.

Operations preserving convexity:

As with the convex set, some operations preserve convexity, among them: (1) non-negative weighted sum; (2) composition with affine mapping; (3) point-wise maximum (and minimum); (4) composition. For instance, the problem on the reals $\boldsymbol{\lambda} \in \mathbb{R}^N$ that seeks the point-wise infimum of the functions $\mathbf{x}^T(\mathbf{A}_0 + \sum_{i=1}^N \lambda_i \mathbf{A}_i)\mathbf{x}$ for a free variable $\mathbf{x}$ is convex since the functions are affine in $\boldsymbol{\lambda}$, see Figure 2.12.

2.3.2 Problem formulation

We consider the problem on the *optimization variable* $\boldsymbol{\phi} \in \mathbb{R}^K$, that minimizes the *objective or cost function* $h(\boldsymbol{\phi})$ such that the variable fulfills the *equality constraints* $p_j(\boldsymbol{\phi}) = 0$ for $j = [L]$. In what follows we consider only the minimization of the cost $h(\boldsymbol{\phi})$ as maximizing a cost $h(\boldsymbol{\phi})$ can be written as the minimization of $-h(\boldsymbol{\phi})$. Further, we only consider equality constraints as they are the only appearing in this thesis, although the theory also applies for inequalities constraints [Boyd and Vandenberghe, 2004].

The *domain* of the problem is the set of points for which the objective and constraints are defined at the same time, whereas the *feasible set* is the set formed by all the points in the domain that fulfill all the constraints. We say the problem is *feasible* if this set is nonempty and *infeasible* otherwise. Formally, we write the problem as

$$h^\star = \inf_{\boldsymbol{\phi} \in \mathbb{R}^K} h(\boldsymbol{\phi}), \qquad \text{subject to } p_j(\boldsymbol{\phi}) = 0,\ j = [L] \tag{OG}$$

For problems whose optimal cost is not known to be finite we follow the notation 'inf' and 'sup' for minimization and maximization, respectively. On the other hand, if the cost is known to be finite we write the problem with 'min' and 'max', respectively. In any case, the *optimal cost* $h^\star$ is the lowest value that takes the objective of the problem, and it can be $+\infty$ when the problem is infeasible, $-\infty$ when the problem is unbounded below or a finite value attained or achieved by an *optimal solution* $\mathbf{x}^\star$, which is necessary feasible and so $f^\star = f(\mathbf{x}^\star)$. For this last case, we say the problem is *solvable* and the set of optimal solutions is called the *optimal set*.

We must mention a family of problems known as *feasibility problems*, where the goal is not to minimize the objective function but to find a feasible point. In this case, the optimal value is zero if such point exists and ∞ otherwise. Formally, this problem is written as

$$\text{find } \boldsymbol{\phi}, \qquad \text{subject to } p_j(\boldsymbol{\phi}) = 0,\ j = [L] \tag{2.59}$$

Standard form of the problem and QCQP:

Our goal is to re-formulate the original problem in Equation (OG) to have quadratic functions both in the cost and the constraints. We transform the original cost in $\boldsymbol{\phi} \in \mathbb{R}^K$ to the general quadratic function in the variable $\mathbf{z} \in \mathbb{R}^N$ as $h(\boldsymbol{\phi}) = f(\mathbf{z}) = \mathbf{z}^T \tilde{\mathbf{C}}_\mathrm{i} \mathbf{z} + 2\tilde{\mathbf{d}}_\mathrm{i}^{\,T} \mathbf{z} + r_i$ with $\tilde{\mathbf{C}}_\mathrm{i} \in \mathbb{S}^N$, $\tilde{\mathbf{d}}_\mathrm{i} \in \mathbb{R}^N$ and $r_i \in \mathbb{R}$ and similarly for the L constraints as $p_j(\boldsymbol{\phi}) = q_j(\mathbf{z}) = \mathbf{z}^T \tilde{\mathbf{A}}_\mathrm{i} \mathbf{z} + 2\tilde{\mathbf{b}_\mathrm{i}}^{\,T} \mathbf{z} + c_i = 0$ for $\tilde{\mathbf{A}}_\mathrm{i} \in \mathbb{S}^N, \tilde{\mathbf{b}}_\mathrm{i} \in \mathbb{R}^N$ and $c_i \in \mathbb{R}$. The transformation $k : \mathbb{R}^K \to \mathbb{R}^N$ as $\mathbf{z} = k(\boldsymbol{\phi})$ depends on the original objective and constraints functions, and it may range from vectorizing the variable, *i.e.*, from $\phi = \mathbf{X} \in \mathbb{R}^{m \times n}$ to the vector $\mathbf{z} \in \mathbb{R}^{mn \times 1}$ to introducing new variables, for example, the matrix $\mathbf{Q} \doteq \mathbf{qr}$

and then, vectorizing it as $\mathbf{z} = \text{vec}(\mathbf{Q})$. Note that new constraints may be required to maintain the original domain defined by the L original constraints to avoid the introduction of new feasible points and so the final problem have, in general, M quadratic constraints. In a similar manner, the cost attained by the original point must be maintained after the transformation, or at most, be shifted by a constant $S \in \mathbb{R}$ and/or scaled by another positive constant $\alpha \in \mathbb{R}_+$ for *all* feasible points, *i.e.*, $f(\mathbf{z}) = \alpha h(\boldsymbol{\phi}) + S$. The non-homogeneous problem is an instance of a **Quadratically Constrained Quadratic Problem** (QCQP) and formally

$$f^\star = \inf_{\mathbf{z}\in\mathbb{R}^N} \mathbf{z}^T \tilde{\mathbf{C}}_\text{i} \mathbf{z} + 2\tilde{\mathbf{d}}_\text{i}^{\ T} \mathbf{z} + r_i, \text{ subject to } \mathbf{z}^T \tilde{\mathbf{A}}_\text{i} \mathbf{z} + 2\tilde{\mathbf{b}}_\text{i}^{\ T} \mathbf{z} + c_i = 0, i = [M], \tag{2.60}$$

where $f^\star$ equals $h^\star$ up to a non-negative scale $\alpha \in \mathbb{R}_+$ and a scalar $S \in \mathbb{R}$ added to the objective function, both constants for all feasible points. We say problem 2.60 is equivalent to OG in the sense that any solution $\mathbf{z}$ of the former can be mapped to the respective solution of the original problem $\boldsymbol{\phi}$ while maintaining the cost up to the above-mentioned factors. Since this holds also for the optimal solution, we can retrieve the optimal $\boldsymbol{\phi}^\star$ from the optimal solution $\mathbf{z}^\star$.

For simplicity, we homogenize the problem on the N-D vector $\mathbf{z}$ by introducing the variable $y = 1$ and the vector $\mathbf{x} \doteq [\mathbf{z}^T, y]^T \in \mathbb{R}^{N+1}$. This allows us to define the matrices $\mathbf{C}$ and $\mathbf{A}_\text{i} \in \mathbb{S}^{N+1}$ for $i = [M]$ for the cost and constraints, respectively, as

$$\mathbf{C} \doteq \begin{pmatrix} \tilde{\mathbf{C}}_\text{i} & \tilde{\mathbf{d}}_\text{i} \\ \tilde{\mathbf{d}}_\text{i}^{\ T} & r_i \end{pmatrix} \text{ and } \mathbf{A} \doteq \begin{pmatrix} \tilde{\mathbf{A}}_\text{i} & \tilde{\mathbf{b}}_\text{i} \\ \tilde{\mathbf{b}}_\text{i}^{\ T} & c_i \end{pmatrix}. \tag{2.61}$$

With these matrices, we re-write the cost as $\mathbf{z}^T \tilde{\mathbf{C}}_\text{i} \mathbf{z} + 2\tilde{\mathbf{d}}_\text{i}^{\ T} \mathbf{z} + r_i = 0 \Leftrightarrow \mathbf{z}^T \tilde{\mathbf{C}}_\text{i} \mathbf{z} + 2y\tilde{\mathbf{d}}_\text{i}^{\ T} \mathbf{z} + r_i y^2 = 0 \Leftrightarrow \mathbf{x}^T \mathbf{C} \mathbf{x} = 0$, and similarly for the constraints as $\mathbf{z}^T \tilde{\mathbf{A}}_\text{i} \mathbf{z} + 2\tilde{\mathbf{b}}_\text{i}^{\ T} \mathbf{z} + c_i = \mathbf{x}^T \mathbf{A} \mathbf{x} = 0$ for $i = [M]$. To ensure the homogeneous variable $y \in \mathbb{R}$ has value one, we need to include an additional constraint as $\mathbf{e}_{N+1}^T \mathbf{x} = 1$ where $\mathbf{e}_{N+1} \doteq [\underbrace{0, \ldots, 0}_{N}, 1]^T \in \mathbb{R}^{N+1}$ is a one-hot vector. The constraint is linear on $\mathbf{x}$ and this may present some problems depending on the approach followed to solve the future relaxation. Specifically, Shor's relaxation [Shor, 1987] has been shown to perform better when all the constraints are quadratic forms, *e.g.*, [Bao et al., 2011, Shor, 1992, Wolkowicz, 2000]. Therefore, we prefer to constrain the homogeneous variable y to have *norm* one $y^2 = 1$,

that is, we allow $y = \pm 1$, and also the vector $\pm\mathbf{x}$. The requirement on y is expressed as the quadratic form $\mathbf{x}^T(\mathbf{e}_{N+1}\mathbf{e}_{N+1}^T)\mathbf{x} = \mathbf{x}^T\mathbf{L}\mathbf{x} = 1$ with

$$\mathbf{L} \doteq \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} \in \mathbb{S}^{N+1} \tag{2.62}$$

Now we can re-formulate the original problem in its standard, homogeneous QCQP form

$$f^\star = \min_{\mathbf{x}\in\mathbb{R}^{N+1}} \mathbf{x}^T\mathbf{C}\mathbf{x}, \quad \text{subject to } \mathbf{x}^T\mathbf{A}_i\mathbf{x} = 0, i = [M], \ \mathbf{x}^T\mathbf{L}\mathbf{x} = 1. \quad \text{(QCQP)}$$

When all the matrices $\tilde{\mathbf{C}}_i, \tilde{\mathbf{A}}_i$ are positive semidefinite, the problem is convex and thus, it can be solved, but otherwise, and in general, problem QCQP is nonconvex and NP-hard to solve. However, some configurations make the problem tractable, although not necessarily convex, and we can solve it efficiently, for instance, problems with only one variable, one constraint, one interval constraint and constraints with only one negative eigenvalue, see [Park and Boyd, 2017] for further details. Nevertheless, in general, we are interested in nonconvex problems and how to solve them or, at least, how to approximate their solutions. For that, we focus on deriving relaxations for them that can be solved and provide some information about the solution to the original problem. The rest of this section elaborates about these relaxations and how to use them to tackle the nonconvex QCQP, whereas Table 2.1 collects the assumptions required by these approaches, as well as their advantages and disadvantages.

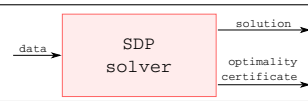
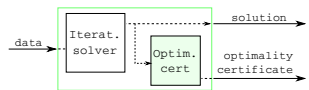
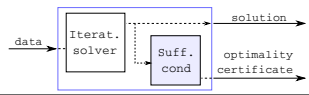
Approach	Advantages	Disadvantages	General scheme
SDP solvers	"straightforward"; can be tightened	polynomial complexity; specific tools	
Certifiable algorithms	more efficient; no specific tools	conditions on formulation; require primal solver	
Sufficient condition	faster; bound optimal region	fewer detected solutions; formulation specific	

Table 2.1: Certifiable methods leveraged in this thesis, with their advantages and disadvantages. The last column shows the general pipeline for these methods.

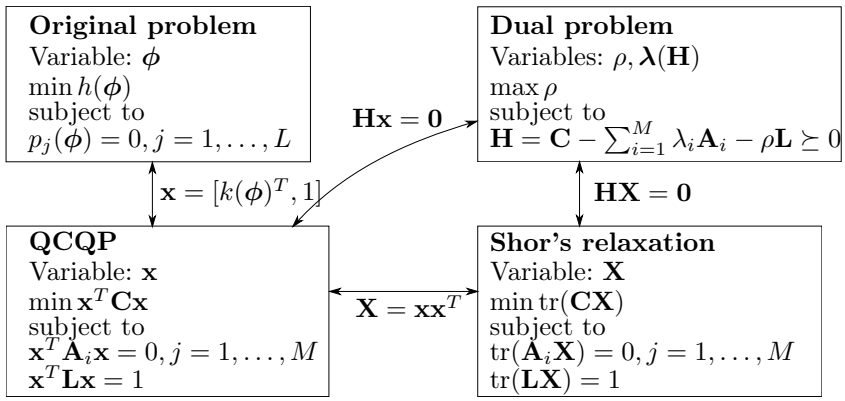


Figure 2.14: Relation between the problems addressed in this thesis and their variables.

2.3.3 Convex relaxations

The goal of these relaxations is to obtain convex, tractable problems that approximate well the original problem QCQP, which we call *primal problem*. Under some conditions, the solution to the relaxation allows us to retrieve and certify the optimal solution for the original problem, and in those cases, we say the relaxation was *tight*. In this thesis, we leverage two well-known convex relaxations, the so-called **dual problem** and the **Shor's relaxation**. Both problems are derived from the homogeneous form in Equation (QCQP) through standard procedure, that can be found *e.g.*, in [Boyd and Vandenberghe, 2004], but we summarize them next for completeness. Last, and since the relations between the problems and their variables are the cornerstone of this thesis, we provide a visual representation in Figure 2.14.

2.3.3.1 Dual problem

The idea behind this dual formulation is to add to the primal cost function a weighted sum of the constraints, see [Boyd and Vandenberghe, 2004, Sec. 5.1] for further information. The resulting function is known as the *Lagrangian* $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \rho)$, which for problem QCQP takes the form

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \rho) \doteq \mathbf{x}^T \mathbf{C} \mathbf{x} - \sum_{i=1}^{M} \lambda_i \mathbf{x}^T \mathbf{A}_i \mathbf{x} + \rho(1 - \mathbf{x}^T \mathbf{L} \mathbf{x}) \tag{2.63}$$

$$= \mathbf{x}^T \big(\mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho \mathbf{L}\big) \mathbf{x} + \rho. \tag{2.64}$$

The variables $\boldsymbol{\lambda} \in \mathbb{R}^M$ and $\rho \in \mathbb{R}$ are known as the *dual variables* or *Lagrange multipliers*. The so-called *dual function* $d(\boldsymbol{\lambda}, \rho)$ seeks the minimum value of the Lagrangian over the variable $\mathbf{x}$, formally

$$d(\boldsymbol{\lambda}, \rho) = \inf_{\mathbf{x} \in \mathbb{R}^{N+1}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \rho) = \inf_{\mathbf{x} \in \mathbb{R}^{N+1}} \mathbf{x}^T \big(\mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho \mathbf{L}\big)\mathbf{x} + \rho \tag{2.65}$$

Since the variable of the problem is the unconstrained vector $\mathbf{x}$, the cost is unbounded below and so the optimal cost $d^\star$ takes values on $-\infty$, see Section 2.3.1.1. Notice the dual function is convex since it is the point-wise infimum of a affine function in $(\boldsymbol{\lambda}, \rho)$, even when the original problem QCQP is nonconvex.

Our interest in the dual function in Equation (2.65) is that its cost $d(\boldsymbol{\lambda}, \rho)$ is always non-greater than the cost $f(\mathbf{x})$ for the primal QCQP, see [Boyd and Vandenberghe, 2004, Sec. 5.1.3]. This is so because for each feasible primal $\mathbf{x}$ the term $\sum_{i=1}^{M} \lambda_i \mathbf{x}^T \mathbf{A}_i \mathbf{x} + \rho(1 - \mathbf{x}^T \mathbf{L} \mathbf{x})$ is zero, and so $d(\boldsymbol{\lambda}, \rho) \leq f(\mathbf{x})$ for every feasible $\mathbf{x}$. Note, however, that if $d(\boldsymbol{\lambda}, \rho) = -\infty$, the lower bound does not provide any useful information about the primal problem. We are interested in these cases in which this lower bound provides some actual, finite bound. For that, we seek the best lower bound, *i.e.*, the maximum cost. The **dual problem** tries to find this value, and formally

$$d^\star = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^M, \rho \in \mathbb{R}} d(\boldsymbol{\lambda}, \rho) = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^M, \rho \in \mathbb{R}} \inf_{\mathbf{x} \in \mathbb{R}^{N+1}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \rho) \tag{D}$$

The optimal solution to this problem $(\boldsymbol{\lambda}^\star, \rho^\star)$ is known as the *dual optimal* or *optimal Lagrange multipliers* and attains the optimal dual cost $d^\star$. The dual problem is always convex as it is the maximum of a convex function and by construction it still provides a lower bound on the optimal cost of the primal QCQP, relation that can be written as the chain of inequalities

$$d(\boldsymbol{\lambda}, \rho) \leq d^\star \leq f^\star \leq f(\mathbf{x}). \tag{2.66}$$

that holds for any feasible primal $\mathbf{x}$ and dual $(\boldsymbol{\lambda}, \rho)$ point. This property is known as *weak duality* and it's key to understand the use of convex relaxations to solve and certify nonconvex problems with the form QCQP. The difference between the optimal primal and dual cost $f^\star - d^\star \geq 0$ is known as the *optimal duality gap*, and we say that *strong duality* holds *iff* $d^\star = f^\star$, that is, the duality gap is zero. This is the situation we seek when deriving convex relaxations, as the optimal dual solution becomes an optimality certificate of the primal solution. Nonetheless, strong duality does not hold in general, although well-known exceptions exist. Among them, we must highlight *Slater's condition* that states that strong duality holds if the primal problem is convex and there exists a strictly feasible point, that is, a point fulfilling the inequality constraints with strict inequality. We elaborate more on these conditions at the end of this section.

Given the form of the primal QCQP, we can re-write the dual problem in a simpler, more explicit form. As the dual function operates on the $(N+1)$-D vector $\mathbf{x}$, the behavior of the function depends on the matrix $\mathbf{H} \doteq \mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho\mathbf{L} \in \mathbb{S}^{N+1}$, known as *the Hessian of the Lagrangian*. The value of the expression $\mathbf{x}^T\mathbf{H}\mathbf{x}$ is not bounded below for generic $\mathbf{H}$ due to $\mathbf{x}$ being unconstrained. For the dual function to attain a finite value, $\mathbf{H}$ is required to be positive semidefinite (PSD), and in this case, the minimum value of the expression $\mathbf{x}^T\mathbf{H}\mathbf{x}$ is finite and zero and the cost for the dual function in Equation (2.65) is ρ. Taking this into account, we re-formulate the dual function explicitly as

$$d(\boldsymbol{\lambda}, \rho) = \inf_{\mathbf{x} \in \mathbb{R}^{N+1}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \rho) = \begin{cases} \rho & \text{if } \mathbf{H} = \mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho\mathbf{L} \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \tag{2.67}$$

Further, since the dual problem in D takes the maximum of the dual function and we seek finite values, we can restrict our interest to the first case with $\mathbf{H} \succeq 0$, and thus we re-formulate the dual problem as the *constrained* problem

$$d^\star = \max_{\boldsymbol{\lambda} \in \mathbb{R}^M, \rho \in \mathbb{R}} \rho \quad \text{subject to } \mathbf{H} = \mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho\mathbf{L} \succeq 0. \tag{DUAL}$$

With this form the dual problem is an instance of a SemiDefinite Positive (SDP) problem that remains convex and can be solved with off-the-shelf tools, see Section 2.3.3.5.

Optimality conditions Therefore, the dual problem in DUAL can be solved directly from scratch, and returns the optimal solution as $(\boldsymbol{\lambda}^\star, \rho^\star)$ and Hessian $\mathbf{H}^\star$. The optimal solutions for the primal QCQP and its dual problem DUAL are related by so-called Karush-Kuhn-Tucker (KKT) conditions as explained in [Boyd and Vandenberghe, 2004, Sec. 5.5.3, 5.9]. These conditions assume strong duality holds and are the cornerstone of certifiable algorithms explained later in Section 2.3.4. Formally

$$\mathbf{H}^\star \mathbf{x}^\star = \mathbf{0}_{N+1} \qquad \text{strong duality} \tag{2.68}$$

$$\mathbf{x}^{\star T} \mathbf{L} \mathbf{x}^\star = 1 \text{ and } \mathbf{x}^{\star T} \mathbf{A}_i \mathbf{x}^\star = 0, i = [M] \qquad \text{primal feasibility} \tag{2.69}$$

$$\mathbf{H}^\star = \mathbf{C} - \sum_{i=1}^{M} \lambda_i^\star \mathbf{A}_i - \rho^\star \mathbf{L} \succeq 0 \qquad \text{dual feasibility} \tag{2.70}$$

These conditions are necessary for any primal and dual solutions to be the optimal solutions for their respective problems, and require the objective and constraints functions to be differentiable. The conditions are also sufficient if the primal problem (here, problem QCQP) is convex.

Retrieving a primal, feasible solution from the dual solution: Further, based on the KKT conditions we can try to retrieve a primal solution

for the original QCQP from the dual optimal solution $(\boldsymbol{\lambda}^\star, \rho^\star)$. In practice, though, off-the-shelf tools solve the dual problem and its dual, which we explain in the next sections. This procedure, however, assumes only the solution to DUAL is available.

For simplicity, the next explanation considers only one of the dual solutions $(\boldsymbol{\lambda}^\star, \rho^\star)$ and its associated Hessian $\mathbf{H}^\star \succeq 0$. The relation between the optimal primal solution $\mathbf{x}^\star$ and the optimal dual solution $(\boldsymbol{\lambda}^\star, \rho^\star)$ comes from the KKT conditions in Eqs. (2.68) to (2.70), which assumes strong duality holds, *i.e.*, $d^\star = f^\star$ [Boyd and Vandenberghe, 2004, Sec. 5.5.2]. In this case, the unknown primal $\mathbf{x}^\star$ is *a* minimizer of the Lagrangian in Equation (2.64) (the function can have more than one minimizer, so this $\mathbf{x}^\star$ is just *one of them*), that is, it's a minimum of the expression $\mathbf{x}^T\mathbf{H}\mathbf{x}$ which is bounded below by zero when $\mathbf{H}$ is PSD, and so $\mathbf{x}^{\star T}\mathbf{H}^\star\mathbf{x}^\star = 0$. Since $\mathbf{H}^\star$ is PSD due to dual feasibility, then $\mathbf{x}^{\star T}\mathbf{H}^\star\mathbf{x}^\star = 0 \Leftrightarrow \mathbf{H}^\star\mathbf{x}^\star = \mathbf{0}_{N+1}$, thus the primal solution $\mathbf{x}^\star$ lays in the nullspace of the Hessian $\mathbf{H}^\star$. If the Hessian has an unique zero eigenvalue, then the primal $\mathbf{x}^\star$ can be retrieved without further operations. Due to numerical instabilities, it's recommended to project the retrieved solution $\mathbf{x}^\star$ to the primal feasible set and check afterwards the cost attained by this solution $\hat{\mathbf{x}}$ as $f(\hat{\mathbf{x}})$. If this cost equals the optimal dual cost $d^\star$, then the projected solution is the global optimum for the original problem QCQP, the optimal dual solution is an *optimality certificate* of this and the strong duality assumption was correct.

Constraint qualifications:

As we mention before, Slater's condition assures that strong duality holds if the primal problem is convex under strict feasibility. This condition is known as a *constraint qualification* (CQ), and it's not the only one for problems like the one at hand. We present next the four most common CQ for problems with equality constraints $h_i(\mathbf{x}) = 0$ and inequality constraints $g_i(\mathbf{x}) \leq 0$ and we refer the reader to [Peterson, 1973] for a review of all CQ.

The Abadie Constraint Qualification (ACQ) holds at the local minimizer $\mathbf{x}$ if $\mathcal{T}(\mathbf{x}) = \mathcal{L}(\mathbf{x})$ where $\mathcal{T}(\mathbf{x})$ is the tangent cone at $\mathbf{x}$ defined by

$$\mathcal{T}(\mathbf{x}) = \left\{ \mathbf{d} \in \mathbb{R}^M \mid \exists \{\mathbf{x}^k\} \subset \Sigma, t_k \downarrow 0 : \mathbf{x}^k \to \mathbf{x} \text{ and } \frac{\mathbf{x}^k - \mathbf{x}}{t_k} \to \mathbf{d} \right\}$$

and $\mathcal{L}(\mathbf{x})$ is the linearized cone at $\mathbf{x}$ defined as

$$\mathcal{L}(\mathbf{x}) = \left\{ \mathbf{d} \in \mathbb{R}^M \mid \nabla g_i(\mathbf{x})^T\mathbf{d} \leq 0 (i = [N]), \nabla h_j(\mathbf{x})^T\mathbf{d} = 0 (j = [P]) \right\},$$

where we only consider the N inequalities constraints such that $g_i(\mathbf{x}) = 0$.

The Guignard Constraint Qualification (GCQ) holds at the local minimizer $\mathbf{x}$ if $\mathcal{T}^*(\mathbf{x}) = \mathcal{L}^*(\mathbf{x})$, that is, their dual cones coincide.

The Mangasarian-Fromovitz Constraint Qualification (MFCQ) holds at a local minimizer $\mathbf{x}$ if the gradients for the equality constraints are linearly independent and there exists a vector $\mathbf{d}$ such that $\nabla h_i(\mathbf{x})^T\mathbf{d} = 0$ and $\nabla g(\mathbf{x})^T\mathbf{d} < 0$.

The Linear Independence Constraint Qualification (LICQ) holds at a local minimizer $\mathbf{x}$ if the gradients of the constraints are linearly independent.

Last, the implication LICQ $\Rightarrow$ MFCQ $\Rightarrow$ ACQ $\Rightarrow$ GCQ holds for every minimizer $\mathbf{x}$. Notice that MFCQ equals LICQ when only equality constraints are present.

2.3.3.2 Problems with one quadratic constraint

QCQP with only one constraint are a special case, as they present strong duality *and* can be solved efficiently. We summarize next the approach proposed by Hmam in [Hmam, 2010] to solve efficiently QCQP problems with one equality constraint. The problem is stated in its non-homogeneous form as

$$f^\star = \min_{\mathbf{z}\in\mathbb{R}^N} \mathbf{z}^T\mathbf{C}\mathbf{z} + 2\mathbf{c}^T\mathbf{z} + c, \quad \text{subject to } \mathbf{z}^T\mathbf{A}\mathbf{z} + 2\mathbf{a}^T\mathbf{z} + a = 0. \tag{2.71}$$

The objective of the approach is to solve the problem through a primal-dual formulation, that is, the primal QCQP and the dual DUAL problems are solved at the same time, hence guaranteeing the optimality of the solution upon convergence. To alleviate algebraic computations, the method re-writes the problem to have a simpler form. Since the coefficient matrix $\mathbf{C}$ is positive definite, its eigendecomposition as $\mathbf{C} = \mathbf{U}\mathbf{D}_\mathrm{C}\mathbf{U}^T$ for $\mathbf{U} \in \mathbb{O}(N)$ and $\mathbf{D}_\mathrm{C}$ diagonal, allows to create the matrix $\mathbf{S} = \mathbf{D}_\mathrm{C}^{-1/2}\mathbf{U}^T \in \mathbb{R}^{N\times N}$ such that $\mathbf{S}\mathbf{C}\mathbf{S}^T = \boldsymbol{I}_N$. Further, since $\mathbf{C}, \mathbf{A}$ commute [Golub and Van Loan, 2013, Sec. 8.7.1] we have that $\mathbf{S}\mathbf{A}\mathbf{S}^T = \mathbf{D}_\mathrm{A}$. The transformation is applied also to the variable $\mathbf{z}$ as $\mathbf{z} = \mathbf{S}^T\mathbf{y}$ and the linear terms as $\mathbf{b} = \mathbf{S}\mathbf{c}$ and $\mathbf{d} = \mathbf{S}\mathbf{a}$. The final change of variables comes as $\mathbf{w} = \mathbf{y} + \mathbf{b} = \mathbf{S}^{-T}\mathbf{z} + \mathbf{S}\mathbf{c}$, which allows us to write the final form of the problem as

$$f^\star = \min_{\mathbf{w}\in\mathbb{R}^N} \mathbf{w}^T\mathbf{w} + r, \quad \text{subject to } \mathbf{w}^T\mathbf{D}_\mathrm{A}\mathbf{w} + 2\mathbf{q}^T\mathbf{w} + q = 0, \tag{2.72}$$

where $\mathbf{q} = \mathbf{d} - \mathbf{D}_\mathrm{A}\mathbf{b}$, $q = a + \mathbf{b}^T\mathbf{D}_\mathrm{A}\mathbf{b} - 2\mathbf{d}^T\mathbf{b}$ and $r = c - \mathbf{b}^T\mathbf{b} = u - \mathbf{c}\mathbf{S}^T\mathbf{S}\mathbf{c}$, and since r does not depend on the original variable $\mathbf{z}$, it can be dropped from the optimization problem. Whereas problem 2.72 and 2.71 are equivalent in the sense that we can map one-to-one their feasible points, the former has a simpler form as the quadratic matrix of the constraint is diagonal, and the cost has only the quadratic term whose coefficient matrix is the identity. We consider that the vector $\mathbf{q} \neq 0$, as otherwise problem 2.72 can be re-formulated as a linear program. The dual problem is obtained through the standard procedure, and the Lagrangian has the form

$$\mathcal{L}(\mathbf{w}, \lambda) = \mathbf{w}^T\mathbf{w} - \lambda(\mathbf{w}^T\mathbf{D}_\mathrm{A}\mathbf{w} + 2\mathbf{q}^T\mathbf{w} + q) = \mathbf{w}^T(\boldsymbol{I}_N - \lambda\mathbf{D}_\mathrm{A})\mathbf{w} - 2\lambda\mathbf{q}^T\mathbf{w} - \lambda q, \tag{2.73}$$

and the dual function takes the form

$$d(\lambda) = \inf_{\mathbf{w}\in\mathbb{R}^N} \mathcal{L}(\mathbf{w},\lambda) = \begin{cases} -\lambda q - \lambda^2 \mathbf{q}^T (\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A})^{-1}\mathbf{q} & \boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A} \succeq 0, \\ & \lambda\mathbf{q} \in \mathcal{R}(\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A}) \\ -\infty & \text{otherwise} \end{cases} . \tag{2.74}$$

The dual problem has the constrained form

$$\begin{aligned} d^\star = \max_{\lambda\in\mathbb{R}} & -\lambda q - \lambda^2 \mathbf{q}^T (\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A})^{-1}\mathbf{q} \\ \text{subject to } & \boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A} \succeq 0,\ \lambda\mathbf{q} \in \mathcal{R}(\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A}). \end{aligned} \tag{2.75}$$

The KKT conditions for this problem are a specification of the general ones in Equations (2.68) to (2.70) with the form

$$(\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A})\mathbf{w} = \lambda\mathbf{q} \tag{2.76}$$

$$\mathbf{w}^T \mathbf{D}_\mathrm{A} \mathbf{w} + 2\mathbf{q}^T\mathbf{w} + q = 0 \tag{2.77}$$

$$\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A} \succeq 0 \tag{2.78}$$

Equation (2.78) allows to express the region of optimality for the optimal multiplier $\lambda^\star$ in terms of the diagonal entries of the matrix $\mathbf{D}_\mathrm{A}$ that are $\mu_{\max} \geq \cdots \geq \mu_{\min}$ as $\lambda^\star \leq 1/\mu_{\max}$ if $\mathbf{D}_\mathrm{A} \succeq 0$ and as $1/\mu_{\min} \leq \lambda^\star \leq 1/\mu_{\max}$ if $\mathbf{D}_\mathrm{A}$ is indefinite. From Equation (2.76), we express the primal solution as $\mathbf{w}^\star = \lambda(\boldsymbol{I}_N - \lambda^\star \mathbf{D}_\mathrm{A})^{-1}\mathbf{q}$ and so the matrix $\boldsymbol{I}_N - \lambda^\star \mathbf{D}_\mathrm{A}$ is required to be full-rank, that is, the optimal multiplier $\lambda^\star$ cannot take the extreme values of the interval. The expression for the optimal $\mathbf{w}^\star$ is inserted into the last condition in Equation (2.77), leading to the form

$$K(\lambda) \doteq \lambda^2 \mathbf{q}^T (\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A})^{-T} \mathbf{D}_\mathrm{A} (\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A})^{-1}\mathbf{q} + 2\lambda \mathbf{q}^T (\boldsymbol{I}_N - \lambda \mathbf{D}_\mathrm{A})^{-1}\mathbf{q} + q = 0. \tag{2.79}$$

which has a zero for the optimal $\lambda^\star$, thus $K(\lambda^\star) = 0$. [Hmam, 2010, Prop. 2] lists under which conditions the polynomial in Equation (2.79) has a unique solution and among these conditions, $\mathbf{A}$ being indefinite is sufficient for uniqueness. Further, note that the algorithm focuses on finding the zero on an interval of the expression in Equation (2.79), which can be done through Newton's or bisection method without specific tools. Last, if the final solution has optimal multiplier $\lambda^\star$ within the region given by the above-mentioned values of $\mu_{\max}$ and $\mu_{\min}$, then the solutions $\mathbf{w}$ and $\lambda^\star$ are the global optima for their respective problems.

2.3.3.3 Shor's relaxation

The second convex relaxation we leverage in this thesis is known as Shor's relaxation [Shor, 1987], and as the dual problem, takes the form of an SDP

problem that can be solved with off-the-shelf tools, see Section 2.3.3.5. Instead of the dual function, this relaxation is derived by dropping (relaxing) the requirement on the rank of the matrix.

First, we use the fact that $\text{tr}(a) = a$ for $a \in \mathbb{R}$ to write the cost and constraints of the primal QCQP as $\mathbf{x}^T\mathbf{C}\mathbf{x} = \text{tr}(\mathbf{x}^T\mathbf{C}\mathbf{x})$. Since the trace operator fulfills the cyclic property $\text{tr}(abc) = \text{tr}(bca)$, the equality $\text{tr}(\mathbf{x}^T\mathbf{C}\mathbf{x}) = \text{tr}(\mathbf{C}\mathbf{x}\mathbf{x}^T)$ holds. We introduce the PSD matrix $\mathbf{X} \doteq \mathbf{x}\mathbf{x}^T \in \mathbb{S}_+^{N+1}$, *i.e.*, the outer product of $\mathbf{x}$ with itself and arrive at the expression $f(\mathbf{x}) = \text{tr}(\mathbf{C}\mathbf{X})$ where $\mathbf{X} \in \mathbb{S}_+^{N+1}$ has rank one, and equivalently for the constraints. We reformulate problem QCQP with this matrix as

$$f^\star = \min_{\mathbf{X}\in\mathbb{S}_+^{N+1}} \text{tr}(\mathbf{C}\mathbf{X}), \text{ subject to } \text{tr}(\mathbf{A}_i\mathbf{X}) = 0, i = [M], \text{tr}(\mathbf{L}\mathbf{X}) = 1, \text{rank}(\mathbf{X}) = 1 \tag{2.80}$$

Problems QCQP and 2.80 are equivalent, in the sense that: (1) their feasible sets are in one-to-one correspondence; and (2) the costs attained by corresponding solutions are the same so are the optimal costs. However, the rank constraint makes problem 2.80 nonconvex, as the original QCQP. The Shor's relaxation is obtained by dropping this constraint, which leads to the relaxation

$$g^\star = \min_{\mathbf{X}\in\mathbb{S}_+^{N+1}} \text{tr}(\mathbf{C}\mathbf{X}) \quad \text{subject to } \text{tr}(\mathbf{A}_i\mathbf{X}) = 0, i = [M], \text{tr}(\mathbf{L}\mathbf{X}) = 1. \quad \text{(Shor)}$$

As the dual problem DUAL, problem Shor is convex and a SDP instance. Further, since the feasible set of Shor contains the set of QCQP but it's also larger since we removed the rank restriction, then the cost $g^\star$ is a lower bound on the primal $f^\star$, formally

$$g(\mathbf{X}) \leq g^\star \leq f^\star \leq f(\mathbf{x}). \tag{2.81}$$

Notice that this chain of inequalities is similar to that for the dual problem in Equation (2.66), so is the purpose behind this relaxation. We say that the relaxation is *tight iff* $g^\star = f^\star$, which provides with an optimality certificate.

Retrieving a primal, feasible solution from the relaxation As with the dual problem, if the relaxation is tight we can retrieve the optimal primal solution $\mathbf{x}^\star$ from the optimum $\mathbf{X}^\star$. The straightforward way to leverage this relaxation is to solve it with any of the off-the-shelf tools available and then try to extract a solution for the original QCQP, as it was done with the dual problem DUAL.

In this case, though, the retrieval relies on the eigendecomposition of the solution $\mathbf{X}^\star$ and we seek the rank-one approximation of $\mathbf{X}^\star$ such that $\mathbf{X} = \mathbf{x}^\star\mathbf{x}^{\star T}$, for $\mathbf{x}^\star \in \mathbb{R}^{N+1}$ and $\mathbf{X}$ is the closest, rank-one PSD matrix in the Frobenius norm sense to $\mathbf{X}^\star$. Although the rank-one approximation is the standard condition, for certain problems the optimal $\mathbf{X}^\star$ has rank larger than one yet still provide with a tight relaxation from which we can extract the optimal primal solution $\mathbf{x}^\star$. For example, if the matrix $\mathbf{X}^\star$ is block-diagonal with non-null blocks,

then the rank will be necessary larger than one. If each block has rank one, we can still retrieve the optimal solution for the primal QCQP by taking the eigen-decomposition of each block separately. Note, however, that this implies we can also retrieved a family of solutions. Let the solution $\mathbf{X}^\star$ have three non-null blocks with rank-one decomposition $\mathbf{x}_1\mathbf{x}_1^T, \mathbf{x}_2\mathbf{x}_2^T$ and $\mathbf{x}_3\mathbf{x}_3^T$, not necessary with the same dimension. Then the matrix $\hat{\mathbf{Y}} = \text{blkdiag}\,(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathbb{R}^{(N+1)\times 3}$ leads to $\mathbf{X}^\star = \hat{\mathbf{Y}}\hat{\mathbf{Y}}^T$, so does any matrix with form $\hat{\mathbf{Y}}\mathbf{O}$ for $\mathbf{O} \in \mathbb{O}(3)$ since $\mathbf{O}\mathbf{O}^T = \boldsymbol{I}_3$. Further, given the cyclic property of the trace we have that $\text{tr}(\hat{\mathbf{Y}}^T\mathbf{C}\hat{\mathbf{Y}}) = \text{tr}(\mathbf{O}^T\hat{\mathbf{Y}}^T\mathbf{C}\hat{\mathbf{Y}}\mathbf{O}) = \text{tr}(\hat{\mathbf{Y}}^T\mathbf{C}\hat{\mathbf{Y}}\mathbf{O}\mathbf{O}^T) = \text{tr}(\hat{\mathbf{Y}}^T\mathbf{C}\hat{\mathbf{Y}})$, that is, there exists a family of feasible, primal solutions, not necessary with a sparse structure. Note that alterations of $\hat{\mathbf{Y}}$ with two and one columns are also feasible. In practice, off-the-shelf tools return the solutions with block-diagonal structure, see Section 2.3.3.5, and therefore, instead of forcing the rank-one requirement, we will say that the relaxation is tight *iff* it is possible to retrieve a primal feasible solution $\mathbf{x}^\star$ from $\mathbf{X}^\star$ that attains the same cost, *i.e.*, $g^\star = \text{tr}(\mathbf{C}\mathbf{X}^\star) = f^\star = \text{tr}\left(\mathbf{x}^{\star T}\mathbf{C}\mathbf{x}^\star\right)$. Notice that this cost condition is the optimality certificate for the retrieved solution $\mathbf{x}^\star$ by Equation (2.81).

2.3.3.4 Relation between relaxations

In the previous sections the dual problem and Shor's relaxation have been presented as isolated relaxations. However, this section shows that the dual problem is the dual of Shor's relaxation and viceversa. The importance of this relation is linked to the conditions that we have considered to call the relaxations tight and to retrieve optimal, primal solutions.

We develop this relation starting from the dual problem in DUAL, although the same result is obtained from Shor. We derive its dual in the same manner we derive the problem for the QCQP. For that, recall that maximizing the cost g is equivalent to minimizing the cost $-g$, that is, $\max \rho = \min(-\rho)$ and thus the dual problem has the form

$$d^\star = \min_{\boldsymbol{\lambda}\in\mathbb{R}^M, \rho\in\mathbb{R}} -\rho \quad \text{subject to } \mathbf{H} = \mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho\mathbf{L} \succeq 0, \tag{2.82}$$

and its associated Lagrangian $\mathcal{K}(\mathbf{X}, \boldsymbol{\lambda}, \rho)$ is defined as

$$\mathcal{K}(\mathbf{X}, \boldsymbol{\lambda}, \rho) \doteq -\rho - \text{tr}(\mathbf{H}\mathbf{X}) = -\rho - \text{tr}\left((\mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_i - \rho\mathbf{L})\mathbf{X}\right) = \tag{2.83}$$

$$= -\text{tr}(\mathbf{C}\mathbf{X}) + \sum_{i=1}^{M} \lambda_i \,\text{tr}(\mathbf{A}_i\mathbf{X}) + \rho(\text{tr}(\mathbf{L}\mathbf{X}) - 1), \tag{2.84}$$

where the last equality holds due to the trace being a linear operator. The dual function minimizes the Lagrangian over the variables $(\boldsymbol{\lambda}, \rho) \in \mathbb{R}^M \times \mathbb{R}$ as

$$h(\mathbf{X}) = \inf_{\boldsymbol{\lambda}\in\mathbb{R}^M, \rho\in\mathbb{R}} \mathcal{K}(\mathbf{X}, \boldsymbol{\lambda}, \rho) \tag{2.85}$$

$$= \inf_{\boldsymbol{\lambda}\in\mathbb{R}^M, \rho\in\mathbb{R}} -\operatorname{tr}(\mathbf{CX}) + \sum_{i=1}^{M} \lambda_i \operatorname{tr}(\mathbf{A}_i\mathbf{X}) + \rho(\operatorname{tr}(\mathbf{LX}) - 1). \tag{2.86}$$

Last, the dual problem for problem DUAL maximizes the dual function $h(\mathbf{X})$ over the variable $\mathbf{X} \in \mathbb{R}^{(N+1)\times(N+1)}$

$$h^\star = \sup_{\mathbf{X}\in\mathbb{R}^{(N+1)\times(N+1)}} h(\mathbf{X}) = \sup_{\mathbf{X}\in\mathbb{R}^{(N+1)\times(N+1)}} \inf_{\boldsymbol{\lambda}\in\mathbb{R}^M, \rho\in\mathbb{R}} \mathcal{K}(\mathbf{X}, \boldsymbol{\lambda}, \rho). \tag{2.87}$$

As with the original problem, by weak duality we know that $h^\star \leq d^\star$, while in this case, both the primal DUAL and derived dual in 2.87 are convex.

For the specific form of problem DUAL and since the Lagrangian is an affine function on the unconstrained variables $(\boldsymbol{\lambda}, \rho)$, the dual function is unbounded below and the cost takes values on $-\infty$, *except* when *all* the values $\operatorname{tr}(\mathbf{A}_i\mathbf{X})$ for $i = [M]$ and $\operatorname{tr}(\mathbf{LX}) - 1$ are zero at the same time. In this case, the minimum of the Lagrangian is $-\operatorname{tr}(\mathbf{CX})$ as the sum $\sum_{i=1}^{M} \lambda_i \operatorname{tr}(\mathbf{A}_i\mathbf{X}) + \rho(\operatorname{tr}(\mathbf{LX}) - 1)$ is zero independently of $\boldsymbol{\lambda}$ and ρ. The dual function $h(\mathbf{X}, \boldsymbol{\lambda}, \rho)$ can be written explicitly as

$$h(\mathbf{X}, \boldsymbol{\lambda}, \rho) = \begin{cases} -\operatorname{tr}(\mathbf{CX}) & \text{if } \operatorname{tr}(\mathbf{A}_i\mathbf{X}) = 0, i = [M],\ \operatorname{tr}(\mathbf{LX}) = 1 \\ -\infty & \text{otherwise} \end{cases} \tag{2.88}$$

Since we are still interested in the finite values for the cost, we restrict our attention to the first case, and considering again that $\max(-x) = \min(x)$, we re-formulate the dual problem as

$$h^\star = \inf_{\mathbf{X}\in\mathbb{R}^{(N+1)\times(N+1)}} \operatorname{tr}(\mathbf{CX}) \text{ subject to } \operatorname{tr}(\mathbf{A}_i\mathbf{X}) = 0, i = [M], \operatorname{tr}(\mathbf{LX}) = 1 \tag{2.89}$$

Notice that, in general, the expression $\operatorname{tr}(\mathbf{CX})$ still does not have a finite minimum *except* when $\mathbf{X}$ is PSD, and thus, the cost is always nonnegative. We restrict once more our attention to this case, which leads to the final form of the dual problem for problem DUAL

$$h^\star = \min_{\mathbf{X}\in\mathbb{S}^{N+1}} \operatorname{tr}(\mathbf{CX}) \text{ subject to } \operatorname{tr}(\mathbf{A}_i\mathbf{X}) = 0, i = [M], \operatorname{tr}(\mathbf{LX}) = 1, \mathbf{X} \succeq 0, \tag{2.90}$$

which is, in fact, Shor's relaxation as in Equation (Shor), see also [Boyd and Vandenberghe, 2004, Ex. 5.11]. Further, since both problems are convex, we know from Slater's condition that strong duality holds if problem DUAL is strictly feasible, *i.e.*, there exists $\mathbf{H} \succ 0$.

Relation through dual cone:
An alternative route to this result leverages the concept of dual cone [Boyd and Vandenberghe, 2004, 2.6], and specifically, the ones for real scalars and PSD matrices, as explained in Section 2.3.1.1. Recall that the dual cone for the unconstrained, real scalars (values from $-\infty$ to ∞) is the singleton 0, whereas for PSD matrices and the PSD cone, the dual cone is itself [Boyd and Vandenberghe, 2004, Ex. 2.24]. By weak duality [Boyd and Vandenberghe, 2004, Sec. 5.9], it is required that the cost attained by the Lagrangian is always non-greater than the primal cost, *i.e.*, $h^\star \leq d^\star$. In this case, the term $\text{tr}\left((\mathbf{C} - \sum_{i=1}^{M} \lambda_\text{i}\mathbf{A}_\text{i} - \rho\mathbf{L})\mathbf{X}\right)$ has to be non-negative to maintain the lower bound, and since $\mathbf{H}$ is PSD, $\mathbf{X}$ must belong to the dual cone, hence being $\mathbf{X}$ also PSD. Further, since the points $(\boldsymbol{\lambda}, \rho)$ are unconstrained in the original problem DUAL, *i.e.*, they belong to the unconstrained, real space, the associated dual cone is the 0 element, that is $\text{tr}(\mathbf{A}_\text{i}\mathbf{X}) = 0$ for all i and $\text{tr}(\mathbf{L}\mathbf{X}) = 1$.

Optimality conditions:
As with the original problem QCQP, we have a set of optimality conditions for the pair of problems DUAL and Shor. Again, we assume strong duality and thus, from [Boyd and Vandenberghe, 2004, Sec. 5.9.2]

$$\text{tr}(\mathbf{H}\mathbf{X}) = 0 \qquad \text{strong duality} \tag{2.91}$$

$$\mathbf{H} = \mathbf{C} - \sum_{i=1}^{M} \lambda_i \mathbf{A}_\text{i} - \rho\mathbf{L} \succeq 0 \qquad \text{primal feasibility} \tag{2.92}$$

$$\text{tr}(\mathbf{L}\mathbf{X}) = 1 \text{ and } \text{tr}(\mathbf{A}_\text{i}\mathbf{X}) = 0, i = [M] \qquad \text{dual feasibility (1)} \tag{2.93}$$

$$\mathbf{X} \succeq 0 \qquad \text{dual feasibility (2)} \tag{2.94}$$

The conditions are necessary and sufficient for optimality as the primal DUAL and dual 2.3.3.3 are convex.

Since $\mathbf{H}$ and $\mathbf{X}$ are PSD for feasible solutions, the strong duality condition $\text{tr}(\mathbf{H}\mathbf{X}) = 0$ implies what we know also as the *complementary condition* $\mathbf{H}\mathbf{X} = \mathbb{S}^{N+1}$, which in turn implies that the matrices $\mathbf{X}$ and $\mathbf{H}$ commute, that is, they have the same set of eigenvectors [Alizadeh et al., 1997]. Formally, let $\mathbf{U} \in \mathbb{O}(N+1)$ be the set of orthogonal eigenvectors such that: (1) $\mathbf{U}^T\mathbf{X}\mathbf{U} = \mathbf{D}_\text{X}$; and (2) $\mathbf{U}^T\mathbf{H}\mathbf{U} = \mathbf{D}_\text{H}$, where $\mathbf{D}_\text{X}, \mathbf{D}_\text{H}$ are diagonal matrices whose diagonal entries are the eigenvalues of $\mathbf{X}, \mathbf{H}$, respectively. We denote the eigenvalues of $\mathbf{X}$ by $\alpha_1, \ldots, \alpha_{N+1} \geq 0$ and the ones for $\mathbf{H}$ by $\rho_1, \ldots, \rho_{N+1} \geq 0$. From the zero duality gap condition in Equation (2.91) we have that

$$\text{tr}(\mathbf{H}\mathbf{X}) = \text{tr}(\mathbf{U}\mathbf{D}_\text{H}\mathbf{U}^T\mathbf{U}\mathbf{D}_\text{X}\mathbf{U}^T) = \text{tr}(\mathbf{D}_\text{H}\mathbf{D}_\text{X}) = \sum_{i=1}^{N+1} \alpha_i\rho_i = 0 \tag{2.95}$$

which implies that $\alpha_i\rho_i = 0$ for all i, and hence at least one of the eigenvalues is zero. Since the rank of the matrix is the number of nonzero eigenvalues, we

have that, if n is the rank of $\mathbf{X}$ and k is the rank of $\mathbf{H}$, then necessarily $n+k \leq N+1$. We say that *strict complementary* holds *iff* $n+k = N+1$. Another important concept associated with these ranks is *maximally complementarity*, which indicates that, among all the solutions $\mathbf{X}$ and $\mathbf{H}$, the particular pair has the largest ranks among all the possibles, *i.e.*, those that maximizes $n+k$. The importance of this relation lies on the process to retrieve primal solutions. Recall that $\mathbf{X}$ is the solution to Shor's relaxation whereas $\mathbf{H}$ is the Hessian for the dual problem, and the procedure to obtain a primal, feasible solution for the original problem in QCQP given the optimal $\mathbf{X}^\star$ or $\mathbf{H}^\star$ involves conditions on the ranks of these matrices. For instance, if $\mathbf{X}$ has a block-diagonal structure with rank $n = 2$, then we expect the Hessian $\mathbf{H}$ to lose rank to maintain the inequality $n+k \leq N+1$ and potentially also show a block-diagonal structure if $\mathbf{X}$ does.

2.3.3.5 Off-the-shelf tools for SDP relaxations

Despite the assumption about the tightness made by the convex relaxations, an extensive literature shows that they can be used in practice to solve problems, *e.g.*, [Agostinho et al., 2022, Zhao et al., 2020, Zhao, 2020, Briales and Gonzalez-Jimenez, 2017b]. Therefore, we believe necessary to at least list the common tools employed for their resolution. For problems with moderate number of variables and constraints, the convex relaxations can be solved with Interior Point Methods (IPM). Implementations as SeDuMi [Sturm, 1999], MOSEK [ApS, 2019], SDPTA [Yamashita et al., 2012] and SDPT3 [Toh et al., 1999] are among the most used. Whereas these tools can be used directly, we believe necessary to mention CVX[1] [Grant and Boyd, 2014] that allows to model the problem in a simple form while choosing the desired solver, and it simplifies the modeling and testing stages of methods based on convex relaxations for nonconvex problems. Note that these tools solve the pair of primal-dual convex problems DUAL and Shor. Thus, in most cases in practice, we can access the optimal solutions for both problems. This implies that we can retrieve the solution for the primal QCQP following any of the two procedures explained in the sections dedicated to these convex relaxations.

Riemannian staircase:

However, and although these tools can be considered generic, there exists more efficient approaches for convex relaxations that exhibit a specific structure. Among them, we must highlight the so-called *Riemannian staircase* first introduce in [Boumal et al., 2016]. The main idea behind this approach is to solve low-rank decompositions of the original Shor's relaxation and obtain an optimality certificate, as we explain in the next section, to verify if the obtained solution is the global optimum, as these subproblems are nonconvex. The rank, and therefore size, of the variable $\mathbf{Y}$ increases with each step, hence

[1] http://cvxr.com/cvx/

the *staircase* terminology. The k-th iteration of the algorithm seeks the matrix $\mathbf{Y}$ with dimension $(N+1) \times r$, that solves the problem

$$h_k^\star = \min_{\mathbf{Y} \in \mathbb{R}^{(N+1)\times r}} \operatorname{tr}(\mathbf{Y}^T \mathbf{C} \mathbf{Y}) \;\text{ subject to }\; \operatorname{tr}(\mathbf{Y}^T \mathbf{A}_\mathrm{i} \mathbf{Y}) = 0, i = [M], \operatorname{tr}(\mathbf{Y}^T \mathbf{L} \mathbf{Y}) = 1, \tag{2.96}$$

and the relation with the original relaxation in Shor is given by $\mathbf{X} = \mathbf{Y}\mathbf{Y}^T$, which is PSD by construction. The low-rank decomposition avoids to solve large-scale SDP problems which may reach the limits of off-the-shelf tools, while still estimating and certifying the optimal solution. If the low-rank subproblems have a special domain, for instance the feasible set is the set of orthogonal matrices and so $\mathbf{Y} \in \mathbb{O}((N+1) \times r)$, then an efficient Riemannian optimization can be deployed, see Section 2.2.5. This approach was followed in, *e.g.* [Rosen et al., 2019, Briales and Gonzalez-Jimenez, 2017a] for pose-graph optimization and was reported to perform really well. However, to derive a fast solver the determinant constraint associated with the rotation was dropped, although empirically the solutions did fulfill this constraint.

2.3.4 Certifiable algorithms

Solving the relaxation from scratch using the above-mentioned tools may be too computationally expensive for large-scale problems or for real-time applications. An alternative approach is the so-called **certifiable algorithms**, as introduced in [Bandeira, 2016]. This approach does not estimate the primal solution but only computes an optimality certificate for it. Since algorithms that estimate the primal solution are common for most computer vision and robotics problems, this approach presents an attractive alternative to solving the relaxations from scratch with off-the shelf tools. In addition, most of these problems involve well-known domains, *e.g.*, rotations, sphere and Euclidean space, which allows us to leverage approaches that don't require reformulating the variables of the problem in their minimal form, which often suffer from singularities [Cayley, 1846], and use non-Euclidean algorithms as we elaborated in Section 2.2.5. This section summarizes the main concepts associated with these certifiable algorithms and their limitations.

Assuming a primal solution is given, our goal is to certify its optimality. The algorithms employed in this thesis are based on the Karush-Kuhn-Tucker (KKT) conditions for problem QCQP and its associated dual problem DUAL, as written in Equations (2.68) to (2.70). The certifier is then built upon the following observation: if for a given primal solution $\mathbf{x} \in \mathbb{R}^{N+1}$, *i.e.*, a solution fulfilling Equation (2.69), we can obtain Lagrange multipliers $(\boldsymbol{\lambda}, \rho)$, *i.e.*, feasible dual points fulfilling Equation (2.70), such that Equation (2.68) holds then: (a) with $d^\star = f^\star$; and (b) the primal and dual solutions are the optimal solutions for their respective problems. Therefore, we have obtained an optimality certificate for the given primal solution in the multipliers $(\boldsymbol{\lambda}, \rho)$.

The first step estimates a candidate to Lagrange multiplier $(\boldsymbol{\lambda}, \rho)$ from the provided primal feasible solution $\mathbf{x}$. For this, we re-write the strong duality condition in Equation (2.68) to relate the multipliers and the primal solution with the definition of $\mathbf{H} = \mathbf{C} - \sum_{i=1}^{M} \lambda_{\mathrm{i}} \mathbf{A}_{\mathrm{i}} - \rho \mathbf{L}$ as

$$\underbrace{[\mathbf{A}_1\mathbf{x}, \ldots, \mathbf{A}_M\mathbf{x}]}_{\mathbf{J}(\mathbf{x})} \boldsymbol{\lambda} = \mathbf{C}\mathbf{x} - \rho\mathbf{L}\mathbf{x} \tag{2.97}$$

where $\mathbf{J}(\mathbf{x}) \in \mathbb{R}^{(N+1)\times M}$ is the *Jacobian* of the constraints evaluated at the primal point $\mathbf{x}$. The second step checks if the solution from this system $(\boldsymbol{\lambda}, \rho)$ makes the Hessian $\mathbf{H}$ PSD, and if positive, then the point is dual feasible, and the above-mentioned conditions regarding optimality hold. For problems with the form in Equation (QCQP) we can further simplify these conditions by imposing explicitly the relation between the multiplier ρ and the potentially optimal costs of the primal and dual problems, as under the assumption of strong duality $d^\star = \rho^\star = f^\star$. Since we can evaluate the primal objective function $f(\mathbf{x})$ for the candidate to optimal primal solution, we can substitute the value of ρ with this cost, and thus, we only need to estimate the M-D vector $\boldsymbol{\lambda}$ in Equation (2.97).

Under some conditions, this type of certifier can be estimated efficiently, and in general the most consuming step is the eigendecomposition of the Hessian $\mathbf{H}$, especially when the dimension of this matrix (and so the $N+1$ variables of the problem QCQP) is large. However, the first step that estimates the Lagrange multipliers with Equation (2.97) is the one that assumes more conditions through the Jacobian $\mathbf{J}(\mathbf{x})$. In the best-case scenario, the Jacobian is tall and skinny, that is, we have more variables $N+1$ than constraints M, and is also full-rank, condition known as Linear Independence Constraint Qualification (LICQ). In this case, the linear system in Equation (2.97) has a unique solution, provided it exists. In practice, due to numerical inaccuracies the solution that minimizes the residual error $\epsilon_{\mathrm{i}} = \|\mathbf{J}(\mathbf{x})\boldsymbol{\lambda} - \mathbf{C}\mathbf{x} + \rho\mathbf{L}\mathbf{x}\|_2$ is considered for these cases. However, if the number of constraints is larger than the number of variables, and so the Jacobian is short and wide, then the linear system is underdetermined and a family of solutions for $\boldsymbol{\lambda}$ exists. In this case, there may or may not exist a dual feasible point, that is, a point $\boldsymbol{\lambda}$ that makes the Hessian PSD, among the elements of the family. Additionally if the Jacobian is not full-rank, then we have again a family of solutions for the multipliers, among which it may or may not exist a dual feasible solution. For some formulations both conditions may fail at the same time and so the Jacobian is (row) rank deficient *and* have more columns M than rows $N+1$. These conditions involving the Jacobian only depend on the constraints employed to define the feasible set of the problem and the primal solution, not the cost. A remarkable example of these conditions holding are problems involving orthonormal matrices with size $(N+1) \times r$ and using $r(r+1)/2$ constraints, which allows to derive fast certification procedures that estimate

the Lagrange multipliers in closed-form, see [Rosen et al., 2019, Briales and Gonzalez-Jimenez, 2017a].

Unfortunately, there exist other problems for which these conditions fail. Formulations with more constraints than variables usually appear when tightening the relaxations, as more constraints have been empirically shown to maintain the relaxations tight even for highly noisy data and in general adversarial problems, see [Briales and Gonzalez-Jimenez, 2017b, Zhao et al., 2020, Briales et al., 2018]. However, even minimally defined formulations may suffer from this, as the number of quadratic constraints required to fully define the space may be larger than the number of variables, for example 3D rotations including the determinant constraints, see *e.g.* [Tron et al., 2015]. On the other hand, some spaces cannot be defined by the minimum number of polynomial constraints, and those that can be are called *complete intersections* [Smith et al., 2004, Sec. 5.5]. The rank of the Jacobian evaluated at feasible points is related to the space and the constraints used to define it, and it can dropped for some feasible points. An upper bound for this rank is found in the difference between the number of variables and the dimension of the underlying space [Decker and Schreyer, 2007, Th. 4.1.12]. For instance, non-symmetric 3×3 matrices with rank one can be written as $\mathbf{Q} = \mathbf{r}\mathbf{q}^T$ and are defined by 9 variables. From [Bruns and Schwänzl, 1990], we need to constrain the entries of $\mathbf{Q}$ with exactly nine constraints, that involve all the 2×2 minors of the matrix and are therefore quadratic in the elements of $\mathbf{Q}$. However, the rank of the 9×9 Jacobian evaluated at a non-null matrix is four and the dimension of the space is five in general [Hauenstein et al., 2012]. However, the zero matrix, which also fulfills the quadratic constraints, reduces the rank of the Jacobian to zero.

Sufficient conditions for optimality:

In spite of these situations, certifiers that estimate the multipliers in closed-form can be derived. However, the certification can be further simplified by developing what we call a *sufficient condition* for optimality. This condition is simpler and more efficient to use than the certifier, although the ratio of detected optimal solutions is theoretically lower. The idea behind these approaches is to bound the region for which the associated Hessian is PSD and express these limits in terms of the Lagrange multipliers and data of the problem. Since the bound is only sufficient for the matrix to be PSD, but not necessary, some situations won't be characterized by this condition, hence the lower ratio of detected optimal solutions. The condition can be further developed to eliminate the dependence on the Lagrange multipliers, leaving the bounds on the eigenvalues as a function of the problem data and the provided primal solution. If the condition is tight, then it certifies solutions without actually computing neither the Lagrange multipliers nor evaluating the Hessian or its eigendecomposition. These approaches are tailored to specific problems and the literature about them is reduced but examples are found in rota-

tion averaging [Eriksson et al., 2018], point cloud registration with missing data [Iglesias et al., 2020] and triangulation [Hartley and Seo, 2008].

Part I

Pose estimation

3

Relative Pose problem between two cameras

You can stay young as long as you learn

Emily Dickinson

3.1 Introduction

Estimating the relative pose as rotation $\mathbf{R} \in \mathbb{SO}(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$ between two calibrated cameras given a set of N pair-wise features correspondences $(\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i})$ is known as the Relative Pose problem (RRp). The problem appears in most applications with cameras, including visual odometry [Nistér et al., 2004, Scaramuzza and Fraundorfer, 2011], Simultaneous Localization and Mapping (SLAM) [Gomez-Ojeda et al., 2019, Mur-Artal et al., 2015, Azzam et al., 2020] and Structure-from-Motion (SfM) [Özyeşil et al., 2017, Triggs et al., 1999], and the literature is therefore extensive. The gold-standard approach to the RRp, known as Bundle-Adjustment (BA) [Hartley and Zisserman, 2003], optimizes both the relative pose between the cameras and the coordinates of the 3D world points that originated the observations by minimizing the reprojection error [Hartley and Zisserman, 2003]. However, this approach turns to be complex, usually nonconvex [Triggs et al., 1999], and thus it requires a good initialization to converge fast and to the global optimum [Spetsakis and Aloimonos, 1992, Özyeşil et al., 2017]. Nevertheless, even obtaining good initial guesses is a non trivial task [Kneip and Lynen, 2013]

and the preferred alternative is to estimate first the relative pose between the cameras without the coordinates of the 3D points and then feed this solution to the BA optimization.

For a general configuration, the relative pose $(\mathbf{R}, \mathbf{t})$ is solved through the essential matrix, see Section 2.2.2, with the form $\mathbf{E} = [\mathbf{t}]_{\times}\mathbf{R} = \mathbf{R}^T[\mathbf{R}^T\mathbf{t}]_{\times}$ that encapsulates the information about the pose and relates the pair-wise observations [Faugeras and Maybank, 1990, Hartley and Zisserman, 2003]. For the central configuration, the essential matrix is only known up to scale and has therefore five degrees of freedom: three for the 3D rotation $\mathbf{R}$ and two for the translation $\mathbf{t} \in \mathbb{S}^2$ that is forced to belong to the sphere. The so-called epipolar relation constrains the features $(\boldsymbol{f}_{\mathrm{i}}, \boldsymbol{f}'_{\mathrm{i}}) \in \mathbb{R}^3 \times \mathbb{R}^3$ corresponding to the same 3D point as

$$\epsilon_{\mathrm{i}} \doteq \boldsymbol{f}_{\mathrm{i}}^T \mathbf{E} \boldsymbol{f}'_{\mathrm{i}} \tag{3.1}$$

where $\epsilon_{\mathrm{i}} = 0$ if the observations are exactly the projections of the same 3D point and/or if at least one of the observations lays on one of the epipoles $\mathbf{t}$ and $\mathbf{R}^T\mathbf{t} \in \mathbb{S}^2$. The epipolar constraints hold for observations belonging to the image plane, *i.e.* the last coordinate of the 3D vectors $\boldsymbol{f}_{\mathrm{i}}, \boldsymbol{f}'_{\mathrm{i}}$ is one (homogeneous) and when they belong to the image sphere, *i.e.* the observations have unit-norm $\boldsymbol{f}_{\mathrm{i}}, \boldsymbol{f}'_{\mathrm{i}} \in \mathbb{S}^2$. We consider only the last case as it is universal to any camera model and the epipolar constraint and error ϵ_{i} can be also found in the literature with the term *normalized* [Lee and Civera, 2020a].

Since the epipolar constraints are linear in the essential matrix, the minimal approach to the central essential matrix requires only five correspondences in general position [Nistér, 2004, Stewenius et al., 2006, Kukelova and Pajdla, 2007, Kukelova et al., 2008]. These minimal approaches can be integrated into robust frameworks [Kneip and Lynen, 2013, Botterill et al., 2011], *e.g.* RANSAC [Fischler and Bolles, 1981] to detect outliers, although the quality of solution returned by the solver depends on the chosen set of observations and is not necessarily accurate for the other inliers. Nonminimal approaches overcome this problem by considering all the observations while estimating the solution, and for the RRP it is a common practice that these nonminimal solvers minimize the sum of the squared (normalized) epipolar errors $\sum_{i=1}^{N} \epsilon_{\mathrm{i}}^2$ which is quadratic in the unknown essential matrix $\mathbf{E}$ [Hartley and Zisserman, 2003, Sec. 9]. Among the nonminimal solvers we need to highlight the Direct Linear Transformation (DLT) that requires eight or more observations to estimate an unique solution [Hartley and Zisserman, 2003]. However, the returned solution is not necessarily an essential matrix, since the approach does not respect the internal constraints and we need to project solution to the essential matrix space which may lead to a suboptimal solution but it is necessary to obtain a valid pose. As an alternative, [Lui and Drummond, 2013, Ma et al., 2001, Tron and Daniilidis, 2017, Helmke et al., 2007, Dubbelman et al., 2012] propose to refine an initial guess directly on the space of essential matrices,

which turns the problem nonconvex and therefore the quality of the solution cannot be guaranteed.

Certifiable approaches seek to overcome this lack of guarantees by obtaining and certifying the global solution. Most of these approaches solve the problem through a technique that guarantees the optimality of the solution, for instance, Branch-and-Bound (BnB) [Kneip and Lynen, 2013, Hartley and Kahl, 2007b], and convex relaxations [Zhao et al., 2020, Briales et al., 2018, Zhao, 2020]. Whereas the former are more computational extensive since they search the entire domain, both approaches are too slow to solve from scratch. Further, convex relaxations rely on the assumption that the derived and to-be-solved relaxation is *tight*, that is, the relaxation approximates well the original, nonconvex problem. Whereas this assumption holds in practice, at least for moderate noise on the data, in most cases it is not known a priori. Alternative and more efficient approaches lay within the family of *certifiable algorithms* that only certify the solution but do not obtain it, as presented in [Bandeira, 2016]. The driven idea behind these approaches is that iterative algorithms tend to work well in practice, usually when a good initial guess is used, but lack optimality guarantees. Therefore, only certifying the solution is often more efficient than solving globally the relaxation from scratch and can achieve a good ratio of detected optimal solutions, as it has been shown for other problems *e.g.*. rotation averaging [Eriksson et al., 2018], pose graph optimization [Briales and Gonzalez-Jimenez, 2017a, Rosen et al., 2019] or SLAM [Briales and Gonzalez-Jimenez, 2016, Carlone et al., 2015, Carlone and Dellaert, 2015]. The certification of a given solution usually rely on the dual problem, a convex relaxation of the original problem, and assumes that this relaxation is tight. Moreover, deriving an efficient certifier requires the formulation to have a set of specific characteristic in most cases as we elaborate in Section 2.3.4. Whereas these conditions are rather strict, they tend to hold for most problems and the certifiers work well as previously reported in the literature. Additionally and under some conditions, it is possible to further simplify this optimality condition, obtaining what we called a *sufficient optimality condition*, that is sufficient but not necessary for optimality of the solution while it is more efficient to compute than the certifier. Ideally this condition is able to detect and certify most optimal solutions for a wide variety of problem instances, although theoretically fewer solutions than the associated certifier. Such sufficient conditions are shown to be useful in practice for other problems, *e.g.* for rotation averaging [Eriksson et al., 2018], point cloud registration with missing data [Iglesias et al., 2020] or triangulation [Hartley and Seo, 2008]. All these approaches, however, are built upon the assumption that the underlying relaxation is *tight*, that is, the optimal solution of the relaxation can be mapped to a feasible, and therefore, optimal solution for the original, nonconvex problem. Whereas this assumption holds for most problem instances, it may fail for problems with highly noisy data and/or low number of correspondences. To overcome this and make the relaxation tight even for this adversary in-

stances, previous works have shown that introducing additional, redundant constraints on the problem tightens the relaxation, see *e.g.* [Ruiz and Grossmann, 2011, Briales and Gonzalez-Jimenez, 2017b, Zhao et al., 2020, Briales et al., 2018, Yang et al., 2020, Anstreicher and Wolkowicz, 2000]. Nevertheless, finding linearly independent yet redundant constraints is a difficult task on its own, while redundant formulations hinder the development of fast certifiable algorithms and sufficient conditions.

3.2 Contribution

Our contribution to the relative pose problem is threefold. First we propose a fast certifier, upon which we build our second contribution, a sufficient but not necessary condition for optimality. Our third proposal improves the tightness of the previous relaxation with a set of formulations with increasing number of constraints.

First, in [Garcia-Salguero et al., 2021] we propose a fast certifiable algorithm that estimates a potential solution to the problem through an on-manifold optimization and then tries to certify it. This certifier, that relies on the dual problem, is built upon the formulation employed by Zhao in [Zhao, 2020] with 7 constraints and 12 variables. However, the formulation lacks LICQ and thus we propose to remove one of the constraints to overcome this limitation, which allows us to estimate the candidate to Lagrange multipliers in closed-form. This approach, together with the iterative algorithm for the solution, makes our approach efficient, estimating and certifying solutions in under 0.5 milliseconds under a MATLAB implementation. We extend this idea in [Garcia-Salguero and Gonzalez-Jimenez, 2021a] by providing a family of certifiers following the same main idea. In the previous work, one of the constraints has to be removed to achieve the LICQ condition and only one of the combinations was used for certification. In this work we provide and use all the six relaxations obtained by dropping only one of the six constraints that make LICQ failed. Checking the optimality of the solution with all these certifiers increases the number of certified solutions, while still being computed in closed-form, therefore running in 500 microseconds, including variable estimation. In addition, we integrate the certifiable solver into the robust paradigm formed by Graduated Non-convexity (GNC) [Blake and Zisserman, 1987] and the Black-Rangarajan duality between robust functions and line processes [Black and Rangarajan, 1996], which is able to detect and remove outliers. Although GNC is shown to be slower than RANSAC, it tolerates up to 60% of outliers.

Our second contribution in [Garcia-Salguero and Gonzalez-Jimenez, 2021b] consists in a sufficient condition for optimality that guarantees the solution is the global optimum without actually computing the certifier. The condition bounds the minimum eigenvalue of the Hessian and define under which conditions it is nonnegative. We consider the six certifiers proposed in our previous works and derive a common form for all them. The condition, that only depends on the problem data and solution, detect optimal solutions up to four times faster than the associated certifier under a MATLAB implementation. As expected, though, the number of certified solutions is lower although for $N = 40$ correspondences and low to moderate noise (up to one pixel), all the problem instances are certified.

Out third and last contribution in [Garcia-Salguero et al., 2022] leverages four different solvers for the relative pose problem, one of them was already

proposed in [Zhao, 2020] and it's included in the work for comparison purposes. These solvers rely on four different formulations with up to 28 constraints and 15 variables, and we show that each of them achieves different performances in terms of tightness. The proposals solve directly the SDP with SDPTA as solver, hence providing us with a primal solution and its optimality certificate when the relaxation is tight. Concretely, the formulation with 28 constraints remains tight even for highly noisy problem instances, while its reduced size allows us to solve it in less than 7 milliseconds.

3.A Certifiable Relative Pose Estimation

Mercedes Garcia-Salguero, Jesus Briales and Javier Gonzalez-Jimenez

Abstract:
In this paper we present the first fast optimality certifier for the non-minimal version of the Relative Pose problem for calibrated cameras from epipolar constraints. The proposed certifier is based on Lagrangian duality and relies on a novel closed-form expression for dual points. We also leverage an efficient solver that performs local optimization on the manifold of the original problem's non-convex domain. The optimality of the solution is then checked via our novel fast certifier. The extensive conducted experiments demonstrate that, despite its simplicity, this certifiable solver performs excellently on synthetic data, repeatedly attaining the (certified *a posteriori*) optimal solution and shows a satisfactory performance on real data.

CRediT: Conceptualization; investigation; methodology; software; validation; visualization; writing - original draft; writing - review & editing

Journal of Image and Vision Computing, 2021
DOI: https://doi.org/10.1016/j.imavis.2021.104142

3.B Fast and Robust Certifiable Estimation of the Relative Pose Between Two Calibrated Cameras

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:
This work contributes an efficient algorithm to compute the Relative Pose problem (RPp) between calibrated cameras and certify the optimality of the solution, given a set of pair-wise feature correspondences affected by noise and probably corrupted by wrong matches. We propose a family of certifiers that is shown to increase the ratio of detected optimal solutions. This set of certifiers is incorporated into a fast essential matrix estimation pipeline that, given any initial guess for the RPp, refines it iteratively on the product space of 3D rotations and 2-sphere. In addition, this fast certifiable pipeline is integrated into a robust framework that combines Graduated Non-convexity and the Black-Rangarajan duality between robust functions and line processes.
We proved through extensive experiments on synthetic and real data that the proposed framework provides a fast and robust relative pose estimation. We make the code publicly available https://github.com/mergarsal/FastCertRelPose.git.

Journal of Mathematical Imaging and Vision, 2021
DOI: https://doi.org/10.1007/s10851-021-01044-0

3.C A Sufficient Condition of Optimality for the Relative Pose Problem between Cameras

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:

The Relative Pose problem (RPp) seeks for the relative rotation and translation between two central, calibrated cameras given a set of pair-wise feature correspondences. The (RPp) is a fundamental block for many 3D computer vision tasks, and hence, the quality of the estimated relative pose is of key importance for the correct performance of these applications. Nonetheless, the (RPp) is a non-convex problem that presents multiple local minima. Recent non-minimal solvers provide relatively fast certifiable solutions, usually relying on a convex relaxation of the problem; however, there is no guarantee a priori that these relaxations return the optimal solution, i.e., are tight.

This work presents a sufficient condition to guarantee that a given solution of the Relative Pose problem (RPp) is the global optimum, in a faster way than evaluating a certifiable algorithm (up to 4 times faster). We state RPp as an optimization problem that minimizes the squared normalized epipolar error over the set of normalized essential matrices. The proposed condition is derived through spectral analysis and builds up on the recently proposed certifiable algorithm in [Garcia-Salguero et al., 2021]. The results of extensive experiments, with both synthetic and real data, support that by using the proposed conditions we can detect a large number of optimal solutions for most common problem instances.

SIAM Journal on Imaging Sciences, 2021
DOI: https://doi.org/10.1137/21M1397970

3.D A Tighter Relaxation for the Relative Pose Problem Between Cameras

Mercedes Garcia-Salguero, Jesus Briales and Javier Gonzalez-Jimenez

Abstract:
This paper tackles the resolution of the Relative Pose problem (RPp) with optimality guarantees by stating it as an optimization problem over the set of essential matrices that minimizes the squared epipolar error. We relax this non-convex problem with its Shor's relaxation, a convex program that can be solved by off-the-shelf tools. We follow the empirical observation that redundant but independent constraints tighten the relaxation. For that, we leverage equivalent definitions of the set of essential matrices based on the translation vectors between the cameras. Over-constrained characterizations of the set of essential matrices are derived by the combination of these definitions.

Through extensive experiments on synthetic and real data, our proposal is empirically proved to remain tight and to require only 7 milliseconds to be solved even for the over-constrained formulations, finding the optimal solution under a wide variety of configurations, including highly noisy data and outliers. The solver cannot certify the solution only in very extreme cases, e.g., noise 100px and number of pair-wise correspondences under 15. The proposal is thus faster than other over-constrained formulations while being faster than the minimal ones, making it suitable for real-world applications that require optimality certification.

CRediT: Conceptualization; investigation; methodology; software; validation; visualization; writing - original draft; writing - review & editing

Journal of Mathematical Imaging and Vision, 2022
DOI: https://doi.org/10.1007/s10851-022-01085-z

4

Relative Pose problem with prior information between two cameras

Codes are a puzzle. A game, just like any other game

Alan M. Turing

4.1 Introduction

As in Section 3.1 the relative pose problem aims to estimate the relative pose $(\mathbf{R}, \mathbf{t})$ between two cameras given a set of N pair-wise feature observations $(\boldsymbol{f}_\mathrm{i}, \boldsymbol{f}'_\mathrm{i})$. However, this is the general problem and for some applications we have prior information, for example, the 3D points that originated the observations belong to a plane or another sensor provides some information about the relative pose, for example an Inertial Measurement Unit (IMU) allows to derive the gravity vector and to align the cameras to this vector. This section assumes the last case as systems with IMU and cameras have received much attention in recent years, *e.g.* [Zhang et al., 2023, Bloesch et al., 2015] As with the general approach, the gold-standard to the RRp is the bundle-adjustment (BA) optimization that jointly estimates the relative pose and 3D points by minimizing the reprojection error together with the prior information provided by the IMU. However, BA is complex and the optimization may get trapped into local minima, therefore the need for a good initialization. We focus on estimating the relative pose between two cameras considering the information

provided by the IMU without estimating the coordinates of the 3D points, as it was done in Chapter 3.

Knowing the gravity vector allows us to align the cameras to this axis, so that the relative pose between the two cameras has as axis the gravity vector and the only unknown is the angle of rotation, therefore reducing the degrees of freedom of the rotation $\mathbf{R}$ from three to one $\mathbf{R}_{\mathrm{y}}$. In turn this implies that fewer points are required to obtain an unique solution for minimal solver, which greatly simplifies robust algorithms such as RANSAC that depend on the minimum number of points. Apart from being more efficient, introducing the information allows to discard outliers that fulfill a general model, but not the one assumed as we indicate in Section 2.2.2.3. For points in general configuration, we can leverage the essential matrix $\mathbf{E} = [\mathbf{t}]_{\mathrm{x}}\mathbf{R}_{\mathrm{y}}$ that relates the observations through the epipolar constraint $\epsilon_{\mathrm{i}} = \boldsymbol{f}_{\mathrm{i}}^{T}\mathbf{E}\boldsymbol{f}'_{\mathrm{i}}$. For central cameras and despite the prior information, the essential matrix is only known up to scale, and thus the translation belongs to the sphere $\mathbf{t} \in \mathbb{S}^2$, that is, the essential matrix has three DoFs: one for rotation and two for the translation. However, it is known that approaching the relative pose through the essential matrix has some associated degeneracies, among then, the problem is ill-defined when all the 3D points that originated the observations belong a plane. In this case, there exists a family of solutions of essential matrices and it's preferred to state the problem through the 3×3 homography matrix $\mathbf{H}$, which, under the rotation assumption made here has the form $\mathbf{H} = \mathbf{R}_{\mathrm{y}} + \frac{1}{d}\mathbf{t}\mathbf{n}^{T}$, where $\mathbf{n} \in \mathbb{S}^2$ is the normal to the plane and $d \in \mathbb{R}$ is the distance to the plane *w.r.t.* the first camera. Without the prior information, *i.e.*, for generic rotation $\mathbf{R}$, the homography matrix has 8 DoF due to the scale ambiguity but given the gravity prior, $\mathbf{H}$ as defined above has only six DoFs. The homography matrix, as the essential matrix, relates the pair-wise observations $(\boldsymbol{f}_{\mathrm{i}}, \boldsymbol{f}'_{\mathrm{i}})$ as $\boldsymbol{f}'_{\mathrm{i}} \sim \mathbf{H}\boldsymbol{f}_{\mathrm{i}}$, where $\sim$ indicates up-to-scale. To eliminate this scale ambiguity, we take the cross-product $\boldsymbol{f}'_{\mathrm{i}} \times \mathbf{H}\boldsymbol{f}_{\mathrm{i}}$ which gives three expressions, although only two of them are algebraically independent. These expressions $\epsilon_{\mathrm{i}1}$ and $\epsilon_{\mathrm{i}2}$ are treated as the error to be minimized and, similar to the epipolar constraint, they are linearly dependent on the unknown $\mathbf{H}$.

Minimal solvers require three points in correspondence for the essential matrix [Fraundorfer et al., 2010, Ding et al., 2020b, Sweeney et al., 2014a] and also for the homography matrix [Saurer et al., 2016, Ding et al., 2019, Ding et al., 2020a] since for the latter each pair-wise observation provides with two constraints. Other solvers based on the homography matrix assume a particular configuration for the translation and/or plane normal, *e.g.*, ground plane, translation is parallel to plane or vertical plane, [Ding et al., 2020a, Guan et al., 2018, Wadenbäck et al., 2016]. Nevertheless, although these solvers can be integrated into RANSAC to find the subset of inliers, their accuracy depends on the set of inliers selected to estimate the (minimal) solution. Nonminimal solvers, on the other hand, seeks to minimize the residuals (errors) for all the correspondences and the preferred approach is to minimize the normalized squared

epipolar $\sum_{i=1}^{N} \epsilon_{\mathrm{i}}^2$ or homography $\sum_{i=1}^{N} \epsilon_{\mathrm{i}1}^2 + \epsilon_{\mathrm{i}2}^2$ errors, which are quadratic in the unknowns $\mathbf{E}$ and $\mathbf{H}$, respectively. For the approach based on the essential matrix we must highlight the DLT method, as under the form of the rotation $\mathbf{R}_{\mathrm{y}}$, we can modify its coefficient matrix to output the non-zero, non-repeated entries of $\mathbf{E}$, see Section 2.2.2.3. Unfortunately, this cannot be done for the homography matrix as it does not inherit any special structure from the prior information, although it's still possible to use the standard method. However, the output of DLT for both the essential and homography matrix is not necessary a valid matrix with the desired structure and pose, and a projection onto the feasible set is required. Alternatively, we can optimize the above-mentioned cost function on the desired domain, which was the approach followed for the general case in, *e.g.* [Helmke et al., 2007, Ma et al., 2001, Lui and Drummond, 2013, Tron and Daniilidis, 2017, Botterill et al., 2011] for the essential matrix, although to the best of our knowledge there do not exist equivalent solvers for the planar case with homography.

However, the problem is still nonconvex and therefore iterative algorithms cannot guarantee the optimality of the solution. Optimal solvers seek to solve globally this problem, for example, deriving a Gröbner bases and solving the polynomial system as the authors did in [Ding et al., 2021] for the essential matrix. However, due to the numerical inaccuracies, the returned solution is shown to achieve larger costs than iterative methods. As an alternative approach, convex relaxations can be leveraged to obtain an approximation of the solution with a certificate and if the relaxation is tight, they can even retrieved the global solution. However solving these relaxations from scratch is slow, specifically for problems with large number of constraints and variables, and thus certifiable algorithms are preferred for time-sensitive applications. These approaches, nevertheless, rely on the possibility of deriving an efficient certifier, which involves certain characteristic on the formulation such as the number of constraints being lower than the variables, as we explain in Section 2.3.4. Further, formulations with redundant constraints, that empirically remain tight even for problem instances with large noise, hinder the derivation of these certifiers in most cases. These conditions turn to be too strict for the problems at hand, and we need to consider alternatives methods that allows to solve efficiently the certification problem. Of special importance for this section is the algorithm introduced in [Burer and Monteiro, 2005] and further develop in [Boumal et al., 2019, Boumal et al., 2016, Boumal et al., 2020] called the *Riemannian staircase*. The idea behind it is to solve low-rank decompositions of the convex relaxation with increasing rank until convergence is reached or the maximum number of steps is exhausted If the low-rank decompositions have a special structure, for example, a matrix whose columns are orthogonal, but not necessarily its rows, *i.e.*, a point on the Stiefel manifold, then efficient algorithms can be devised based on the Riemannian machinery [Absil et al., 2009, Boumal, 2023b]. The intermediate problems are nonconvex and thus fast certifier are used to check optimality during each step. This approach

has been exploited for large-scale problems, for instance, pose-graph optimization [Rosen et al., 2019, Briales and Gonzalez-Jimenez, 2017a].

4.2 Contribution

We contribute certifiable solvers for the relative pose problem under the assumption that the gravity vector has been provided, for example, by an IMU, considering points in general and in planar configurations.

First, we propose in [Garcia-Salguero and Gonzalez-Jimenez, 2023b] a fast, two-step certifiable solver for the relative pose problem based on the essential matrix with known gravity vector. We estimate the pose through an on-manifold optimization minimizing the squared, normalized epipolar error on the product space of rotation and sphere that allows us to introduce the prior information about the rotation. We then try to certify the solution as the global optimum. Based on our observations in [Garcia-Salguero et al., 2022], we adapt the four different formulations to this problem by considering the pattern of the essential matrix. As with the original form, all the formulations lack LICQ and two of them have more constraints than variables, which blocks closed-form certifiers. To solve the certification, we propose an iterative certifier based on a low-rank decomposition of the Hessian. The proposal runs in milliseconds for all the formulations and certifies more than 90% in both synthetic and real data. Additionally, we observe the rank of the 9×9 Hessian drops to five for all formulations, in contrast with the results obtained by standard off-the-shelf tools that usually attains rank eight.

Motivated by our previous results, our second contribution in [Garcia-Salguero and Gonzalez-Jimenez, a] provides a two-step certifiable solver for the relative pose problem for planar scenes based on the estimation of the homography matrix instead. We estimate the pose and normal of the plane through an on-manifold optimization, minimizing the squared error associated to the homography, see Chapter 4, and then we try to certify the solution as optimal. We exploit four different formulations for this problem with up to 96 constraints and 21 variables, all of them lacking LICQ and two of them having more constraints than variables. We leverage the low-rank iterative certifier from [Garcia-Salguero and Gonzalez-Jimenez, 2023b] and observed a similar performance with more than 90% of solutions certified by all formulations. Additionally, we notice the Hessians for the formulations with 12 variables attain rank eight while the ranks for the formulations with 21 variables depend on the form of the cost, and go to rank 8 or rank 16. On the other hand, solving the same problems under the same cost with off-the-shelf tools returns Hessian with ranks 10 and 19, respectively. Notice that this implies the existance of more than one optimal dual solution.

4.A Fast certifiable relative pose estimation with gravity prior

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:

Redundant and complementary information from different types of sensors boosts the robustness of autonomous systems, making them more reliable and safer. In particular, inertial measurement units (IMUs) are increasingly being integrated with cameras for that purpose, since the information provided by the IMU helps to simplify some visual problems and improves the accuracy of the results. In the context of estimating the motion of a camera, which is the problem we address in this work, the gravity vector delivers by the IMU reduces the unknown rotation to only one degree of freedom instead of three, hence simplifying the relative pose problem (RRp). Despite this simplification, the RRp is still nonconvex, therefore the quality (optimality) of the solution returned by iterative solvers cannot be guaranteed. These suboptimal solutions may have serious consequences for applications that have this solver as a key block, and may even cause their complete failure. In this paper, we contribute a certifiable solver for the RRp with gravity prior. We propose an iterative certifier that does not assume any condition on the problem, and returns an optimality certification even for an overconstrained formulation with 28 constraints in less than 1.5 milliseconds. Since the certifier doesn't obtain the solution to the problem, we also provide a fast, iterative on-manifold estimation of the relative pose, which is shown to return solutions with lower costs than other nonminimal solvers in less time. We make the code available at https://www.github.com/mergarsal

Artificial Intelligence, 2023 DOI:
https://doi.org/10.1016/j.artint.2023.103862

4.B Certifiable Planar Relative Pose Estimation with Gravity Prior

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:

In this work we propose a certifiable solver for the relative pose problem between two calibrated cameras under the assumptions that the unknown 3D points lay on an unknown plane and the axis of rotation is given, *e.g.*by an IMU. The problem is stated in terms of the rotation, translation and plane parameters and solved iteratively by an on-manifold optimization. Since the problem is nonconvex, we then try to certify this solution as the global optimum. For that, we leverage four different definitions for the search space that provide us with different certification capabilities. Since the formulations lack the Linear Independence Constraint Qualification and two of them have more constraints than variables, we cannot derive a closed-form certifier. Instead, we leverage the iterative algorithm proposed in our previous work [Garcia-Salguero and Gonzalez-Jimenez, 2023b] that does not assume any condition on the problem formulation. Our evaluation on synthetic and real data shows that the smaller formulations are enough to certify most of the solutions, whereas the redundant ones certify all of them, including problem instances with highly noisy data. Code can be found in https://github.com/mergarsal.

Computer Vision and Image Understanding, 2024 DOI: https://doi.org/10.1016/j.cviu.2023.103887

5 Absolute Pose problem

You are always a little bit wrong

Hank Green

5.1 Introduction

Estimating the absolute pose $(\mathbf{R}, \mathbf{t})$ between a camera $\{\mathbf{C}\}$ and the world frame $\{\mathbf{W}\}$ given the coordinates of 3D points *w.r.t.* the world frame and their observations on the camera is known as the resectioning or Perspective-n-Point (PnP) problem. The problem is a cornerstone of more complex tasks, such as SLAM or SfM, but it is devised also as the final goal, for example, for a visual odometer where the map is known, see [Aqel et al., 2016] for a review. The resectioning problem can be extended to noncentral cameras, *i.e.*, a rig of M rigidly attached cameras, see Section 2.2.1, which also generalized the problem as a central camera can be seen as the specific case for $M = 1$. Therefore, we consider the noncentral configuration for the rest of this part. Further, to accommodate all camera models, even those without an image plane see Section 2.2.1, we let the observations $\boldsymbol{f}_\mathrm{i}$ have norm one, that is, a point on the sphere $\mathbb{S}^2$, hence working with the image sphere. With these, we have that for the general case the i-th observation $\boldsymbol{f}_\mathrm{i} \in \mathbb{S}^2$ associated with the 3D point $\mathbf{P}_\mathrm{i} \in \mathbb{R}^3$ *w.r.t.* the world frame $\{\mathbf{W}\}$ is captured by the j-th camera with pose $(\mathbf{R}_\mathrm{j}, \mathbf{c}_\mathrm{j})$ *w.r.t.* the camera center $\{\mathbf{C}\}$. The PnP problem seeks the absolute

pose $(\mathbf{R}, \mathbf{t})$ that transform each point $\mathbf{P}_\mathrm{i}$ as $\mathbf{Q}_\mathrm{i} = \mathbf{R}_\mathrm{j}^T(\mathbf{R}\mathbf{P}_\mathrm{i} + \mathbf{t} - \mathbf{c}_\mathrm{j}) \in \mathbb{R}^3$ on the camera frame to match the observation $\boldsymbol{f}_\mathrm{i}$.

Previous works have tackled this problem by either minimizing the reprojection error *e.g.*, [Cifuentes, 2021, Yang et al., 2011], which makes the cost rational, or through the point-line error, *e.g.* [Lourakis and Terzakis, 2021] making the error quadratic in the unknown pose. The first case requires as well to select which metric is used to measure distance, for both the image plane and sphere, see Section 2.2.1. In this thesis, we avoid rational errors as they hinder the development of efficient certifiable algorithms and opt for the point-line error as it is already quadratic in the unknowns. In this case, the observation is seen as the ray emanating from the camera center towards the observation, that is, it is considered a line in the space and instead of the reprojection, the error measures the distance (perpendicular) between the 3D point and the observation (line). Another interesting point for this approach is that the literature for registration problems is wide, and previous approaches have relied on Gröbner-basis methods *e.g.* [Malis, 2023, Zhou et al., 2020, Wientapper et al., 2018], Branch-and-Bound methods *e.g.* [Olsson and Eriksson, 2008] or convex relaxations as SDP solvers *e.g.* [Briales and Gonzalez-Jimenez, 2017b].

Minimal solvers require three points to estimate an unique solution for both the central [Kneip et al., 2011, Ke and Roumeliotis, 2017, Wang et al., 2018, Gao et al., 2003] and noncentral [Nistér and Stewénius, 2007, Miraldo and Araujo, 2014, Ventura et al., 2014], and can be used to detect outliers, for instance, with RANSAC. However, for noisy data these solutions may be skew and its preferred to use nonminimal solvers that leverage all the correspondences to estimate a solution. Since one of the unknowns is a rotation, previous works have exploited minimal parameterization for it in order to solve the problem *e.g.* [Hesch and Roumeliotis, 2011]. Nonetheless, these characterizations are affected by singularities [Cayley, 1846] but fortunately, rotations are a well-known domain and there exists alternative approaches that respect the internal structure while being efficient. Previous works [Zhou and Kaess, 2019, Campos et al., 2019, Schweighofer and Pinz, 2006] state the problem as to respect the domain, although due to the non-convexity of the problem [Schweighofer and Pinz, 2006] the proposals cannot guarantee the optimality of the retrieved solution. Other approaches [Hesch and Roumeliotis, 2011, Sweeney et al., 2014b, Zheng et al., 2013, Wu et al., 2022, Kneip et al., 2014, Kneip et al., 2013] obtain the solutions to a polynomial system, and keep the one with lower cost, *i.e.*, the optimal solution. Despite being optimal in theory, these methods may suffer from numerical instabilities [Byröd et al., 2009]. Another approach that allows to estimate and certify, under some conditions, the optimality of the solution relies on convex relaxations of the original problem and Lasserre's hierarchy [Lasserre, 2001], as it was done in *e.g.* [Agostinho et al., 2022, Schweighofer and Pinz, 2008, Sun and Deng, 2020].

Nevertheless, the off-the-shelf tools employed to solve these relaxations have an upper limit in the number of constraints and variables, while being

too computational expensive even for smaller problems. Previous works have tried to overcome this by relaxing the domain from the rotation set to the orthogonal set, hence allowing reflections, *i.e.*, matrices with determinant -1, unblocking certifiers that estimate the candidate to Lagrange multipliers in closed-form. Whereas this approach has been shown to work well for most problem instances, we observe its failure for some instances of the PnP problem, therefore our interest in providing an efficient certifiable algorithm for the original formulation with a rotation. The PnP problem can be stated only in terms of the rotation, as the unconstrained translation can be marginalized as it was done in [Briales and Gonzalez-Jimenez, 2017b]. This reduces the number of variables of the problem and the final form only involves the 3D rotation matrix $\mathbf{R}$, formally

$$f^\star = \min_{\mathbf{R}\in\mathbb{R}^{3\times 3}} \operatorname{vec}(\mathbf{R})^T \mathbf{C} \operatorname{vec}(\mathbf{R}) + 2\mathbf{c}^T \operatorname{vec}(\mathbf{R}) + c \text{ subject to } \mathbf{R} \in \mathbb{SO}(3) \tag{PROB-R}$$

where $\mathbf{C} \in \mathbb{S}^9, \mathbf{c} \in \mathbb{R}^9$ and $c \in \mathbb{R}$ depend on the camera configuration (central or noncentral) and the observation-point correspondences. Notice that the formulation in Equation (PROB-R) matches other problems, for example point-point and point-plane registration as it was shown in *e.g.* [Olsson et al., 2008, Briales and Gonzalez-Jimenez, 2017b]. The problem is nonconvex due to the constraints required by the rotation, and whereas the number of variables of the problem is nine, the number of constraints depend on which definition we use for the rotation space $\mathbb{SO}(3)$. As we commented above, previous works have relaxed the problem to the orthogonal set, hence retaining six constraints, making either the columns or rows orthonormal. However, as shown in previous works, *e.g.* [Briales and Gonzalez-Jimenez, 2017b, Garcia-Salguero et al., 2022, Zhao et al., 2020], introducing additional constraints into the problem tend to tighten the relaxation, and in terms of certification, it may increase the number of detected global optima.

5.2 Contribution

We contribute a certifiable algorithm in [Garcia-Salguero et al.,] for the resectioning problem for central and noncentral cameras that obtains and certifies the optimal solution even for problems with large noise. While we devised our proposal for the PnP problem, the performance of the algorithm on random data motivates its use for other problems that estimate rotations, which are common in computer vision and robotics.

We minimize the point-line error, that is, the perpendicular distance between the 3D point and the ray associated to the observation and simplify the formulation to only involve the rotation as in Equation (PROB-R). The first step of the proposal estimates a solution through an on-manifold optimization, which may be a suboptimal solution. The second step tries to certify it as the global optimum, and for that we rely on formulations of the problem with the standard definitions of the rotation space. The constraints for orthogonal rows and columns, each with 6 expressions, can be joined to create another (redundant) set with 11 constraints as one of them is linearly dependent. The positive determinant requirement is enforced by nine additional constraints, and thus the final redundant formulation has 20 constraints. The formulations with redundant constraints fail the LICQ condition while having more constraints than variables. This precludes the development of efficient, closed-form certifiers, and we need to consider alternative approaches for certification. To solve the certification in an efficient manner, we reformulate it as an eigenvalue problem, *e.g.* [Shapiro and Fan, 1995, Lewis and Wylie, 2019] that maximizes the minimum eigenvalue(s) of the Hessian. The certification is convex but in general nonsmooth if the number of zero eigenvalues is larger than one. Empirically we observe this to be the case even for random data, although applying a parameter $\alpha = 2$ during the optimization reduces this number to one, making the optimization also smooth. While the number of zero eigenvalues does not hinder the performance of the proposal, it shows the existence of more than one dual optimal solution. The resulting certifier handles the redundant formulations and returns positive certificates even for highly noisy data in microseconds, while requiring only a library for linear algebra for its implementation.

5.A Fast certifiable algorithm for the absolute pose estimation of a camera

Mercedes Garcia-Salguero, Elijs Dima, André Mateus and Javier Gonzalez-Jimenez

Abstract:
Estimating the absolute pose of a camera given a set of N points and their observations is known as the resectioning or Perspective-n-Point (PnP) problem. It is at the core of most computer vision applications and it can be stated as an instance of 3D registration with point-line distances, making the error quadratic in the unknown pose. The PnP problem, though, is nonconvex due to the constraints associated with the rotation, and iterative algorithms may get trapped into any suboptimal solutions without notice. This work proposes an efficient certification algorithm for central and noncentral cameras that either confirms the optimality of a solution or is inconclusive. We exploit different sets of constraints for the rotation to assess their performance in terms of certification. Two of the formulations lack the Linear Independence Constraint Qualification (LICQ) while one of them has more constraints than variables. This hinders the usage of the standard procedure which estimates the Lagrange multipliers in closed-form. To overcome that, we formulate the certification as an eigenvalue optimization and solve it through a line-search method. Our evaluation on synthetic and real data shows that minimal formulations certify most solutions (more than 90% on real data) whereas redundant formulations are able to certify all of them and even random problem instances. The proposed algorithm runs in microseconds for all these formulations.

CRediT: Conceptualization; investigation; methodology; software; validation; visualization; writing - original draft; writing - review & editing

Submitted (2023)

Part II

Triangulation

6 Euclidean Triangulation

Still, no gardener would be a gardener if he did not live in hope.

Vita Sackville-West

6.1 Introduction

Reconstructing the 3D points of a scene given their observations and the poses of the associated cameras is known as the triangulation problem. In this context, the triangulation provides a sparse point cloud, see Figure 6.1, as the reconstructed points are associated with relevant features, for example corners. The scene can be therefore difficult to distinguish at plain sight, as some zones may lack characteristics features and being absent from the reconstruction output. However, even the sparse cloud turns out to be relevant for some applications, for example, visual odometry, while being more efficient to estimate and store as only a finite number of points have to be considered. In the rest of this section, we focus on the retrieval of this sparse point cloud under different configurations and number of cameras.

The 3D coordinates of the point $\mathbf{Q}$ are obtained through the linear method explained in Section 2.2.3 from a set of N observations $\boldsymbol{f}_{\mathrm{i}}$ and camera poses $\mathbf{P}_i$. This approach requires the decomposition of the matrix formed by the data to obtain the solution, which we consider to be exact if the least singularvalue

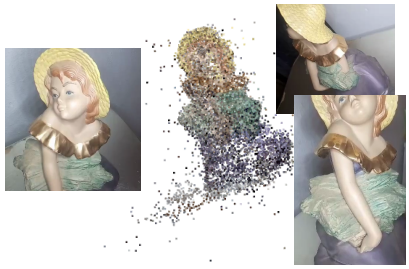


Figure 6.1: Sparse point cloud generated from a set of images.

of the matrix is zero up to some tolerance. Here, the term "exact" implies that the 3D point that originates the observations exists, and so the rays emanating from each camera center towards the corresponding observations meet at a single 3D point. This condition, however, only holds for noiseless observations and in practice, the rays do not meet in a single point, even when two views $N = 2$ are considered. The triangulation topic focuses on these cases and how to obtain the 3D point. A common approach is the midpoint method, that, in its basic form, takes the geometric midpoint from each ray [Hartley and Zisserman, 2003]. This approach achieves good results, specially if some weights are considered [Yang et al., 2019] and/or the camera poses are not accurate. However, the so-called "optimal" approach for the triangulation problem consists in correcting the observations so that the linear method can retrieve the 3D point exactly. The correction of the observations is sought to be minimal, leading to different solvers depending on which metric is measuring distance, *i.e.* [Lee and Civera, 2019a, Hartley and Sturm, 1997, Oliensis, 2002, Chum et al., 2005] for fewer than four views and [Hartley and Kahl, 2007a, Lee and Civera, 2020b, Aholt et al., 2012, Chen et al., 2020] for N-view.

The correction, however, requires the corrected observations to be originated by a single point, which introduces some constraints on the problem. As it was shown in [Cifuentes, 2021] the N-view triangulation can be tackle with the rational constraint, although the number of constraints limit the number of views that can be considered. These constraints, however, can be substituted in some cases. The concept of multiview ideal or joint image, *e.g.* [Agarwal et al., 2019, Trager et al., 2015] explains the underlying relations between observations of the scene points and the camera poses. For the triangulation problem our interest focuses on which relations need to be fulfilled by the observations so the 3D point is real. In those references and also empirically in [Aholt et al., 2012] the authors conclude that: (1) for two views, the relation rising from the epipolar matrix $\mathbf{E}$ (either the essential or fundamental matrix) is sufficient; (2) for three views, the relation from the trifocal tensor must be included; and (3) for four or more views, the relation from all the combinations of two views are

sufficient, provided some conditions holds. The last conclusion simplifies the formulation of the triangulation problem as the epipolar relation $\boldsymbol{f}_{\mathrm{i}}^T \mathbf{E} \boldsymbol{f}'_{\mathrm{i}} = 0$ is bilinear on the unknowns. However, theoretically the relation only holds if the cameras are not all coplanar (including linear) and the observations lay far from the epipoles. Assuming these conditions, we can correct the observations so that the $\binom{N}{2} = N(N-1)/2$ epipolar constraints are fulfilled, which guarantees the retrieved 3D point exists. Nevertheless, for N large the number of constraints cannot be handled by off-the-shelf tools, as it was shown in [Aholt et al., 2012].

On the other hand, these relations define a general scene, and additional information, for example the point belonging to a plane, is not necessary included and will not be fulfilled in general. Instead and for the specific case of planar scenes, the homography matrix $\mathbf{H}$ is used as it induces a point-to-point relation between the observations $\boldsymbol{f}_{\mathrm{i}}, \boldsymbol{f}'_{\mathrm{i}}$ with $\boldsymbol{f}_{\mathrm{i}} \sim \mathbf{H}\boldsymbol{f}'_{\mathrm{i}}$ where $\sim$ indicates equality up-to-scale. The concept behind the optimal triangulation approach is the same, though, and we seek the correction $\Delta\boldsymbol{f}_{\mathrm{i}}, \Delta\boldsymbol{f}'_{\mathrm{i}}$ with minimum norm that makes the equality exact. To remove the scale ambiguity from the constraint $\boldsymbol{f}_{\mathrm{i}} \sim \mathbf{H}\boldsymbol{f}'_{\mathrm{i}}$, we take the cross-product as $\boldsymbol{f}_{\mathrm{i}} \times \mathbf{H}\boldsymbol{f}'_{\mathrm{i}} = \mathbf{0}_{3\times 1}$ which gives three bilinear expressions in the unknowns. However, only two are algebraically independent, *i.e.*, each pair of observations contributes with two constraints to the problem. If the observations can be limited to the image plane, then the 2-view problem has four variables and two constraints. This is the approach considered by Chum *etal.* in [Chum et al., 2005] which they solved through a Gröbner basis. Extending this case to N-view implies the study of how many constraints one needs to fully defined the space, that is, the returned solution is *guaranteed* to retrieve a 3D point. The minimal, general graph is the linear one where each image (node) is related (link) only to the next image (node) thus providing $N-1$ relations and $2(N-1)$ constraints for N-views and $2N$ variables. On the other hand, we can also leverage the over-constrained problem with all the combinations $\binom{N}{2} = N(N-1)/2$ and thus, $2\binom{N}{2} = N(N-1)$ constraints for $2N$ variables. As for the general case, problems with N large reach the limits of current off-the-shelf solvers and alternative approaches should be used.

6.2 Contribution

Our contribution for the triangulation problem is three-fold and focuses on planar and general scenes for two and N views.

First, in [Garcia-Salguero and Gonzalez-Jimenez, 2022] we tackle the 2-view triangulation problem assuming the unknown 3D point belongs to a plane, therefore relying on the homography matrix to relate the observations. We formulate the triangulation problem as the minimization of the norm-2 of the correction such that the corrected observations fulfill the two constraints provided by the homography matrix. Since the cost and constraints are quadratic, we derive a QCQP with four variables and two constraints. The problem is close to the formulation with one constraint that is tackled in [Hmam, 2010] and motivate by their results, which we summarize in Section 2.3.3.2, we opt for applying the approach to this case. In summary, Hmam proposed a primal-dual method based on the dual problem where convergence implies optimality as, upon algebraic manipulation, the problem is reduced to a zero-finding algorithm in one variable. For the problem with one constraint, the existence of a real zero can be proved, implying the existence of the optimality certificate and the fulfillment of the strong duality condition. Whereas this is a known fact for problems with one constraint, see [Boyd and Vandenberghe, 2004, App. B.1] and Section 2.3.3, for the general problem with two constraints this cannot be guaranteed a priori. However, the algorithm can be applied to the 2-view planar triangulation problem in a straightforward manner. Following this approach, we derive a primal-dual algorithm where convergence also implies optimality and strong duality. Our evaluation shows a great performance in terms of certification and computational time. However, and although this primal-dual approach works well, we further extend our proposal with an optimality certifier that does not estimate the solution. Since the primal solution can be obtained by any means, we believe the usage of the certifier alone is also relevant. For that, we leverage once again the dual problem and derive the certifier with the standard procedure. The formulation fulfills LICQ and has fewer constraints than variables, allowing us to derive a closed-form expression for the candidate to Lagrange multipliers. The algorithm is shown to certify all solutions in a fraction of time. Our last contribution is a sufficient condition for optimality, as in Section 3.2, which is only sufficient but not necessary. Despite being by construction looser than the certifier, that is, in general it detects fewer optimal solutions than the certifier, it is faster to estimate and is a good alternative for time critical applications. The condition relies of bounds of the minimum eigenvalue of the Hessian, and empirically, we observe that it detects most optimal solutions while running in microseconds.

Our second contribution in [Garcia-Salguero and Gonzalez-Jimenez, 2023a] focuses on the N-view triangulation problem for general scenes without assuming any condition on the 3D points. We pose the triangulation problem as the norm-2 correction of the observations such that the constraints, given

by the $\binom{N}{2}$ epipolar relations are fulfilled. Since the number of constraints quickly reaches the limit of current off-the-shelf solvers, we propose a two step certifiable algorithm that is implemented with the library for linear algebra EIGEN [Guennebaud et al., 2010]. The first stage is an iterative algorithm that estimates a local solution of the problem based on Newton's method for multivariate problems. For the second stage, the certification of this solution, we observe that LICQ is not fulfilled and the number of constraint is greater than the number of variables for $N > 4$. However, our evaluation shows that the Lagrange multipliers with minimum-norm are in fact the optimal dual solution, *i.e.*, the associated Hessian is PSD. This observation allows us to propose a closed-form certifier, that is also implemented with EIGEN, and therefore the full proposal does not require any additional specific tool. Empirically, the proposal estimates and certifies more than 99% of the solutions while surpassing the limits of current approaches [Aholt et al., 2012, Cifuentes, 2021]. We also report that the coplanar condition that will make the formulation fail in theory only holds up to some noise, which is further supported by the great performance of the proposal on CORRIDOR [1], a real sequence where the movement is linear.

Our third and last contribution in [Garcia-Salguero and Gonzalez-Jimenez, b], motivated by the previous works, tackles the N-view planar triangulation problem. In this case, the relation between two corresponding observations is given by the homography matrix. An important first observation is that it suffices to consider $N - 1$ links between images, contrary to the $\binom{N}{2}$ required by the general case. Therefore, it is simplest form the problem has $2N$ variables and $2(N - 1)$ constraints. As with the general N-view, we rely on a two-step certifiable algorithm and use the same approach. For the minimal formulation, the candidate to Lagrange multipliers is estimated in closed-form since LICQ holds and we have fewer constraints than variables. Our evaluation shows an excellent performance of this approach, except in a few cases on real data. For those cases, we include all $\binom{N}{2}$ relations, leading to $N(N - 1)$ constraints. Despite the size of the problem, our proposal is able to handle it and returns the global, certified solution in seconds. However, under this redundant formulation LICQ fails and the number of constraints is larger than the number of variables. Interesting enough, we observe in this case as well that the Lagrange multipliers with minimum-norm are the optimal dual solution with PSD Hessian, which allows us to estimate the certifier in closed-form even for this redundant formulation. We observe that the problem instances on real data that couldn't be certified by the first approach are certified by this formulation, that is, the solution returned by the formulation with $2(N - 1)$ constraints was actually the global optimum and the associated certifier fails in those cases. Since we use the same approach than in [Garcia-Salguero and

[1] https://www.robots.ox.ac.uk/~vgg/data/mview/

Gonzalez-Jimenez, 2023a], this problem can be solved without specific tools relying only on Eigen.

6.A Certifiable algorithms for the two-view planar triangulation problem

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:
Planar scenes predominate in man-made environments, *e.g.*interior or facades of buildings and in ground images from aerial vehicles. Points lying on those surfaces can be reconstructed from their observations in two images. However, generic reconstruction algorithms output 3D points not lying on the plane, thus obtaining inaccurate reconstructions. The problem also turns to be non-convex with many local minima, hence hindering the performance of iterative method. Therefore, being able to obtain *and* certify the optimal solution to this problem is of special relevant for these applications. In this paper we first propose a fast and certifiable algorithm that both estimates and certifies the optimal solution to the triangulation problem. From this formulation, we also present an optimality certificate that tells us whether a given solution (obtained by any solver) is the global optimum. Last, from this certificate we derive a sufficient (but not necessary) optimality condition that allows us to certify optimality in less than one microsecond. We test the proposed algorithms on extensive experiments on both synthetic and real data. Code is made available at https://github.com/mergarsal.

Computer Vision and Image Understanding, 2022. DOI: https://doi.org/10.1016/j.cviu.2022.103570

6.B Certifiable solver for real-time N-view triangulation

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:

Cutting-edge field robotic systems, such as UAV or autonomous cars, demand fast and optimal solutions for any component at the core of their critical navigational tasks. Among them, we focus on the triangulation of image points from multiple views, which is a cornerstone for more complex tasks such as visual localization and SLAM. In this paper we present a fast and certifiable solver for the N-view triangulation problem that doesn't require any specific optimization software package and can be implemented with any linear algebra library. The proposal relies on a series of linear convexifications which, in the limit, recovers the original problem, allowing us to solve problem instances with $N = 10$ views in 150 microseconds on a standard desktop computer. On real data our solver obtains and certifies the optimal solution in more than 99% of the problem instances. We make the code available at https://github.com/mergarsal.

IEEE Robotics and Automation Letters (RA-L), 2023 DOI: 10.1109/LRA.2023.3245408

6.C A fast certifiable algorithm for the N-view planar triangulation

Mercedes Garcia-Salguero and Javier Gonzalez-Jimenez

Abstract:

Planes are ubiquitous in man-made environments: facades, walls, floors, tables, etc. The visual reconstruction of planar primitives is accomplished by the triangulation of feature points from multiple views. However in the presence of noise these features don't originate a single 3D point and the "optimal" approach, which under the ℓ_2 is non-convex in general, is to correct these observations. In this paper, we contribute for this problem a fast iterative solver accompanied by an optimality certifier that can handle large-scale problems in milliseconds. The proposal builds upon operations available in any algebra library and is particularly suitable for real-time applications that have very limited computational resources available. Our evaluation on real data shows that the proposal estimates a feasible solution for all problem instances, whereas the certifier certifies more than 96.67% solutions.

Moreover, for applications with no severe speed limitations, we extend the above formulation to achieve the certification of all (100%) the solutions by incorporating a much higher number of (redundant) constraints in the problem. This comes at the expense of more computational cost (around six milliseconds for $N = 30$ views). We make the code available at https://www.github.com/mergarsal.

Submitted (2023)

7

Conclusions and future works

This thesis has proposed different certifiable algorithms for two of the main problems in visual-based pipelines: the pose and the point cloud (triangulation) estimation under different assumptions of the problem and/or 3D point distribution. In spite of being hot topics of research during the past decades, the common approach for these tasks states them as nonconvex problems, thus having in general more than one local minima, and leverages iterative algorithms to solve these formulations. Due to the iterative nature of these approaches, the algorithms cannot guarantee that the returned solution is the global optimum, although empirically it has been observed that they perform well and output good solutions, specially when the initial guesses are also good enough. Nevertheless, the quality of these solutions may affect other blocks/tasks of the pipeline and the errors may cascade into the final solution of the application. Hence, being able to guarantee the optimality of each task acquires special importance for the safe operation and reliability of the system. The so-called global methods estimate the global optimum with guarantees, although the standard tools may be computational expensive, making them unsuitable for real-time applications in most cases. On the other hand, the so-called certifiable algorithms only certify solutions but do not estimate them. These methods are motivated by the observation that iterative algorithms tend to perform well, despite not providing guarantees of the quality of the solution. Hence, only certifying the solution is usually faster, although the certification may fail for different reasons and the approach requires another algorithm to estimate a (local) solution of the problem.

In this thesis we focus on global methods and specifically, efficient certifiable algorithms. We tackle problems that can be formulated as the minimization of a quadratic cost subject to a set of quadratic constraints, allowing us to leverage two well-known convex relaxations: the dual problem and Shor's

relaxation. These relaxation can be solved from scratch by off-the-shelf tools with polynomial complexity in the number of variables and constraints, although for some specific formulations more efficient approaches as the low-rank decomposition in [Burer and Monteiro, 2005] can be leveraged. Moreover fast certifier can be derived from these relaxations although they come with some limitations, namely the number of constraints must be smaller than the number of variables and the constraints must fulfill the so-called Linear Independence Constraint Qualification (LICQ) that involves the Jacobian of the constraints. These two conditions are shown to be too strict for the common computer vision problems tackled in this thesis, hence motivating the research of alternative certifiable approaches.

The contributions of this thesis are grouped into two topics: pose and triangulation. Next, we summarize the proposed methods and their drawbacks.

Pose estimation: The first part of this thesis focuses on pose estimation problems, both relative (between two cameras) and absolute (between a camera and a fixed frame). These problems are the cornerstone of more complex pipelines and are often the first task to be solved, although even the basic formulations suffer from local minima. We start with the the relative pose problem between two cameras and contribute a set of closed-form optimality certifiers in [Garcia-Salguero et al., 2021, Garcia-Salguero and Gonzalez-Jimenez, 2021a]. Whereas these algorithms allow to certify solutions in an efficient manner, they relax the original formulation before deriving the dual problem, which is also a relaxation. This implies that failure of the certification may come from the solution being suboptimal and/or any of the relaxations not being tight. However, in practice the algorithms work well, specially for problems without highly noisy observations. This performance motivates our work in [Garcia-Salguero and Gonzalez-Jimenez, 2021b] in which we derive a sufficient condition for optimality that does not require to obtain the certificate from scratch. Thus, the approach is faster than the previous certifiers and empirically is able to detect most optimal solutions. Nevertheless, as it is a relaxation of the original relaxations, the ratio of certified solutions is lower than the original approaches. We overcome the limitations of the previous works by proposing in [Garcia-Salguero et al., 2022] a set of formulations whose relaxations are tighter but have to be, unfortunately, solved with off-the-shelf tools. Although this approach is slower than the previous works, our empirically evaluation shows that the redundant formulations can estimate and certify problem instances with highly noisy observations.

Based on these results, our next work in [Garcia-Salguero and Gonzalez-Jimenez, 2023b] leverages these redundant formulations for the configuration where the axis of rotation is known, for example, from an IMU. The original constraints can be adapted to this case, and their performance in terms of certification is similar to the original case. We go one step further and propose as well a fast iterative certifier that can handle these formulations that have

redundant constraints and lack of LICQ. Our evaluation shows that this algorithm certifies most solutions while being faster than solving the relaxation from scratch as in [Garcia-Salguero et al., 2022]. However, the certifier requires an on-manifold optimization and therefore a specific library, which may not be suitable for devices with memory restrictions. This approach, however, is shown to perform well even for other formulations, and in our next work in [Garcia-Salguero and Gonzalez-Jimenez, a] we consider the relative pose between cameras with known axis of rotation, but in this case, the homography matrix relates the observations as they are all originated from 3D points on a plane. We provide four different formulations that have redundant constraints and lack of LICQ and leverage the previous iterative certifier to certify solutions efficiently. Our evaluation shows similar results whereas we observe different performances of the certifier with the type of cost matrix. The drawback of this approach remains in the tools required to solve the certification, and our results about the cost open the question about the reason behind this behavior and whether it can be seen and used in other problems.

Our last work [Garcia-Salguero et al.,] in the topic of camera pose proposes a certifier for the resectioning, PnP or absolute pose problem for central and noncentral cameras. The problem is stated in its quadratic form for both cases and is re-arranged to depend only on the 3D rotation matrix of the pose. As with our previous works, the redundant formulation for this problem has more constraints than variables and lack LICQ. Our iterative certifier in [Garcia-Salguero and Gonzalez-Jimenez, 2023b] shows Hessians with deficient rank although it is slow due to the bad initialization. With this in mind, we propose another iterative certifier that states the problem as an unconstrained eigenvalue optimization. The problem, which is in general convex but nonsmooth, can be solved with any library for linear algebra. Empirically the proposal can estimate and certify solutions even for random problem instances, yet it may be difficult a priori to extend the idea to more than one rotation matrix.

Point estimation: The second part of our contribution focuses on the triangulation problem through so-called "optimal" approach in which the observations are corrected to triangulate an unique 3D point. First, our work in [Garcia-Salguero and Gonzalez-Jimenez, 2022] tackles this problem for the planar case with two views and seeks the correction for the observations that fulfills the homography constraint. For this problem, we propose a primal-dual algorithm that estimates the solution where convergence implies optimality. Based on this same formulation of the problem, we propose a closed-form optimality certifier that is faster than the primal-dual algorithm but does not estimate the solution. We further expand this by providing a sufficient condition for optimality, which is faster than the certifier but fails for some problem instances. Our proposals, however, are tailored to the 2-view triangulation problem and cannot be transferred to the N-view configuration in a straightforward manner.

Our second work in [Garcia-Salguero and Gonzalez-Jimenez, 2023a] is motivated by this drawback, and we tackle the N-view triangulation but for a generic configuration of the points, that is, not all the points lay on a plane. We rely on the formulation in [Aholt et al., 2012] that leverages the essential or fundamental matrices instead of the projection matrices to constrain the corrected observations. This approach makes the restriction of the problem quadratic in the unknowns, but requires $\binom{N}{2}$ constraints for N views, which quickly reaches the limit of off-the-shelf tools. To avoid these tools, we propose a two-step certifiable algorithm that first estimates the solution through Newton's method and then, tries to certify it as the global optimum. As the formulation has more constraints than variables and also lacks LICQ for more than four views, a closed-form certifier cannot be derived. However, empirically we observe that the minimum-norm solution for the Lagrange multipliers actually certifies the solution, observation which is backed up by our extensive evaluation on synthetic and real data. Nonetheless, this closed-form certifier is just an empirical observation and a rigorous proof of its suitably has not been found yet. Our last contribution in [Garcia-Salguero and Gonzalez-Jimenez, b] builds upon the previous two works and tackles the planar, N-view triangulation. For this case, a minimally-constrained formulation can be used, although it fails to remain tight for some problem instances. Fortunately, the redundant formulation with all the available $2\binom{N}{2}$ constraints success at this to the expense of computational time due to the number restrictions. As for the generic case, the two-step algorithm with the combination of Newton's method and the optimality certifier works well in synthetic and real data, whereas the minimum-norm solution for the Lagrange multipliers for the formulation with redundant constraints also returns positive certifications. As with the previous work, this behavior is still an empirical observation, but also motivates further research about it and opens the question about other problems showing the same characteristic.

Future work

We conclude this thesis with two potential lines of research we consider of special relevance. Both topics are complimentary and could be used jointly as an alternative to the off-the-shelf tools employed to solve convex relaxations as the ones leveraged in this work. Despite the unquestionable relevance of these tools, current performance requirements call for more efficient and maybe specific algorithms.

Efficient certifiers for non-minimally constrained formulations: As we have seen through this thesis, formulations lacking LICQ and with more constraints than variables appear in more problems than expected, while tend to be the tightest relaxations. Whereas the off-the-shelf tools can be used to solve globally these formulations, the lack of fast certifiers is a major drawback, specially for real-world applications. Deriving new (or even specific) methods and algorithms to certify these problems, independently of the lack of LICQ

and the number of constraints, is essential for the introduction of optimality certificates in common applications. Moreover, making these certifiers efficient but also easy to implement without requiring additional libraries may expand the range of devices that can leverage them. Another line of work should be centered at the initial guess for these certifiers, specially those initializations that can be estimated in closed-form. Not only good initial guesses will speed up the convergence if the certifier is iterative, but it can also shred some light onto why some algorithms may not be suitable for some problems.

Primal solution extraction from a failed certificate: The first line of research tries to certify a given solution despite the conditions of the formulation. Whereas the certification may fail due to the underlying relaxation being not tight, it can also happen when the solution is not the global optimum. This second line of research tries to work with this information, and instead of finding a new solution, it tries to find a primal solution from the failed certificate that can be used as initial guess for iterative, primal solvers. Whereas this process can be relatively easy to solve when the nullspace of the Hessian is one-dimensional, as it has been done in the literature, the problem gets more complex when this is not the case, and at least one feasible solution must be retrieved from a n-dimensional space which may or may not contain a feasible solution. However, solving this will provide with strong initial guesses for the primal iterative algorithms, and, when pair with the certifier, it will unblock a primal-dual algorithm robust to lack of constraint qualifications and redundant constraints.

And with that, best of luck.

Bibliography

[Absil et al., 2009] Absil, P.-A., Mahony, R., and Sepulchre, R. (2009). *Optimization algorithms on matrix manifolds*. Princeton University Press.

[Agarwal et al., 2019] Agarwal, S., Pryhuber, A., and Thomas, R. R. (2019). Ideals of the multiview variety. *IEEE transactions on pattern analysis and machine intelligence*.

[Agostinho et al., 2022] Agostinho, S., Gomes, J., and Del Bue, A. (2022). Cvxpnpl: A unified convex solution to the absolute pose estimation problem from point and line correspondences. pages 1–21.

[Aholt et al., 2012] Aholt, C., Agarwal, S., and Thomas, R. (2012). A qcqp approach to triangulation. In *European Conference on Computer Vision*, pages 654–667. Springer.

[Alizadeh et al., 1997] Alizadeh, F., Haeberly, J.-P. A., and Overton, M. L. (1997). Complementarity and nondegeneracy in semidefinite programming. *Mathematical programming*, 77(1):111–128.

[Anstreicher and Wolkowicz, 2000] Anstreicher, K. and Wolkowicz, H. (2000). On lagrangian relaxation of quadratic matrix constraints. *SIAM Journal on Matrix Analysis and Applications*, 22(1):41–55.

[ApS, 2019] ApS, M. (2019). *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*

[Aqel et al., 2016] Aqel, M. O., Marhaban, M. H., Saripan, M. I., and Ismail, N. B. (2016). Review of visual odometry: types, approaches, challenges, and applications. *SpringerPlus*, 5:1–26.

[Azzam et al., 2020] Azzam, R., Taha, T., Huang, S., and Zweiri, Y. (2020). Feature-based visual simultaneous localization and mapping: a survey. *SN Applied Sciences*, 2(2):1–24.

[Bandeira, 2016] Bandeira, A. S. (2016). A note on probably certifiably correct algorithms. *Comptes Rendus Mathematique*, 354(3):329–333.

[Bao et al., 2011] Bao, X., Sahinidis, N. V., and Tawarmalani, M. (2011). Semidefinite relaxations for quadratically constrained quadratic programming: A review and comparisons. *Mathematical programming*, 129:129–157.

[Black and Rangarajan, 1996] Black, M. J. and Rangarajan, A. (1996). On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *International journal of computer vision*, 19(1):57–91.

[Blake and Zisserman, 1987] Blake, A. and Zisserman, A. (1987). *Visual reconstruction*. MIT press.

[Bloesch et al., 2015] Bloesch, M., Omari, S., Hutter, M., and Siegwart, R. (2015). Robust visual inertial odometry using a direct ekf-based approach. In *2015 IEEE/RSJ international conference on intelligent robots and systems (IROS)*, pages 298–304. IEEE.

[Botterill et al., 2011] Botterill, T., Mills, S., and Green, R. (2011). Refining essential matrix estimates from ransac. In *Proceedings Image and Vision Computing New Zealand*, pages 1–6.

[Boumal, 2023a] Boumal, N. (2023a). *An introduction to optimization on smooth manifolds*. Cambridge University Press.

[Boumal, 2023b] Boumal, N. (2023b). *An introduction to optimization on smooth manifolds*. Cambridge University Press.

[Boumal et al., 2019] Boumal, N., Absil, P.-A., and Cartis, C. (2019). Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33.

[Boumal et al., 2016] Boumal, N., Voroninski, V., and Bandeira, A. (2016). The non-convex burer-monteiro approach works on smooth semidefinite programs. *Advances in Neural Information Processing Systems*, 29.

[Boumal et al., 2020] Boumal, N., Voroninski, V., and Bandeira, A. S. (2020). Deterministic guarantees for burer-monteiro factorizations of smooth semidefinite programs. *Communications on Pure and Applied Mathematics*, 73(3):581–608.

[Boyd and Vandenberghe, 2004] Boyd, S. and Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.

[Boyd and Vandenberghe, 2006] Boyd, S. and Vandenberghe, L. (2006). Subgradients. notes for ee364b. *Stanford University, Winter*, 7:2008.

[Briales and Gonzalez-Jimenez, 2016] Briales, J. and Gonzalez-Jimenez, J. (2016). Fast global optimality verification in 3d slam. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4630–4636. IEEE.

[Briales and Gonzalez-Jimenez, 2017a] Briales, J. and Gonzalez-Jimenez, J. (2017a). Cartan-sync: Fast and global se (d)-synchronization. *IEEE Robotics and Automation Letters*, 2(4):2127–2134.

[Briales and Gonzalez-Jimenez, 2017b] Briales, J. and Gonzalez-Jimenez, J. (2017b). Convex global 3d registration with lagrangian duality. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4960–4969.

[Briales et al., 2018] Briales, J., Kneip, L., and Gonzalez-Jimenez, J. (2018). A certifiably globally optimal solution to the non-minimal relative pose problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 145–154.

[Bruns and Schwänzl, 1990] Bruns, W. and Schwänzl, R. (1990). The number of equations defining a determinantal variety. *Bulletin of the London Mathematical Society*, 22(5):439–445.

[Burer and Monteiro, 2005] Burer, S. and Monteiro, R. D. (2005). Local minima and convergence in low-rank semidefinite programming. *Mathematical programming*, 103(3):427–444.

[Byröd et al., 2009] Byröd, M., Josephson, K., and Åström, K. (2009). Fast and stable polynomial equation solving and its application to computer vision. 84:237–256.

[Cai et al., 2019] Cai, Q., Wu, Y., Zhang, L., and Zhang, P. (2019). Equivalent constraints for two-view geometry: Pose solution/pure rotation identification and 3d reconstruction. *International Journal of Computer Vision*, 127:163–180.

[Campos et al., 2019] Campos, J., Cardoso, J. R., and Miraldo, P. (2019). Poseamm: A unified framework for solving pose problems using an alternating minimization method. pages 3493–3499.

[Carlone and Dellaert, 2015] Carlone, L. and Dellaert, F. (2015). Duality-based verification techniques for 2d slam. In *2015 IEEE international conference on robotics and automation (ICRA)*, pages 4589–4596. IEEE.

[Carlone et al., 2015] Carlone, L., Rosen, D. M., Calafiore, G., Leonard, J. J., and Dellaert, F. (2015). Lagrangian duality in 3d slam: Verification techniques and optimal solutions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 125–132. IEEE.

[Cayley, 1846] Cayley, A. (1846). About the algebraic structure of the orthogonal group and the other classical groups in a field of characteristic zero or a prime characteristic. *Reine Angewandte Mathematik*, 32(1846):6.

[Chen et al., 2020] Chen, J., Wu, D., Song, P., Deng, F., He, Y., and Pang, S. (2020). Multi-view triangulation: Systematic comparison and an improved method. *IEEE Access*, 8:21017–21027.

[Chum et al., 2005] Chum, O., Pajdla, T., and Sturm, P. (2005). The geometric error for homographies. *Computer Vision and Image Understanding*.

[Cifuentes, 2021] Cifuentes, D. (2021). A convex relaxation to compute the nearest structured rank deficient matrix. *SIAM Journal on Matrix Analysis and Applications*, 42(2):708–729.

[Decker and Schreyer, 2007] Decker, W. and Schreyer, F.-O. (2007). Varieties, groebner bases, and algebraic curves. *To appear*.

[Ding et al., 2020a] Ding, Y., Barath, D., and Kukelova, Z. (2020a). Homography-based egomotion estimation using gravity and sift features. In *Proceedings of the Asian Conference on Computer Vision*.

[Ding et al., 2021] Ding, Y., Barath, D., Yang, J., Kong, H., and Kukelova, Z. (2021). Globally optimal relative pose estimation with gravity prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 394–403.

[Ding et al., 2020b] Ding, Y., Yang, J., and Kong, H. (2020b). An efficient solution to the relative pose estimation with a common direction. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 11053–11059. IEEE.

[Ding et al., 2019] Ding, Y., Yang, J., Ponce, J., and Kong, H. (2019). An efficient solution to the homography-based relative pose problem with a common reference direction. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1655–1664. IEEE.

[Dubbelman et al., 2012] Dubbelman, G., Dorst, L., and Pijls, H. (2012). Manifold statistics for essential matrices. In *European Conference on Computer Vision*, pages 531–544. Springer.

[Eriksson et al., 2018] Eriksson, A., Olsson, C., Kahl, F., and Chin, T.-J. (2018). Rotation averaging and strong duality. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 127–135.

[Faugeras and Maybank, 1990] Faugeras, O. D. and Maybank, S. (1990). Motion from point matches: multiplicity of solutions. *International Journal of Computer Vision*, 4(3):225–246.

[Fischler and Bolles, 1981] Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

[Fraundorfer et al., 2010] Fraundorfer, F., Tanskanen, P., and Pollefeys, M. (2010). A minimal case solution to the calibrated relative pose problem for the case of two known orientation angles. In *European Conference on Computer Vision*, pages 269–282. Springer.

[Gao et al., 2003] Gao, X.-S., Hou, X.-R., Tang, J., and Cheng, H.-F. (2003). Complete solution classification for the perspective-three-point problem. *IEEE transactions on pattern analysis and machine intelligence*, 25(8):930–943.

[Garcia-Salguero et al., 2021] Garcia-Salguero, M., Briales, J., and Gonzalez-Jimenez, J. (2021). Certifiable relative pose estimation. *Image and Vision Computing*, (63).

[Garcia-Salguero et al., 2022] Garcia-Salguero, M., Briales, J., and Gonzalez-Jimenez, J. (2022). A tighter relaxation for the relative pose problem between cameras. *Journal of Mathematical Imaging and Vision*, (64):493–505.

[Garcia-Salguero et al.,] Garcia-Salguero, M., Dima, E., Mateus, A., and Gonzalez-Jimenez, J. Fast certifiable algorithm for the absolute pose estimation of a camera.

[Garcia-Salguero and Gonzalez-Jimenez, a] Garcia-Salguero, M. and Gonzalez-Jimenez, J. Certifiable planar relative pose estimation with gravity prior.

[Garcia-Salguero and Gonzalez-Jimenez, b] Garcia-Salguero, M. and Gonzalez-Jimenez, J. A fast certifiable algorithm for the n-view planar triangulation.

[Garcia-Salguero and Gonzalez-Jimenez, 2021a] Garcia-Salguero, M. and Gonzalez-Jimenez, J. (2021a). Fast and robust certifiable estimation of the relative pose between two calibrated cameras. *Journal of Mathematical Imaging and Vision*, 63:1036–1056.

[Garcia-Salguero and Gonzalez-Jimenez, 2021b] Garcia-Salguero, M. and Gonzalez-Jimenez, J. (2021b). A sufficient condition of optimality for the relative pose problem between cameras. *SIAM Journal on Imaging Sciences*, 14(4):1617–1634.

[Garcia-Salguero and Gonzalez-Jimenez, 2022] Garcia-Salguero, M. and Gonzalez-Jimenez, J. (2022). Certifiable algorithms for the two-view planar triangulation problem. *Computer Vision and Image Understanding*.

[Garcia-Salguero and Gonzalez-Jimenez, 2023a] Garcia-Salguero, M. and Gonzalez-Jimenez, J. (2023a). Certifiable solver for real-time n-view triangulation. *IEEE Robotics and Automation Letters*, 8(4):1999–2005.

[Garcia-Salguero and Gonzalez-Jimenez, 2023b] Garcia-Salguero, M. and Gonzalez-Jimenez, J. (2023b). Fast certifiable relative pose estimation with gravity prior. *Artificial Intelligence*.

[Golub and Van Loan, 2013] Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. JHU press.

[Gomez-Ojeda et al., 2019] Gomez-Ojeda, R., Moreno, F.-A., Zuñiga-Noël, D., Scaramuzza, D., and Gonzalez-Jimenez, J. (2019). Pl-slam: A stereo slam system through the combination of points and line segments. *IEEE Transactions on Robotics*, 35(3):734–746.

[Grant and Boyd, 2014] Grant, M. and Boyd, S. (2014). CVX: Matlab software for disciplined convex programming, version 2.1. http://cvxr.com/cvx.

[Guan et al., 2018] Guan, B., Vasseur, P., Demonceaux, C., and Fraundorfer, F. (2018). Visual odometry using a homography formulation with decoupled rotation and translation estimation using minimal solutions. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2320–2327. IEEE.

[Guennebaud et al., 2010] Guennebaud, G., Jacob, B., et al. (2010). Eigen v3. http://eigen.tuxfamily.org.

[Hartley and Kahl, 2007a] Hartley, R. and Kahl, F. (2007a). Optimal algorithms in multiview geometry. In *Asian conference on computer vision*. Springer.

[Hartley and Seo, 2008] Hartley, R. and Seo, Y. (2008). Verifying global minima for l 2 minimization problems. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE.

[Hartley et al., 2013] Hartley, R., Trumpf, J., Dai, Y., and Li, H. (2013). Rotation averaging. *International journal of computer vision*, 103:267–305.

[Hartley and Zisserman, 2003] Hartley, R. and Zisserman, A. (2003). *Multiple view geometry in computer vision*. Cambridge university press.

[Hartley and Kahl, 2007b] Hartley, R. I. and Kahl, F. (2007b). Global optimization through searching rotation space and optimal estimation of the essential matrix. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.

[Hartley and Sturm, 1997] Hartley, R. I. and Sturm, P. (1997). Triangulation. *Computer vision and image understanding*, 68(2).

[Hauenstein et al., 2012] Hauenstein, J., Rodriguez, J., and Sturmfels, B. (2012). Maximum likelihood for matrices with rank constraints. *arXiv preprint arXiv:1210.0198*.

[Helmke et al., 2007] Helmke, U., Hüper, K., Lee, P. Y., and Moore, J. (2007). Essential matrix estimation using gauss-newton iterations on a manifold. *International Journal of Computer Vision*, 74(2):117–136.

[Hesch and Roumeliotis, 2011] Hesch, J. A. and Roumeliotis, S. I. (2011). A direct least-squares (dls) method for pnp. pages 383–390.

[Hmam, 2010] Hmam, H. (2010). Quadratic optimisation with one quadratic equality constraint. Technical report, DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION EDINBURGH (AUSTRALIA) ELECTRONIC ....

[Iglesias et al., 2020] Iglesias, J. P., Olsson, C., and Kahl, F. (2020). Global optimality for point set registration using semidefinite programming. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8287–8295.

[Ke and Roumeliotis, 2017] Ke, T. and Roumeliotis, S. I. (2017). An efficient algebraic solution to the perspective-three-point problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7225–7233.

[Kneip et al., 2013] Kneip, L., Furgale, P., and Siegwart, R. (2013). Using multi-camera systems in robotics: Efficient solutions to the npnp problem. In *2013 IEEE International Conference on Robotics and Automation*, pages 3770–3776. IEEE.

[Kneip et al., 2014] Kneip, L., Li, H., and Seo, Y. (2014). Upnp: An optimal o (n) solution to the absolute pose problem with universal applicability. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pages 127–142. Springer.

[Kneip and Lynen, 2013] Kneip, L. and Lynen, S. (2013). Direct optimization of frame-to-frame rotation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2352–2359.

[Kneip et al., 2011] Kneip, L., Scaramuzza, D., and Siegwart, R. (2011). A novel parametrization of the perspective-three-point problem for a direct computation of absolute camera position and orientation. In *CVPR 2011*, pages 2969–2976. IEEE.

[Kukelova et al., 2008] Kukelova, Z., Bujnak, M., and Pajdla, T. (2008). Polynomial eigenvalue solutions to the 5-pt and 6-pt relative pose problems. In *BMVC*, volume 2, page 2008.

[Kukelova and Pajdla, 2007] Kukelova, Z. and Pajdla, T. (2007). Two minimal problems for cameras with radial distortion. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE.

[Lasserre, 2001] Lasserre, J. B. (2001). Global optimization with polynomials and the problem of moments. 11(3):796–817.

[Lee and Civera, 2019a] Lee, S. H. and Civera, J. (2019a). Closed-form optimal two-view triangulation based on angular errors. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

[Lee and Civera, 2019b] Lee, S. H. and Civera, J. (2019b). Triangulation: why optimize? *arXiv preprint arXiv:1907.11917*.

[Lee and Civera, 2020a] Lee, S. H. and Civera, J. (2020a). Geometric interpretations of the normalized epipolar error. *arXiv preprint arXiv:2008.01254*.

[Lee and Civera, 2020b] Lee, S. H. and Civera, J. (2020b). Robust uncertainty-aware multiview triangulation. *arXiv preprint arXiv:2008.01258*.

[Lewis and Wylie, 2019] Lewis, A. and Wylie, C. (2019). A simple newton method for local nonsmooth optimization. *arXiv preprint arXiv:1907.11742*.

[Lourakis and Terzakis, 2021] Lourakis, M. and Terzakis, G. (2021). A globally optimal method for the pnp problem with mrp rotation parameterization. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3058–3063. IEEE.

[Luenberger et al., 1984] Luenberger, D. G., Ye, Y., et al. (1984). *Linear and nonlinear programming*, volume 2. Springer.

[Lui and Drummond, 2013] Lui, V. and Drummond, T. (2013). An iterative 5-pt algorithm for fast and robust essential matrix estimation. In *BMVC*.

[Ma et al., 2001] Ma, Y., Košecká, J., and Sastry, S. (2001). Optimization criteria and geometric algorithms for motion and structure estimation. *International Journal of Computer Vision*, 44(3):219–249.

[Ma et al., 2004] Ma, Y., Soatto, S., Košecká, J., and Sastry, S. (2004). *An invitation to 3-d vision: from images to geometric models*, volume 26. Springer.

[Malis, 2023] Malis, E. (2023). Complete closed-form and accurate solution to pose estimation from 3d correspondences. 8(3):1786–1793.

[Malis and Vargas, 2007] Malis, E. and Vargas, M. (2007). *Deeper understanding of the homography decomposition for vision-based control*. PhD thesis, INRIA.

[Micusik and Wildenauer, 2017] Micusik, B. and Wildenauer, H. (2017). Plane refined structure from motion. In *Scandinavian Conference on Image Analysis*, pages 29–40. Springer.

[Miraldo and Araujo, 2014] Miraldo, P. and Araujo, H. (2014). A simple and robust solution to the minimal general pose estimation. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2119–2125. IEEE.

[Motwani and Raghavan, 1996] Motwani, R. and Raghavan, P. (1996). Randomized algorithms. *ACM Computing Surveys (CSUR)*, 28(1):33–37.

[Mur-Artal et al., 2015] Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D. (2015). Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163.

[Nistér, 2004] Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):756–770.

[Nistér et al., 2004] Nistér, D., Naroditsky, O., and Bergen, J. (2004). Visual odometry. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. Ieee.

[Nistér and Stewénius, 2007] Nistér, D. and Stewénius, H. (2007). A minimal solution to the generalised 3-point pose problem. *Journal of Mathematical Imaging and Vision*, 27(1):67–79.

[Nocedal and Wright, 1999] Nocedal, J. and Wright, S. J. (1999). *Numerical optimization*. Springer.

[Oliensis, 2002] Oliensis, J. (2002). Exact two-image structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(12).

[Olsson and Eriksson, 2008] Olsson, C. and Eriksson, A. (2008). Solving quadratically constrained geometrical problems using lagrangian duality. pages 1–5. IEEE.

[Olsson et al., 2008] Olsson, C., Kahl, F., and Oskarsson, M. (2008). Branch-and-bound methods for euclidean registration problems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(5):783–794.

[Özyeşil et al., 2017] Özyeşil, O., Voroninski, V., Basri, R., and Singer, A. (2017). A survey of structure from motion*. *Acta Numerica*, 26:305–364.

[Park and Boyd, 2017] Park, J. and Boyd, S. (2017). General heuristics for nonconvex quadratically constrained quadratic programming. *arXiv preprint arXiv:1703.07870*.

[Peterson, 1973] Peterson, D. W. (1973). A review of constraint qualifications in finite-dimensional spaces. *Siam Review*, 15(3):639–654.

[Poullis and You, 2011] Poullis, C. and You, S. (2011). 3d reconstruction of urban areas. In *2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, pages 33–40. IEEE.

[Rockafellar, 1993] Rockafellar, R. T. (1993). Lagrange multipliers and optimality. *SIAM review*, 35(2):183–238.

[Rosen et al., 2019] 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.

[Ruiz and Grossmann, 2011] Ruiz, J. P. and Grossmann, I. E. (2011). Using redundancy to strengthen the relaxation for the global optimization of minlp problems. *Computers & Chemical Engineering*, 35(12):2729–2740.

[Saurer et al., 2016] Saurer, O., Vasseur, P., Boutteau, R., Demonceaux, C., Pollefeys, M., and Fraundorfer, F. (2016). Homography based egomotion estimation with a common direction. *IEEE transactions on pattern analysis and machine intelligence*, 39(2):327–341.

[Scaramuzza and Fraundorfer, 2011] Scaramuzza, D. and Fraundorfer, F. (2011). Visual odometry [tutorial]. *IEEE robotics & automation magazine*, 18(4):80–92.

[Schweighofer and Pinz, 2006] Schweighofer, G. and Pinz, A. (2006). Robust pose estimation from a planar target. 28(12):2024–2030.

[Schweighofer and Pinz, 2008] Schweighofer, G. and Pinz, A. (2008). Globally optimal o (n) solution to the pnp problem for general camera models. pages 1–10.

[Shapiro and Fan, 1995] Shapiro, A. and Fan, M. K. (1995). On eigenvalue optimization. *SIAM Journal on Optimization*, 5(3):552–569.

[Shor, 1987] Shor, N. Z. (1987). Quadratic optimization problems. *Soviet Journal of Computer and Systems Sciences*, 25:1–11.

[Shor, 1992] Shor, N. Z. (1992). Dual estimates in multiextremal problems. *Journal of Global optimization*, 2:411–418.

[Smith et al., 2004] Smith, K., Kahanpää, L., Kekäläinen, P., and Traves, W. (2004). *An invitation to algebraic geometry*. Springer Science & Business Media.

[Spetsakis and Aloimonos, 1992] Spetsakis, M. E. and Aloimonos, Y. (1992). Optimal visual motion estimation: A note. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):959–964.

[Stewenius et al., 2006] Stewenius, H., Engels, C., and Nistér, D. (2006). Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60(4):284–294.

[Sturm, 1999] Sturm, J. F. (1999). Using sedumi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1-4):625–653.

[Sun and Deng, 2020] Sun, L. and Deng, Z. (2020). Certifiably optimal and robust camera pose estimation from points and lines. *IEEE Access*, 8:124032–124054.

[Sweeney et al., 2014a] Sweeney, C., Flynn, J., and Turk, M. (2014a). Solving for relative pose with a partially known rotation is a quadratic eigenvalue problem. In *2014 2nd International Conference on 3D Vision*, volume 1, pages 483–490. IEEE.

[Sweeney et al., 2014b] Sweeney, C., Fragoso, V., Höllerer, T., and Turk, M. (2014b). gdls: A scalable solution to the generalized pose and scale problem. pages 16–31.

[Toh et al., 1999] Toh, K.-C., Todd, M. J., and Tütüncü, R. H. (1999). Sdpt3—a matlab software package for semidefinite programming, version 1.3. *Optimization methods and software*, 11(1-4):545–581.

[Trager et al., 2015] Trager, M., Hebert, M., and Ponce, J. (2015). The joint image handbook. In *Proceedings of the IEEE international conference on computer vision*, pages 909–917.

[Triggs et al., 1999] Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (1999). Bundle adjustment—a modern synthesis. In *International workshop on vision algorithms*, pages 298–372. Springer.

[Tron and Daniilidis, 2017] Tron, R. and Daniilidis, K. (2017). The space of essential matrices as a riemannian quotient manifold. *SIAM Journal on Imaging Sciences*, 10(3):1416–1445.

[Tron et al., 2015] Tron, R., Rosen, D. M., and Carlone, L. (2015). On the inclusion of determinant constraints in lagrangian duality for 3d slam. In *Robotics: Science and Systems (RSS), Workshop "The problem of mobile sensors: Setting future goals and indicators of progress for SLAM*, volume 4.

[Ventura et al., 2014] Ventura, J., Arth, C., Reitmayr, G., and Schmalstieg, D. (2014). A minimal solution to the generalized pose-and-scale problem. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 422–429.

[Wadenbäck et al., 2016] Wadenbäck, M., Åström, K., and Heyden, A. (2016). Recovering planar motion from homographies obtained using a 2.5-point solver for a polynomial system. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2966–2970. IEEE.

[Wang et al., 2018] Wang, P., Xu, G., Wang, Z., and Cheng, Y. (2018). An efficient solution to the perspective-three-point pose problem. *Computer Vision and Image Understanding*, 166:81–87.

[Wientapper et al., 2018] Wientapper, F., Schmitt, M., Fraissinet-Tachet, M., and Kuijper, A. (2018). A universal, closed-form approach for absolute pose problems. 173:57–75.

[Wolkowicz, 2000] Wolkowicz, H. (2000). Semidefinite and lagrangian relaxations for hard combinatorial problems. In *System Modelling and Optimization: Methods, Theory and Applications. 19 th IFIP TC7 Conference on System Modelling and Optimization July 12–16, 1999, Cambridge, UK 19*, pages 269–309. Springer.

[Wu et al., 2022] Wu, J., Zheng, Y., Gao, Z., Jiang, Y., Hu, X., Zhu, Y., Jiao, J., and Liu, M. (2022). Quadratic pose estimation problems: Globally optimal solutions, solvability/observability analysis, and uncertainty description. 38(5):3314–3335.

[Yamashita et al., 2012] Yamashita, M., Fujisawa, K., Fukuda, M., Kobayashi, K., Nakata, K., and Nakata, M. (2012). Latest developments in the sdpa family for solving large-scale sdps. In *Handbook on semidefinite, conic and polynomial optimization*, pages 687–713. Springer.

[Yang et al., 2020] Yang, H., Shi, J., and Carlone, L. (2020). Teaser: Fast and certifiable point cloud registration. *arXiv preprint arXiv:2001.07715*.

[Yang et al., 2019] Yang, K., Fang, W., Zhao, Y., and Deng, N. (2019). Iteratively reweighted midpoint method for fast multiple view triangulation. *IEEE Robotics and Automation Letters*, 4(2):708–715.

[Yang et al., 2011] Yang, L., Ying, X., Hou, L., Kong, J., and Zha, H. (2011). Camera resectioning from image edges with the linfty-norm using linear programming. In *BMVC*, pages 1–11. Citeseer.

[Zhang et al., 2023] Zhang, H., Wang, D., and Huo, J. (2023). A visual-inertial dynamic object tracking slam tightly coupled system. *IEEE Sensors Journal*.

[Zhao, 2020] Zhao, J. (2020). An efficient solution to non-minimal case essential matrix estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

[Zhao et al., 2020] Zhao, J., Xu, W., and Kneip, L. (2020). A certifiably globally optimal solution to generalized essential matrix estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12034–12043.

[Zheng et al., 2013] Zheng, Y., Kuang, Y., Sugimoto, S., Astrom, K., and Okutomi, M. (2013). Revisiting the pnp problem: A fast, general and optimal solution. pages 2344–2351.

[Zhou and Kaess, 2019] Zhou, L. and Kaess, M. (2019). An efficient and accurate algorithm for the perspecitve-n-point problem. pages 6245–6252.

[Zhou et al., 2020] Zhou, L., Wang, S., and Kaess, M. (2020). A fast and accurate solution for pose estimation from 3d correspondences. pages 1308–1314.