

分类号: TM863

单位代码: 10335

学 号: 21621094

浙江大学

硕士学位论文



论文题目: 面向视觉-惯性 SLAM 的通用
增量式集束调整框架

作者姓名: 王儒

指导教师: 章国锋

学科 (专业): 计算机科学与技术

研究方向: 计算机视觉

所在学院: 计算机科学与技术学院

提交日期

A Dissertation Submitted to Zhejiang
University for the Degree of
Master of Engineering



TITLE: A General Incremental Bundle
Adjustment Framework for
Visual-Inertial SLAM

Author: Ru Wang

Supervisor: Guofeng Zhang

Subject: Computer Science and Technology

Research Domain: Computer Vision

College: College of Computer Science and Technology

Submitted Date

浙江大学研究生学位论文独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 浙江大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

学位论文作者签名：

签字日期： 年 月 日

学位论文版权使用授权书

本学位论文作者完全了解 浙江大学 有权保留并向国家有关部门或机构送交本论文的复印件和磁盘，允许论文被查阅和借阅。本人授权浙江大学可以将学位论文的全部或部分内 容编入有关数据库进行检索和传播，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书)

学位论文作者签名：

导师签名：

签字日期： 年 月 日 签字日期： 年 月 日

摘要

随着人工智能概念的兴起和 AR、无人机、移动机器人、自动驾驶等行业的发展,工业界与学术界对高精度、高效率、高鲁棒性的三维感知的需求越来越大。在 SLAM 算法中,状态估计算法的速度和精度极大地制约了 SLAM 应用的性能。传统的批量式集束调整求解器为了保证通用性,牺牲了一定的效率,而现有的基于增量式集束调整的求解器虽然效率很高,但由于耦合程度过高,难以在其他 SLAM 系统中使用。本文基于以上分析,提出了适用于 VISLAM 的增量式集束调整通用求解器,提供了基于 IMU 预积分的目标函数接口和重投影误差目标函数接口,以及多种线性求解算法接口,在保证效率和精度的基础上保持了很好的通用性。

关键词: SLAM, 增量式, 集束调整, 状态估计

Abstract

With the rise of artificial intelligence concepts and the development of industries such as AR, drones, mobile robots, and autonomous driving, there is a growing demand for high-precision, high-efficiency, and highly robust 3D sensing in industry and academia. In the SLAM algorithm, the speed and accuracy of the state estimation algorithm greatly limits the performance of the SLAM application. The traditional batch bundle adjustment solver sacrifices efficiency in order to ensure versatility. Existing solver based on incremental bundle adjustment is very efficient, but difficult to integrate into other SLAM systems due to the high degree of coupling. Based on the above analysis, this paper proposes a general incremental bundle adjustment solver for VISLAM, which provides built-in cost functions based on IMU preintegration and reprojection error, while retaining the interfaces for customized cost function types. Meanwhile, we provide a variety types of linear solver algorithm, which ensure efficiency and versatility.

Keywords: SLAM, incremental, bundle adjustment, state estimation

目录

摘要 i

Abstract..... ii

第 1 章 绪论..... 1

 1.1 基于扩展卡尔曼滤波的状态估计 2

 1.1.1 状态预测 3

 1.1.2 状态更新 3

 1.2 基于非线性优化方法的状态估计 4

 1.2.1 基于图优化的状态估计 4

 1.2.2 非线性最小二乘 5

 1.2.3 高斯-牛顿法 6

 1.2.4 莱文贝格-马夸特方法 7

 1.2.5 Dog-Leg 法..... 9

 1.2.6 舒尔补..... 9

 1.2.7 鲁棒估计方法..... 12

 1.2.8 边缘状态估计..... 14

 1.3 视觉惯性 SLAM 相关工作 15

 1.3.1 基于滤波法的 SLAM 系统..... 15

 1.3.2 基于优化法的 SLAM 系统..... 17

 1.3.3 基于直接法的 SLAM 系统..... 19

 1.3.4 增量式集束调整方法..... 19

 1.3.5 VISLAM 系统基本框架 22

 1.4 本文内容及结构..... 23

第 2 章 基于 IMU 预积分的惯性导航 25

 2.1 IMU 基本模型 25

 2.1.1 离散时间噪声和连续时间噪声 27

 2.1.2 IMU 状态传播..... 27

2.2 基于 IMU 预积分技术的 VISLAM..... 29

2.3 预积分项的误差模型 30

2.4 偏移量的线性修正 30

2.5 非线性优化 31

2.6 本章小结..... 32

第 3 章 基于增量式集束调整的高效状态估计 33

3.1 增量舒尔补算法..... 33

3.2 标记脏子图 34

3.3 增量更新舒尔补..... 36

3.4 状态增广 38

3.5 线性求解..... 38

3.6 本章小结..... 41

第 4 章 实验和结论..... 42

4.1 测试设置..... 42

4.2 结果对比..... 43

4.3 本章小结..... 43

第 5 章 总结与展望..... 44

5.1 总结 44

5.2 展望 44

参考文献..... 45

插图

1.1 因子图示例 [1]: 紫色节点代表位姿状态, 绿色节点代表三维点状态, 黑色节点代表状态之间的约束因子。	5
1.2 信赖域与 Dog-Leg 法迭代步 [2].....	10
1.3 一般的稀疏矩阵结构: 左上角部分为三维点状态对应的海森矩阵, 呈对角块稀疏状; 右下角部分为相机状态对应的海森矩阵, 呈条带稀疏状; 左下角矩阵为三维点状态和相机状态的相关海森矩阵, 只有具有观测关系的三维点和相机之间的相关部分为非零。	10
1.4 常用的两种鲁棒函数	13
1.5 信息矩阵稀疏结构 [3]: 左图代表按照一般的消元顺序得到的平方根信息矩阵, 右图代表先使用 COLAMD 算法选择一个较优的消元顺序, 然后得到的平方根信息矩阵。可见, 变量消去顺序对矩阵分解产生的填充现象有巨大影响。	21
1.6 VINS-Mobile[4] 的模块: 逻辑上被划分为初始化、前端视觉惯性跟踪、后端非线性优化以及回路闭合四个模块。	24
2.1 一般的 IMU 误差模型 [5]	26
2.2 OKVIS 因子图示例图 [6].....	31
2.3 IMU 预积分因子图示例 [7].....	31
3.1 因子图.....	33
3.2 正规方程: 左上角块对角矩阵部分对应三维点状态 1 至 6 的信息矩阵 P , 右下角部分为相机状态 1 至 5 对应的信息矩阵 C , 左下角和右上角为矩阵 W 和 W^T ; 中间列向量从上至下依次对应三维点状态 1 至 6 和相机状态 1 至 5 的迭代步; 右侧部分从上至下依次对应三维点状态 1 至 6 和相机状态 1 至 5。	34
3.3 舒尔补: 高亮部分为相机状态对应舒尔补方程, 其余部分为原方程。	34
3.4 计算舒尔补的一次迭代: 从左上角开始依次迭代计算每一个三维点对应的舒尔补部分, 并累加到右下角相机部分中。	35

3.5 因子图待更新部分：相机状态 5 为脏变量；与其直接相连的因子为脏因子；相机状态 4 和三维点状态 6 为只有梯度发生了变化的脏变量。这些节点构成了一个脏子图。 35

3.6 图 3.5 对应的舒尔补方程中的脏块：相机状态 4、5 和三维点状态 6 构成的线性子系统。 36

3.7 更新舒尔补 38

3.8 增广因子图 38

3.9 增广正规方程 40

3.10 更新正规方程 40

第1章 绪论

随着人工智能概念的兴起和增强现实 (Augment Reality, 以下简称 AR)、无人机、移动机器人、自动驾驶等行业的发展, 工业界与学术界对高精度、高效率、高鲁棒性的三维感知的需求越来越大。这类应用通常需要实时并持续地获取场景空间的三维结构信息以及自身的位姿 (朝向和位置) 和运动信息。传统的三维感知方案受限于自身的局限性, 很难同时满足精度、效率和鲁棒性的需求。如普通的 GPS 定位由于频率和精度较低且无法在室内或恶劣天气中使用; 高精度 GNSS 系统 (Global Navigation Satellite System) 和惯性导航系统 INS (Inertial Navigation System) 虽然在精度、效率和鲁棒性上都能达到要求, 但由于售价高达数千至数十万美元, 通常只能在高端领域中使用; 廉价的消费级惯性测量单元 (Inertial Measurement Unit, 以下简称 IMU) 能以非常高的频率实时获取自身的运动信息且具备非常高的稳定性, 但由于其传感器噪声过大, 误差累积严重, 通常只能用来测量旋转信息或初步地测量平移信息。

同时定位和建图 (Simultaneous Localization and Mapping, 以下简称 SLAM) 算法指在未知环境中, 让搭载传感器的设备从某一位置出发对场景进行观测, 在运动和观测过程中计算出自身的位姿 (朝向和位置), 同时逐渐构建场景地图的过程, 其中传感器通常指消费级的相机设备, 也可能包括激光雷达、IMU 等其他传感器。目前在计算机视觉领域, 基于多视图几何和运动恢复结构 (Structure from Motion, 以下简称 SfM) 的算法^[8;9] 已经能做到非常精确的场景地图重建和传感器的定位, 却由于其只能离线工作而不能满足实时性要求高的应用。视觉 SLAM (Visual SLAM, 以下简称 VSLAM) 可以被认为是在线版本的 SfM, 是使用相机 (单目相机或多目相机) 作为唯一的传感器, 利用视觉信息进行定位和定位和建图的方法。相对于 SfM, VSLAM 利用了连续图像帧的局部性, 逐步、增量地计算出位姿并恢复场景的三维结构, 因此能比较容易地达到较高的精度和实时的效率。然而, 单目 VSLAM 系统存在一定的局限性, 即不能恢复场景的绝对尺度信息。另外, VSLAM 非常依赖相机的成像质量, 在图像质量不佳的时候则较难保证算法的鲁棒性。以 V-LOAM^[10] 为代表方法一类激光 SLAM 系统使用了昂贵的激光雷达传感器。由于激光雷达能够获取高精度的三维点云, 使用激光雷达的 SLAM 系统能够获取非常高精度的建图和定位结果, 同时能获得场景的绝对尺度。但是受限于成本、功耗和体积, 激光雷达难以使用在移动 AR、无人机或中小型移动机器人等应用上。视觉惯性 SLAM (Visual Inertial SLAM, 以下简称 VISLAM) 则是使用了相机和 IMU 作为传感器, 基于 IMU 提供的角速度测量和加速度测

量, VISLAM 原生就具备了估计绝对尺度的能力。同时, 受益于 IMU 传感器的高效和稳定, 通过融合视觉信息和惯性信息, VISLAM 比 VSLAM 具有更好的鲁棒性。

此外, VISLAM 所需的相机和消费级 IMU 的价格低廉, 容易获取, 目前大部分手机都配备可以供 SLAM 算法使用的相机和 IMU。综合成本、体积、功耗和性能, VISLAM 与传统的三维感知方案或其他类型的 SLAM 方案相比, 更适合于移动终端的 AR、小型移动机器人、无人机等应用场景, 是未来移动的智能设备上不可或缺的一项关键技术。

在 SLAM 算法中, 状态估计算法的速度和精度极大地制约了 SLAM 应用的性能。目前主流的开源 SLAM 系统的状态估计部分, 大多数基于开源的通用批量式求解器, 如 Ceres-Solver^[11]、g²o^[12] 等, 难以在移动设备上达到实时的性能。而最新的基于增量式求解的求解器, 比如 SLAM++^[13] 和 iSAM2^[1], 又由于与其他模块高度耦合, 难以通用。本文提出了基于增量式舒尔补的通用集束调整求解器, 在保证一定精度的情况下将大规模集束调整的求解性能提升数倍, 并保证了一定的通用性, 方便移植复用。

1.1 基于扩展卡尔曼滤波的状态估计

在 SLAM 领域中, 主流的状态估计算法有两类: 基于扩展卡尔曼滤波(Extended Kalman Filter, 以下简称 EKF) 的滤波法和基于非线性最小二乘(Nonlinear Least Squares) 的优化法。两类方法没有本质上的区别, 都是使用最大化后验概率(Maximum A-Posteriori estimation) 的思想求解非线性系统的最优状态分布。本节将介绍 EKF 的基本原理。

EKF 将状态估计的过程分为状态预测(predict) 和状态更新(update) 两个步骤, 分别对应状态的转移模型(propagation model) 和观测模型(observation model)。在状态预测阶段, EKF 通过状态转移模型预测当前系统状态的先验分布, 而在状态更新阶段, EKF 通过当前时刻的观测模型和观测值对状态的分布进行修正, 得到后验状态分布:

$$\begin{aligned} \text{转移模型} \quad \mathbf{x}_{k+1} &= f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \\ \text{观测模型} \quad \mathbf{z}_{k+1} &= h(\mathbf{x}_{k+1}, \mathbf{v}_{k+1}) \end{aligned} \quad (1.1)$$

其中 \mathbf{u}_k 为输入控制变量, \mathbf{w}_k 和 \mathbf{v}_{k+1} 分别为独立的系统噪声和观测噪声, 均符合零均值高斯分布:

$$\begin{aligned} p(\mathbf{w}_k) &\sim \mathcal{N}(\mathbf{0}, \mathbf{Q}) \\ p(\mathbf{v}_{k+1}) &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \end{aligned} \quad (1.2)$$

1.1.1 状态预测

对于一个离散时间的非线性系统，EKF 假设已知 t_k 时刻系统的状态 \mathbf{x}_k 符合高斯分布：

$$p(\mathbf{x}_k) \sim \mathcal{N}(\hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \quad (1.3)$$

由于系统噪声 \mathbf{w}_k 未知，在 t_{k+1} 时刻可以通过转移模型预测当前系统状态的先验分布：

$$\begin{aligned} p(\mathbf{x}_{k+1}) &\sim \mathcal{N}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{P}_{k+1|k}) \\ \hat{\mathbf{x}}_{k+1|k} &= f(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, \mathbf{0}) \\ \mathbf{P}_{k+1|k} &= \mathbf{A}_{k+1} \mathbf{P}_{k|k} \mathbf{A}_{k+1}^\top + \mathbf{W}_{k+1} \mathbf{Q}_k \mathbf{W}_{k+1}^\top \end{aligned} \quad (1.4)$$

其中

$$\begin{aligned} \mathbf{A}_{k+1} &\doteq \frac{\partial f}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, \mathbf{0}) \\ \mathbf{W}_{k+1} &\doteq \frac{\partial f}{\partial \mathbf{w}}(\hat{\mathbf{x}}_{k|k}, \mathbf{u}_k, \mathbf{0}) \end{aligned} \quad (1.5)$$

1.1.2 状态更新

假设在 t_{k+1} 时刻获得了对系统的观测 \mathbf{z}_{k+1} 而观测噪声 \mathbf{v}_{k+1} 未知，根据观测模型可以得到观测值的条件概率分布：

$$\begin{aligned} p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1}) &\sim \mathcal{N}(\hat{\mathbf{z}}_{k+1}, \mathbf{S}_{k+1}) \\ \hat{\mathbf{z}}_{k+1} &= h(\hat{\mathbf{x}}_{k+1|k}, \mathbf{0}) \\ \mathbf{S}_{k+1} &= \mathbf{V}_{k+1} \mathbf{R}_{k+1} \mathbf{V}_{k+1}^\top \end{aligned} \quad (1.6)$$

其中

$$\begin{aligned} \mathbf{H}_{k+1} &\doteq \frac{\partial h}{\partial \mathbf{x}}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{0}) \\ \mathbf{V}_{k+1} &\doteq \frac{\partial h}{\partial \mathbf{v}}(\hat{\mathbf{x}}_{k+1|k}, \mathbf{0}) \end{aligned} \quad (1.7)$$

为了简洁表示，以下省略下标。由于后验概率 $p(\mathbf{x}|\mathbf{z}) \propto p(\mathbf{x})p(\mathbf{z}|\mathbf{x})$ ，可以构建最大化后验概率问题：

$$\begin{aligned} \max_{\mathbf{x}} & \frac{1}{\sqrt{2\pi}\mathbf{P}^{1/2}} \cdot \exp\left(-\frac{1}{2} \|\mathbf{x} - \hat{\mathbf{x}}\|_{\mathbf{P}^{-1}}^2\right) \\ & \cdot \frac{1}{\sqrt{2\pi}\mathbf{S}^{1/2}} \cdot \exp\left(-\frac{1}{2} \|\mathbf{z} - \hat{\mathbf{z}}\|_{\mathbf{S}^{-1}}^2\right) \end{aligned} \quad (1.8)$$

EKF 假设状态先验 $\hat{\mathbf{x}}$ 足够接近最优解 \mathbf{x} ，且 $\mathbf{e} \doteq \mathbf{x} - \hat{\mathbf{x}}$ 。估最大化上式相当于最小化：

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{e}\|_{\mathbf{P}^{-1}}^2 + \frac{1}{2} \|\mathbf{z} - h(\hat{\mathbf{x}}, \mathbf{0})\|_{\mathbf{S}^{-1}}^2 \quad (1.9)$$

$$\begin{aligned} \mathbf{e} &= (\mathbf{P} + \mathbf{H}^\top \mathbf{S}^{-1} \mathbf{H})^{-1} \mathbf{H}^\top \mathbf{S}^{-1} (\mathbf{z} - h(\hat{\mathbf{x}}, \mathbf{0})) \\ &= \mathbf{P} \mathbf{H}^\top (\mathbf{H} \mathbf{P} \mathbf{H}^\top + \mathbf{S})^{-1} (\mathbf{z} - h(\hat{\mathbf{x}}, \mathbf{0})) \end{aligned} \quad (1.10)$$

记 t_{k+1} 时刻的卡尔曼增益 (Kalman gain) 为

$$\mathbf{K}_{k+1} \doteq \mathbf{P}_{k+1|k} \mathbf{H}_{k+1}^\top (\mathbf{H}_{k+1} \mathbf{P}_{k+1|k} \mathbf{H}_{k+1}^\top + \mathbf{S}_{k+1})^{-1} \quad (1.11)$$

可以对状态进行更新:

$$\hat{\mathbf{x}}_{k+1|k+1} = \hat{\mathbf{x}}_{k+1|k} + \mathbf{K}_{k+1} (\mathbf{z}_{k+1} - h(\hat{\mathbf{x}}_{k+1|k}, \mathbf{0})) \quad (1.12)$$

同时, 要更新后验协方差矩阵:

$$\mathbf{P}_{k+1|k+1} = (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1}) \mathbf{P}_{k+1|k} (\mathbf{I} - \mathbf{K}_{k+1} \mathbf{H}_{k+1})^\top + \mathbf{K}_{k+1} \mathbf{S}_{k+1} \mathbf{K}_{k+1}^\top \quad (1.13)$$

因此, 在状态更新阶段, 可以得到状态的后验概率分布:

$$p(\mathbf{x}_{k+1}) \sim \mathcal{N}(\hat{\mathbf{x}}_{k+1|k+1}, \mathbf{P}_{k+1|k+1}) \quad (1.14)$$

1.2 基于非线性优化方法的状态估计

继上一节介绍了基于 EKF 的状态估计方法, 本节将继续介绍基于非线性最小二乘的状态估计方法, 包括几种主流的求解最小二乘估计的方法以及加速求解的算法, 并从理论角度分析滤波法和优化法的优劣。

目前绝大多数的 SLAM 算法通过集束调整^[14] (Bundle Adjustment) 方法来对状态进行估计。在传统的视觉 SfM 算法中, 集束调整是指通过联合优化所有的视觉观测误差来同时求解最优的相机位姿状态和三维点状态的方法。而对于使用了更多传感器的 SLAM 系统, 除了三维点和位姿状态, 集束调整算法还需要实时地、持续对系统的其他状态比作出估计。对于 VISLAM, 集束调整通常会通过联合优化所有的视觉观测和惯性观测来同时求解相机或 IMU 的位姿、速度以及 IMU 的 bias 状态。

1.2.1 基于图优化的状态估计

图优化方法是分析并求解 SLAM 后端状态估计问题的常用工具。最早在机器人领域, 为了解决多段激光传感器数据融合时的全局一致性问题,^[15;16] 提出了基于位姿图的优化方法。^[17] 在此基础上进一步提出了 GraphSLAM, 使用由位姿状态和三维点状态以及它们

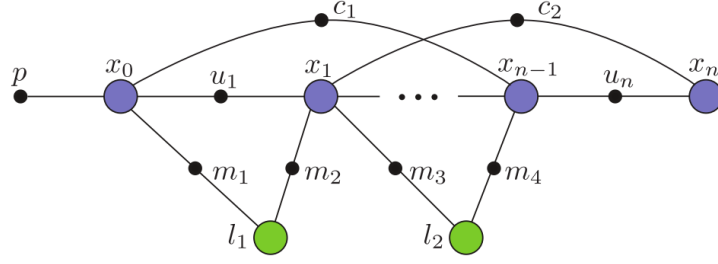


图 1.1 因子图示例 [1]: 紫色节点代表位姿状态, 绿色节点代表三维点状态, 黑色节点代表状态之间的约束因子。

之间的约束构成的因子图来描述集束调整问题。因子图可以很直观地描述集束调整问题中状态和约束之间的关联, 有助于分析集束调整问题的各种性质。经过长时间的发展, 图优化方法已经成为 SLAM 领域的经典方法。

图 1.1 展示了一类常见的因子图结构。按照图示的例子, 一个 SLAM 问题中可能存在的约束有状态变量的先验约束 p , 相邻状态之间的相对位姿约束 $u_1 \dots u_n$, 非相邻状态之间的回路闭合约束 c_1, c_2 以及视觉观测约束 $m_1 \dots m_4$ 等。而求解完整集束调整的过程就相当于最大化整个因子图中的所有约束的集合 \mathcal{Z} 关于所有状态集合 \mathcal{X} 的联合条件概率:

$$P(\mathcal{Z}|\mathcal{X}) = \prod_{f_i \in \mathcal{Z}} P(f_i(\mathcal{X}_i)) \quad (1.15)$$

其中 \mathcal{X}_i 代表所有与约束节点 f_i 相邻的状态节点。通常会假设这是一个高度非线性系统, 并且所有的约束 f_i 的噪声服从高斯分布 $f_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ 。这样的概率最大化问题通常被转化为非线性最小二乘问题来求解:

$$\begin{aligned} \mathcal{X}^* &= \arg \max_{\mathcal{X}} \prod_{f_i \in \mathcal{Z}} P(f_i(\mathcal{X}_i)) \\ &= \arg \min_{\mathcal{X}} \sum_{f_i \in \mathcal{Z}} \frac{1}{2} \|f_i(\mathcal{X}_i)\|_{\sigma_i^2}^2 \end{aligned} \quad (1.16)$$

1.2.2 非线性最小二乘

非线性最小二乘问题难以被直接求解, 通常要使用迭代求解的方法来逼近局部最优解。如果变量的初值在全局最优解的附近, 则一般可以使用牛顿法 (Newton's method) 迭代求解。如下给出了而一个典型的非线性最小二乘问题 $F(\mathbf{x})$ 的形式:

$$F(\mathbf{x}) = \frac{1}{2} \|\mathbf{f}(\mathbf{x})\|^2 \quad (1.17)$$

其中变量 $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f}(\cdot) : \mathbb{R}^n \mapsto \mathbb{R}^m$ 为关于变量 \mathbf{x} 的非线性代价函数, 即目标函数。在局部最优解附近, 最小化能量函数 $F(\cdot)$:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} F(\mathbf{x}) \quad (1.18)$$

等同于求解其梯度函数 $\mathbf{g}(\cdot)$ 的零点 (对于线性的函数则为唯一的零点), 即:

$$\mathbf{g}(\mathbf{x}) = \nabla F(\mathbf{x}) = \left[\frac{\partial F}{\partial x_1} \quad \frac{\partial F}{\partial x_2} \quad \dots \quad \frac{\partial F}{\partial x_n} \right]^\top = \mathbf{0} \quad (1.19)$$

采用牛顿法求解上述零点问题, 记梯度函数 $\mathbf{g}(\cdot)$ 的一阶导数为海森矩阵 (Hessian matrix):

$$\mathbf{H} = \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \frac{\partial^2 F}{\partial x_1 \partial x_2} & \dots & \dots & \frac{\partial^2 F}{\partial x_1 \partial x_n} \\ \frac{\partial^2 F}{\partial x_2 \partial x_1} & \frac{\partial^2 F}{\partial x_2^2} & & & \frac{\partial^2 F}{\partial x_2 \partial x_n} \\ \vdots & & \ddots & & \vdots \\ \vdots & & & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_n \partial x_1} & \frac{\partial^2 F}{\partial x_n \partial x_2} & \dots & \dots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix} \quad (1.20)$$

求解如下线性系统:

$$\boldsymbol{\delta} = \mathbf{H} \setminus \boldsymbol{\eta} \quad (1.21)$$

其中 $\boldsymbol{\eta} = -\mathbf{g}(\mathbf{x})$ 。使用 $\boldsymbol{\delta}$ 更新变量: $\mathbf{x} \leftarrow \mathbf{x} + \boldsymbol{\delta}$ 。式(1.21)通常称为正规方程 (normal equation), 如上求解非线性最小二乘的方法则为正规方程法。线性的问题中, 矩阵 $\mathbf{H}(\mathbf{x})$ 至少为半正定矩阵, 给定任意初值 \mathbf{x} (这里选择了初值为 $\mathbf{0}$) 只需一步就可解得式(1.17)的全局最优解。而对于非线性的情况, 则需要多次迭代才能解得局部最优解。

1.2.3 高斯-牛顿法

对于非线性最小二乘问题, 牛顿法中计算函数 $F(\cdot)$ 的准确海森矩阵 \mathbf{H} 的代价往往很大, 甚至没有闭解。一些算法会寻求使用近似方法计算每次迭代的海森矩阵, 比如拟牛顿法^[2] (Qausi-Newton method)。拟牛顿法适用于一般的最小化问题, 而对于非线性最小二乘问题, 通常可以高斯-牛顿法 (Gauss-Newton method) 求解。高斯-牛顿法是用于求解非线性最小二乘问题的经典迭代算法, 也是众多其他迭代算法的基础。

依然考虑非线性最小二乘问题(1.17), 使用高斯-牛顿法迭代求解, 首先要给定变量的初值 $\mathbf{x} = \mathbf{x}_0$ 。对其在当前点处进行泰勒展开, 并忽略高阶项, 可得目标函数 $\mathbf{f}(\cdot)$ 在 \mathbf{x} 处的线性近似:

$$\mathbf{f}(\mathbf{x} + \boldsymbol{\delta}) \simeq \mathbf{f}(\mathbf{x}) + \mathbf{J}\boldsymbol{\delta} \quad (1.22)$$

其中 $J \in \mathbb{R}^{m \times n}$ 是目标函数 $f(\cdot)$ 在 \mathbf{x} 处的雅各比矩阵 (Jacobian)。相应地,

$$\begin{aligned} F(\mathbf{x} + \boldsymbol{\delta}) &\simeq L(\boldsymbol{\delta}) = \frac{1}{2} \|\mathbf{f}(\mathbf{x}) + J\boldsymbol{\delta}\|^2 \\ &= F(\mathbf{x}) + J^\top \mathbf{f} + \frac{1}{2} \boldsymbol{\delta}^\top J^\top J \boldsymbol{\delta} \end{aligned} \quad (1.23)$$

式(1.17)被转化为局部 \mathbf{x} 处的线性最小二乘子问题, 使用牛顿法求解这个线性最小二乘子问题

$$\min_{\boldsymbol{\delta}} \frac{1}{2} \|\mathbf{f}(\mathbf{x}) + J\boldsymbol{\delta}\|^2 \quad (1.24)$$

需要构建正规方程:

$$H_{gn} = J^\top J, \quad \boldsymbol{\eta} = -J^\top \mathbf{f}(\mathbf{x}) \quad (1.25)$$

再求解, 得到高斯-牛顿法一次迭代的结果:

$$\boldsymbol{\delta}_{gn} = H_{gn} \setminus \boldsymbol{\eta} \quad (1.26)$$

将此结果更新至变量: $\mathbf{x} \leftarrow \mathbf{x} + \boldsymbol{\delta}$, 然后不断重复以上过程, 直至结果收敛。

式(1.25)和(1.21)的形式类似, 也可以认为高斯-牛顿法和牛顿法的最大区别就是高斯-牛顿法使用了 $J^\top J$ 来近似海森矩阵。

1.2.4 莱文贝格-马夸特方法

除了高斯-牛顿法之外, 还有一种经典的求解非线性最小二乘优化的算法: 莱文贝格-马夸特 (Levenberg-Marquardt) 法。它也被认为是带有阻尼因子的高斯-牛顿法, 也是信赖域方法 (trust region method) 的前身^[18]。

前面提到, 一般的纯视觉的 SLAM 算法中, 尺度是不可观测的, 也就是说 VSLAM 的状态估计中, 对全局所有的三维点坐标和相机位置进行统一的缩放, 不会引起重投影误差的改变 (在不考虑数值误差的情况下)。从求解 VSLAM 集束调整问题的角度来看, 这一性质造成了在直接使用高斯-牛顿法时海森矩阵的秩亏现象, 即矩阵 $J^\top J$ 为半正定的情况。此时直接使用矩阵分解求逆的方法求解线性最小二乘子问题就会得到数值不稳定的结果, 比如解得的步长在某个方向过大等。

莱文贝格法

针对这一情况, 莱文贝格在基础的高斯-牛顿法中引入了对迭代步长的直接约束, 将原先每次迭代的线性最小二乘子问题改写为:

$$\min_{\boldsymbol{\delta}} \frac{1}{2} (\|\mathbf{f}(\mathbf{x}) + J\boldsymbol{\delta}\|^2 + \mu \|\boldsymbol{\delta}\|^2), \quad \mu > 0 \quad (1.27)$$

这样一来，原先高斯-牛顿法求解的正规方程就变成了

$$H_l = (J^T J + \mu I), \quad \eta = -J^T f(x) \quad (1.28)$$

$$\delta_l = H_l \setminus \eta \quad (1.29)$$

其中的 μ 就是所谓的阻尼因子 (damping factor)。

阻尼因子控制住了每次迭代计算的步长的模。通常阻尼因子 μ 具备下面三个性质^[2]:

1. 对于任意的 $\mu > 0$ ，海森矩阵 $(J^T J + \mu I)$ 正定，这一点保证了求得的步长 δ 处于能量下降的方向；
2. 当 $\mu \rightarrow \infty$ 时，求得的步长 δ 接近于最速下降方向 $-\frac{1}{\mu} J^T f(x)$ ，当当前的变量 x 的值距离最优解较远时，最速下降的方向更容易符合预期；
3. 当 $\mu \rightarrow 0$ 时，则求得的步长 δ 更接近于高斯-牛顿法的结果，如果当前变量 x 的值较为接近最优解，则这样的步长更容易符合预期，因为在收敛点附近，真实的海森矩阵更接近标准的二次型形式。

同时，由于正则项 μI 的加入，修改后的海森矩阵 $(J^T J + \mu I)$ 一定为正定二次型矩阵，提升了求解的数值稳定性。

莱文贝格-马夸特法

莱文贝格在海森矩阵的每一个主元上都施加了同样的阻尼限制 μ ，某种程度上将是对步长 δ 在各个梯度方向上施加了相同的限制，这显然不是最合理的做法。当 μ 非常大的时候，一定程度上由原始海森矩阵 $J^T J$ 提供的信息就被抑制了，难以被利用。马夸特在莱文贝格的方法上进行了改进，根据海森矩阵的主元大小，对步长的不同梯度方向施加不同的阻尼：

$$H = (J^T J + \mu \text{diag}(J^T J)), \quad \eta = -J^T f(x) \quad (1.30)$$

$$\delta_{lm} = H \setminus \eta \quad (1.31)$$

这样一来，在梯度较小的方向阻尼较小， δ_{lm} 的分量也就较大，反之则阻尼较大， δ_{lm} 的分量较小。得益于更合理的阻尼设置，相对于单纯的莱文贝格法，莱文贝格-马夸特方法在实际应用中的收敛效率要更好。

1.2.5 Dog-Leg 法

Dog-Leg 法是另一个经典的求解非线性最小二乘优化的算法，也是信赖域方法的代表。莱文贝格-马夸特方法通过阻尼因子来间接地控制每一次迭代的步长，而 Dog-Leg 法则显式地使用了信赖域来约束每一次迭代的步长。下面给出传统的使用 Dog-Leg 法求解问题(1.17)的步骤^[2]。首先按照常规的求解步骤进行线性化，得到线性最小二乘子问题(1.24)。Dog-Leg 法首先使用式(1.26)求解高斯-牛顿迭代步 δ_{gn} ，然后使用最速下降法求解一阶迭代步：

$$\delta_{sd} = -\frac{\|g\|^2}{\|Jg\|^2}g \quad (1.32)$$

和莱文贝格-马夸特法类似，Dog-Leg 法也是一种结合了高斯-牛顿法和最速下降法的方法。如式(1.33)所示，

$$\delta_{dl} := \begin{cases} \delta_{gn}, & \text{if } \|\delta_{gn}\| \leq \Delta \\ \frac{\Delta}{\|\delta_{sd}\|} \delta_{sd}, & \text{elseif } \|\delta_{sd}\| \geq \Delta \\ \delta_{sd} + \beta(\delta_{gn} - \delta_{sd}), & \text{elsewise} \end{cases} \quad (1.33)$$

Dog-Leg 法根据当前的信赖域半径 Δ 和以上求解的两个步长来计算综合的迭代步 δ_{dl} ：

1. 如果高斯-牛顿迭代步 δ_{gn} 的模小于信赖域半径，则认为可以直接选择 δ_{gn} ；
2. 如果高斯-牛顿迭代步 δ_{gn} 的模和最速下降迭代步 δ_{sd} 都大于信赖域半径，则选择 δ_{sd} 并将其缩放到信赖域大小；
3. 如果高斯-牛顿迭代步 δ_{gn} 的模大于信赖域半径而最速下降迭代步的模小于信赖域半径，则需要取 δ_{gn} 和 δ_{sd} 连线与信赖域的交点作为迭代步（如图 1.2）。

这样，当 Δ 较大时，Dog-Leg 法更倾向于使用高斯-牛顿法的迭代步 δ_{gn} ；当 Δ 较小时则更倾向于最速下降法的迭代步 δ_{sd} 。算法 1 详细描述了 Dog-Leg 法的迭代过程。和高斯-牛顿法一样，Dog-Leg 法在求解式(1.26)的时候也可能遇到海森矩阵 H 半正定的情况，此时应该选择数值稳定的线性求解方法来求解海森矩阵的伪逆 H^\dagger 。

1.2.6 舒尔补

不论是使用高斯-牛顿法，莱文贝格-马夸特方法还是 Dog-Leg 法，在每一轮迭代过程中，都需要进一步通过求解正规方程这个线性系统来得到每一轮迭代的步长（如式(1.26)，(1.31)，(1.33)）。根据经验，SLAM 问题具有良好的稀疏性：通常三维点状态的数量要远大

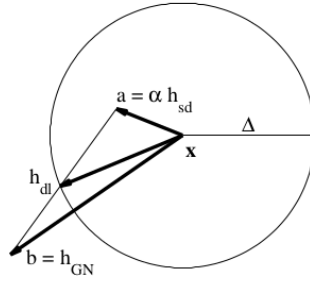


图 1.2 信赖域与 Dog-Leg 法迭代步 [2]

于相机状态的数量，而且通常三维点之间没有直接联系；同时，只有物理上距离比较接近的三维点状态和相机状态之间有相互约束。因此，由集束调整问题构建的海森矩阵 H 也具有稀疏的性质，通过对变量适当的排序，将三维点状态和相机状态分别分组，一般可以得到如图 1.3所示的特殊的稀疏结构。

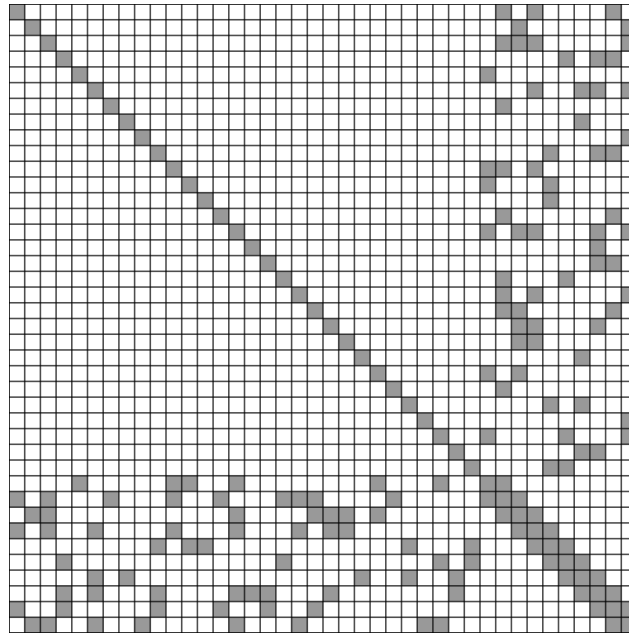


图 1.3 一般的稀疏矩阵结构：左上角部分为三维点状态对应的海森矩阵，呈对角块稀疏状；右下角部分为相机状态对应的海森矩阵，呈条带稀疏状；左下角矩阵为三维点状态和相机状态的相关海森矩阵，只有具有观测关系的三维点和相机之间的相关部分为非零。

对于这样具备特殊稀疏结构的线性系统，可以用舒尔补 (Schur complement) 来加速求解。舒尔补是一种常用的加速求解稀疏线性系统的方法，特别适用于集束调整问题中的正规方程的求解。其本质上是一种基于高斯消元的分块求解线性系统的方法。首先根据图 1.3对

算法 1 Dog-Leg 法**输入:** 初始信赖域半径 Δ_0 , 初值 \mathbf{x}_0 **输出:** 优化结果 \mathbf{x} $\Delta := \Delta_0, \mathbf{x} := \mathbf{x}_0$ **for** $k = 0 \rightarrow k_{max}$ **do** $k := k + 1$

使用式(1.24)计算线性化子问题

使用式(1.26)计算高斯-牛顿迭代步 δ_{gn} 使用式(1.32)计算最速下降迭代步 δ_{sd} 使用式(1.33)计算 Dog-Leg 法步长 δ_{dl} **if** {步长 δ_{dl} 收敛} **then**

结束优化

end if计算迭代步质量 ϱ :

$$\varrho = \frac{F(\mathbf{x}) - F(\mathbf{x} + \delta_{dl})}{L(\mathbf{0}) - L(\delta_{dl})}$$

if { $\varrho > 0.0$ } **then** $\mathbf{x} := \mathbf{x} + \delta_{dl}$ **end if****if** { $\varrho > \epsilon_0$ } **then** $\Delta := \text{fmax}\{\Delta, 3 \|\delta_{dl}\|\}$ **else if** { $\varrho < \epsilon_1$ } **then** $\Delta := \Delta/2$ **end if****if** {信赖域 Δ 收敛} **then**

结束优化

end if**end for**

正规方程进行分块:

$$\mathbf{H} = \begin{bmatrix} \mathbf{P} & \mathbf{W}^\top \\ \mathbf{W} & \mathbf{C} \end{bmatrix}, \quad \boldsymbol{\delta} = \begin{bmatrix} \delta_p \\ \delta_c \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \mathbf{p} \\ \mathbf{c} \end{bmatrix} \quad (1.34)$$

其中 \mathbf{P} 是三维点状态对应的海森矩阵, \mathbf{C} 是相机状态对应的海森矩阵, \mathbf{W} 是三维点状态和

相机状态的海森矩阵。求解三维点状态和相机状态之前，先通过行变换构建舒尔补方程：

$$(C - WP^{-1}W^T) \delta_c = c - WP^{-1}p \quad (1.35)$$

记舒尔补矩阵为 $S \doteq C - WP^{-1}W^T$ ，右侧为 $b \doteq c - WP^{-1}p$ 。如前面提到的，集束调整问题中的三维点状态数量远多于相机状态数量，故矩阵 S 的规模远小于 P 的规模，通过求解下舒尔补方程：

$$\delta_c = S \setminus b \quad (1.36)$$

可以快速先得到相机状态的迭代步。而又因为矩阵 P 通常呈块对角状，其逆矩阵的求解 P^{-1} 也非常容易，通过变量回代又可以高效地求解剩余的三维点状态的迭代步：

$$\delta_p = P \setminus (p - W^T \delta_c) \quad (1.37)$$

舒尔补可以加速稀疏线性系统的求解，但需要注意的是，舒尔补一般会造成矩阵的填充效应 (fill in)^[3]，即破坏矩阵稀疏性的情况。在使用舒尔补的时候如果不注意变量的排序，则很容易造成严重的填充。

1.2.7 鲁棒估计方法

最小二乘非常容易受数据中的外点 (outliers) 影响。也就是说，基于非线性最小二乘的因子图优化对于问题中的错误约束非常敏感。考虑一般的 VSLAM，一种常见的产生错误约束的情况就是问题中特征误匹配带来的错误视觉观测约束。错误的约束通常对应着更大的目标函数，为了降低错误约束的影响，通常要引入鲁棒估计的方法 (robust estimation) 来进一步限制问题。一个带有鲁棒项的非线性最小二乘优化问题 (robustified nonlinear least squares) 将基本的非线性最小二乘问题(1.17)修改为了如下形式：

$$R(\mathbf{x}) = \frac{1}{2} \rho(F(\mathbf{x})) = \frac{1}{2} \rho(\|\mathbf{f}(\mathbf{x})\|^2) \quad (1.38)$$

其中标量函数 $\rho(\cdot) : \mathbb{R} \mapsto \mathbb{R}$ 为鲁棒函数 (robust function)，通常为高度非线性。

由于鲁棒函数 $\rho(\cdot)$ 给原先非线性最小二乘问题又引入了高度的非线性性质，为了保证计算的正确性，求解鲁棒非线性最小二乘问题的时候不能忽略 $\rho(\cdot)$ 二阶导数^[14]：

$$\begin{aligned} H &= J^T (\rho' I + 2\rho'' \mathbf{f}(\mathbf{x}) \mathbf{f}^T(\mathbf{x})) J \\ \boldsymbol{\eta} &= -\rho' J^T \mathbf{f}(\mathbf{x}) \end{aligned} \quad (1.39)$$

可见，鲁棒函数 $\rho(\cdot)$ 会根据每一个目标函数的模其对正规方程的贡献权重，而且随着优化的进行，这个权重会在每一轮迭代中动态地调整。为了达到合理降低外点权重的目标，需要合理选择鲁棒函数形式。由式(1.39)可见，鲁棒函数的选择尽可能地要满足如下几个条件^[14;19]：

1. $\rho(0) = 0$
2. $\rho'(0) = 1$
3. 当 $\|\mathbf{f}(\mathbf{x})\|^2$ 的值处于外点区域时， $\rho'(\|\mathbf{f}(\mathbf{x})\|^2) < 1$
4. 当 $\|\mathbf{f}(\mathbf{x})\|^2$ 的值处于外点区域时， $\rho''(\|\mathbf{f}(\mathbf{x})\|^2) < 0$

其中前两点是为了保证鲁棒函数不对收敛的目标函数造成影响，后两点是为了保证鲁棒函数能有效地降低处于外点的目标函数的权重。常用的鲁棒函数有 Cauchy 函数：

$$\rho(s) = \ln(1 + s) \quad (1.40)$$

和 Huber 函数：

$$\rho(s) = \begin{cases} s, & s \leq 1 \\ 2\sqrt{s} - 1, & s \geq 1 \end{cases} \quad (1.41)$$

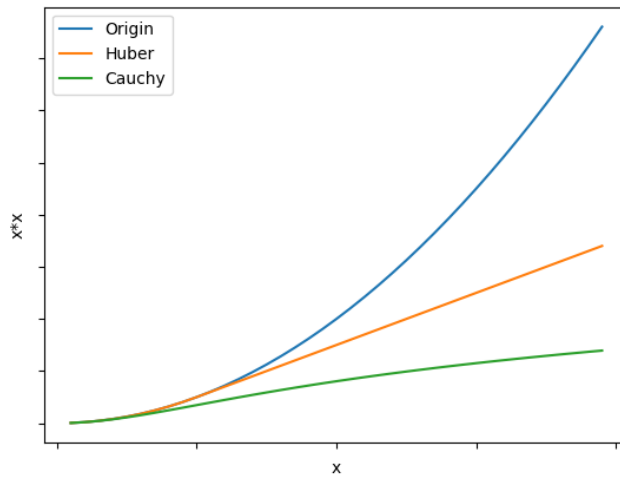


图 1.4 常用的两种鲁棒函数

1.2.8 边缘状态估计

对于比较大的系统，状态数量比较多，一般需要通过边缘化对状态进行精简。在图优化中对应的就是通过删减状态节点调整因子图的过程。假设状态 \mathbf{x} 和 \mathbf{y} 符合联合高斯分布：

$$p(\mathbf{x}, \mathbf{y}) \sim \mathcal{N} \left(\begin{bmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{yx}^\top \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix} \right) \quad (1.42)$$

对于高斯分布，总是可以使用舒尔补将上面的联合概率分布拆分成两个独立分布的乘积 $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$ ：

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &\sim \mathcal{N}(\boldsymbol{\mu}_x + \Sigma_{yx}^\top \Sigma_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y), \Sigma_{xx} - \Sigma_{yx}^\top \Sigma_{yy}^{-1} \Sigma_{yx}) \\ p(\mathbf{y}) &\sim \mathcal{N}(\boldsymbol{\mu}_y, \Sigma_{yy}) \end{aligned} \quad (1.43)$$

其中 $p(\mathbf{y})$ 就是变量 \mathbf{y} 的边缘概率，而 $p(\mathbf{x}|\mathbf{y})$ 则是变量 \mathbf{x} 关于 \mathbf{y} 的条件概率。在最小二乘问题中，正规方程的海森矩阵 \mathbf{H} 和右侧 $\boldsymbol{\eta}$ 对应的是信息矩阵 (information matrix) 和信息向量 (information vector)^[20]。信息矩阵也对应于状态的协方差矩阵的逆 Σ^{-1} ，而正规方程的解对应与变量的期望。以状态 \mathbf{x} 和 \mathbf{y} 为例，可以构建正规方程：

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{xx} & \mathbf{H}_{yx}^\top \\ \mathbf{H}_{yx} & \mathbf{H}_{yy} \end{bmatrix} \setminus \begin{bmatrix} \boldsymbol{\eta}_x \\ \boldsymbol{\eta}_y \end{bmatrix} \quad (1.44)$$

易得联合概率分布：

$$p(\mathbf{x}, \mathbf{y}) \sim \mathcal{N} \left(\begin{bmatrix} \mathbf{H}_{xx} & \mathbf{H}_{yx}^\top \\ \mathbf{H}_{yx} & \mathbf{H}_{yy} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\eta}_x \\ \boldsymbol{\eta}_y \end{bmatrix}, \begin{bmatrix} \mathbf{H}_{xx} & \mathbf{H}_{yx}^\top \\ \mathbf{H}_{yx} & \mathbf{H}_{yy} \end{bmatrix}^{-1} \right) \quad (1.45)$$

根据式(1.43)，同样使用舒尔补可以计算得到状态 \mathbf{y} 的边缘概率分布：

$$\begin{aligned} p(\mathbf{y}) &\sim \mathcal{N} \left(\underbrace{(\mathbf{H}_{yy} - \mathbf{H}_{yx} \mathbf{H}_{xx}^{-1} \mathbf{H}_{yx}^\top)^{-1}}_{\text{边缘期望}} \underbrace{(\boldsymbol{\eta}_y - \mathbf{H}_{yx} \mathbf{H}_{xx}^{-1} \boldsymbol{\eta}_x)}_{\text{边缘信息向量}}, \right. \\ &\quad \left. \underbrace{(\mathbf{H}_{yy} - \mathbf{H}_{yx} \mathbf{H}_{xx}^{-1} \mathbf{H}_{yx}^\top)^{-1}}_{\text{边缘协方差}} \right) \end{aligned} \quad (1.46)$$

如1.2.6节中提到的，舒尔补一般会造成稀疏矩阵的填充效应，在使用舒尔补进行边缘化操作的时候，同样会遇到这样的情况。因此，在使用边缘化进行状态删减的时候，也需要注意使用合理的策略选择被删减的状态，尽量减少填充。

1.3 视觉惯性 SLAM 相关工作

VSLAM 和 VISLAM 都需要根据系统的先验估计、传感器观测来对系统的状态进行估计,可能包括实时的位姿状态、传感器状态、场景三维点状态以及其他可能的信息。主流的 SLAM 方法根据状态估计的方法可以分为基于扩展卡尔曼滤波 (Extended Kalman Filter, 以下简称 EKF) 的滤波法和基于非线性最小二乘 (Nonlinear Least Squares) 的优化法。两类方法没有本质上的区别,都是使用最大化后验概率 (Maximum A-Posteriori estimation, 以下简称 MAP) 的思想求解非线性系统的最优状态估计,但是在求解的速度、精度以及算法的可扩展性上存在一定差异。根据对视觉信息的利用方法又可以分为直接法 (direct methods) 和特征法 (feature-based methods)。使用关键点进行图像匹配,并使用重投影误差 (reprojection error) 作为视觉约束求解三维结构的方法称为特征法。直接使用稠密或稀疏的图像像素信息,并使用光度误差 (photometric error) 作为视觉约束的方法称为直接法。

1.3.1 基于滤波法的 SLAM 系统

一般来讲滤波法的时间复杂度小于同规模的优化法,早期由于计算能力的限制,选择滤波法进行状态估计是一种合理的选择。MonoSLAM^[21] 是最早的单目 VSLAM 系统之一。MonoSLAM 使用了 EKF 求解相机状态和三维点状态,也是最早的滤波法 VSLAM 系统之一。在状态估计阶段, MonoSLAM 快速地将旧的相机状态通过边缘化操作消去,将系统的状态数量限制在 $O(N)$, 其中 N 为三维点的数量,同时状态估计的时间复杂度被限制在了 $O(N^3)$ 。因此 MonoSLAM 的运行时间得到了限制,不至于随着时间无限增长。但由于 MonoSLAM 过早地将相机状态进行边缘化,一部分尚未收敛的相机状态就将错误的信息留在了系统的状态中,导致了较大的误差累积。另一方面,频繁的边缘化操作也就导致了系统状态的协方差矩阵变得稠密,无法利用稀疏的求解方法进行加速,这一点也是基于 EKF 的算法的通病。

另一个经典的基于 EKF 的 SLAM 系统是 MSCKF^[22]。MSCKF 全称是多状态约束卡尔曼滤波 (Multi-State Constraints Kalman Filter)。与 MonoSLAM 不同的是, MSCKF 使用了相机状态窗口的概念,计算大小为 M 的系统的 EKF 更新,其中 M 为状态窗口的大小。因此 MSCKF 将状态估计的实现复杂度限制在了 $O(M^3)$ 。由于在 SLAM 问题中,三维点的数量往往远远大于相机状态的数量: $N \gg M$, MSCKF 算法可以达到实时性的要求,效率要高于 MonoSLAM。又由于 MSCKF 保留了一定数量的历史相机状态,而不是尽早

地将旧的状态消去，因此误差累积更小，更适合长时间、大尺度的应用。

此外，MSCKF 还使用了 IMU 信息，因此属于 VISLAM。MSCKF 使用了一般的迭代式 IMU 积分技术，即直接通过对 IMU 读数进行积分，得到最新相机状态的先验分布，而将图像特征信息作为观测，经过 EKF 更新融合视觉信息和 IMU 信息。MSCKF 的算法的框架如下：

1. 状态传播：对于 IMU 读数，使用龙格库塔法对其进行积分，同时根据 IMU 噪声参数更新状态的先验协方差矩阵；
2. 图像注册：每当得到新的图像时，根据当前的最新状态以及 IMU-相机外参对状态进行增广，得到状态的先验。并对图像进行处理，提取视觉特征，更新特征跟踪信息等；
3. 状态更新：选择合适的时机，通过 EKF 更新融合视觉信息和 IMU 信息，得到后验状态分布。

MSCKF 的状态窗口包括最新的相机状态和保留在窗口内的部分历史相机状态。在 k 时刻其状态变量定义如下：

$$\mathbf{X}_k \doteq [\mathbf{X}_{\text{IMU}_k}, {}^G\mathbf{R}_{C_1}, {}^G\mathbf{p}_{C_1}, \dots, {}^G\mathbf{R}_{C_N}, {}^G\mathbf{p}_{C_N}] \quad (1.47)$$

除了 MSCKF，后续的一些主流的 VISLAM 也大多使用了如上的相机-IMU 状态定义。

如前面描述的，MSCKF 算法是无三维结构信息的，也就是只保留相机状态，而不估计三维点的状态。在状态估计前选取一部分视觉特征，快速地通过三角化计算出对应的三维点状态，然后快速地通过边缘化操作消去，最后再对相机状态进行更新。MSCKF 依据一定的策略来选取特征，以下两条规则会触发状态更新：

1. 这种情况触发得最频繁：当一个视觉特征在被连续跟踪数帧后丢失（移出相机视野）时会触发一次状态更新。这种情况下，由于在最新的相机中特征已经丢失，因此后续它不会再更新，可以通过边缘化操作将它消去。
2. 每次得到图像时，MSCKF 都会增广当前的状态。为了尽可能保留长的基线和几何信息，当相机状态数量达到上限 M_{max} 时，MSCKF 选择从状态中第二旧的相机开始，平均地选择 $M_{max}/3$ 个相机状态。使用被选中的相机观测到的特征跟踪来进行 EKF 更新，最后消去对应的状态。

MSCKF 求解的状态估计仍然不是全局最优的。前面提到,被消去的状态的信息要先以边缘化的方式将保留下来,即得到剩余状态的边缘概率分布。一部分剩余状态的雅各比矩阵,其线性化点的被永远固定在了发生边缘化的时刻。后续即使状态值发生了改变,旧的雅各比矩阵由于已经被编码进了先验状态分布中而不能再改变,而随后得到的新的观测信息总是会根据最新的状态值计算雅各比矩阵,这就造成了所谓的信息不一致 (inconsistency)。客观上,信息不一致造成了 MSCKF 的状态估计在某些不可观测的自由度上引入了一些不必要的误差。为了解决信息一致性的问题,后续版本的 MSCKF2.0^[23] 引入了 FEJ 技术 (First Estimate Jacobian) 技术^[24],在计算新的观测关于旧的变量的雅各比矩阵时,总是使用状态边缘化发生时刻的线性化点,一定程度上缓解了一个问题,提高了精度。另一些侧重于解决信息不一致性的 SLAM 相关工作有^[25] 和^[26],通过选用特殊的线性化点的方式保持了系统的不可观测的自由度,以及使用了一个变种 EKF 算法弥补错误不可观测自由度的^[27],也可以提高状态估计的精度。

1.3.2 基于优化法的 SLAM 系统

随着硬件能力的发展,状态优化的计算效率已经越来越不再是 SLAM 系统的瓶颈所在。一方面,由于优化法和滤波法并没有时间复杂度上的差异,也有许多算法能够利用 SLAM 问题的稀疏性和局部性减少优化的计算量,比如使用舒尔补、稀疏矩阵分解甚至增量式优化的方法。一些高效的基于优化法的 SLAM 系统已经能在效率上逼近甚至超越基于滤波法的 SLAM 系统。另一方面,优化法在精度上相对于滤波法有着不可比拟的优势。尽管有很多的手段提升滤波法的精度,但是由于难以求解全局最优解,滤波法仍然会有较大的误差累积。前面提到,滤波法和优化法在概率上都是 MAP 估计,由于每次求解状态估计,EKF 都只做一次线性化并进行一次更新,而优化法则通过迭代不断更新线性化点求解直至收敛,因此滤波法可以认为是使用了一轮迭代的优化法。显而易见,相比于优化法,滤波法虽然速度更快,但不保证状态收敛。而且基于优化法的 SLAM 可以在基础的视觉观测约束、IMU 约束上其他约束,进一步消除误差累积,基于滤波法的 SLAM 则在这方面的可扩展性上稍差。因此,越来越多的主流 SLAM 系统已经在使用基于优化法的状态估计。

PTAM^[28] 是最经典的基于优化法的 VSLAM 系统的代表。在系统架构上,PTAM 创新性地将局部的相机跟踪定位和全局优化分发到两个线程中运行。在前端相机跟踪线程中,PTAM 对系统的移动速度进行了建模,使用估计的速度预测相机的位姿。借助于预测的位姿,可以得到旧的视觉特征点在新的相机图像中的投影,从而缩小特征匹配时的搜索范围。给定特征匹配,通过最小化重投影误差可以得到相机位姿的粗略估计。由于前端线程

只求解位姿，这一步可以实时运行。当前端线程还会筛选出质量较高的图像帧作为关键帧(keyframe) 加入到后端全局优化线程中。而在后端全局优化线程中，PTAM 使用了集束调整来求解全局的地图。全局地图的求解时间会随关键帧的增长而呈 $O(M^3)$ 级别的增长，受限与此，PTAM 将关键帧的数量限制在了 100 帧。但即便如此，将相机跟踪和全局优化分发到不同的线程中执行的思想被证明可以大幅提高基于优化法的 SLAM 的性能，在随后的 SLAM 研究工作中，这种思想被大量借鉴。目前的主流 SLAM 系统大多数都使用了多线程的方法提高系统的效率。

ORB-SLAM^[29;30] 是另一个经典的 VLSAM 工作，也是目前最先进的 SLAM 系统之一。ORB-SLAM 使用了 ORB 特征来提升跟踪质量。在 PTAM 的基础上，ORB-SLAM 还额外利用了全局地图的信息，加入了重定位和回路闭合模块。同样的，ORB 也使用了多线程架构：局部相机跟踪、地图优化和回路闭合被分发到三个不同的线程中执行。局部相机跟踪线程维护了一个由关键帧组成的滑动窗口，会适时地优化最新关键帧和与其具有共视关系的一系列关键帧。由于局部窗口大小固定，局部跟踪线程的计算时间被限制在一个可接受的水平。为了尽量避免全局优化，在回路闭合线程中，ORB-SLAM 使用了最小生成树建立的 Essential Graph 的方法来构建关键帧之间的回路约束。回路闭合的过程中也没有包括三维点状态的求解，Essential Graph 以较低的计算开销获取了较好的回路闭合性能。

OKVIS^[6] 是另一个基于优化法的 SLAM 系统。同时还是 VISLAM 的经典框架之一。与 ORB-SLAM 不同的是，OKVIS 仅维护了一个局部的滑动窗口优化。由于不包含全局地图优化，OKVIS 并不具备回路闭合和重定位能力。OKVIS 在相机跟踪部分使用了和 MSCKF 类似的迭代式 IMU 积分技术，随后通过滑动窗口优化，同时最小化 IMU 运动误差和重投影误差，得到包含比较准确的尺度信息的状态估计结果。同时，OKVIS 在保持优化的稀疏性的前提下使用了包含边缘化的滑动窗口优化，以提升优化的精度。滑动窗口会随着时间增长逐渐滑动，滑出窗口的状态一般需要进行边缘化，为了尽可能保持优化的稀疏性，OKVIS 会根据滑出窗口的是否为关键帧来判断是否要进行完整的边缘化操作。

VINS-Mono^[31] 则是最新的一个较为完整的 VISLAM 系统实现。相对于以上的 SLAM 系统，VINS-Mono 的框架更为完整，包含了鲁棒的初始化模块、包含边缘化的滑动窗口优化模块、重定位模块和全局优化模块。同样，这些模块被分发到了不同的线程中执行，以提升效率。

VINS-Mono 使用了一个松耦合的方式对系统的状态进行初始化。在初始化阶段，VINS-Mono 使用视觉 SfM 方法由一系列初始帧构建一个小规模的地图，然后通过与独立的 IMU 积分算法构建的初始位姿进行对齐，获取初始的地图状态和位姿状态，以及一系列初始的

IMU 状态参数。在滑动窗口模块，VINS-Mono 首先通过 KLT 跟踪获取每一帧的初始位姿估计。在优化部分，VINS-Mono 借鉴了 OKVIS 的做法，使用稀疏求解器求解状态估计，同时使用了类似 OKVIS 的选择性边缘化策略来保持优化的稀疏性。不同的是，VINS-Mono 使用了基于 IMU 预积分^[21] 技术，而不是传统的迭代式 IMU 积分，以提升 IMU 状态的优化效果。对于滑出窗口的状态，如果判断为关键帧，则会被加入到后端的全局优化中。全局优化运行于独立的线程，VINS-Mono 使用了位姿图优化的方式对这些历史关键帧进行进一步更新，提供给回路闭合模块使用。VINS-Mono 在系统的功能、速度和精度上达到了较高的水准。而且其移动端的版本 VINS-Mobile^[4] 经过精简，可以在移动设备上达到实时的性能。

1.3.3 基于直接法的 SLAM 系统

以上介绍的 SLAM 系统均基于特征法。另一类直接法 SLAM 系统近年来也获得了较大的关注。特征法和直接法也没有绝对的优劣之分，特征法利用了图像中的关键信息，因此通常对于几何误差（如相机内参的误差）和图像噪声（如光照变化、卷帘快门相机的图像撕裂）更为鲁棒，但是关键点的提取和匹配通常比较耗时；直接法不需要提取和匹配关键点，因此速度较快，但对图像噪声较为敏感；另一方面，由于直接法通常利用了更多图像信息，因此在弱纹理情况下表现更好，而且其构建的地图往往比较稠密，可以进一步提供给三维重建算法使用。

DSO (Direct Sparse Odometry)^[32] 是目前最先进的直接法 VSLAM 系统之一。DSO 使用了类似^[33] 提出的稀疏直接法进行跟踪，这一点与以往的大部分直接法 VSLAM 都不同。DSO 也使用了优化法进行状态估计。除了传统直接法中使用的光度误差，DSO 还对场景中的光照进行建模，提出了使用曝光时间、镜头晕影 (lens vignetting) 和非线性相应函数。这一改进使得 DSO 在光照变化的场景下具有更好的精度和鲁棒性。

类似 OKVIS，DSO 也使用了包含边缘化的滑动窗口优化方法。为了保证滑动窗口中的相机状态在三维空间中有较好的分布，DSO 设计了一个特殊设计的评分函数对关键帧进行评分。当状态数量达到上限时，DSO 会根据评分挑选需要消去的相机状态。为了保持稀疏性，DSO 也会有选择地对消去的状态进行边缘化。

1.3.4 增量式集束调整方法

在实现 SLAM 算法时，需要在算法的精度和性能两者中做出权衡。完整的集束调整包括了对所有历史状态的所有观测，虽然可以获得最优的状态估计，但其计算代价往往是难

以接受的。在一些对性能要求高的应用场景，例如 AR 应用和自动驾驶应用中，算法的性能往往决定了它的可用性甚至安全性。随着这类应用对实时的、高效 SLAM 算法需求的日益增加，一些致力于在保证一定精度的前提下降低计算代价的集束调整算法应运而生，其主要通过两种策略来减少算法的计算量。一类是针对应用场景的特点减小集束调整问题的规模，基于前几节的介绍，可以总结出以下几种方式：

- 全局优化的规模：只保留状态变量，而不保留三维点变量，如 ORB-SLAM 的 Essential Graph 和 VINS-Mono 的位姿图优化；
- 基于历史状态窗口：只估计最近的数个历史状态或一系列选定的数个历史状态，OKVIS、VINS-Mono 等的滑动窗口优化；
- 基于关键帧：只估计一部分选定的携带了足够信息的历史状态，而放弃一些冗余历史状态，如 OKVIS、VINS-Mono 等。

以上的方法通常也可以结合，或搭配多线程技术使用。

另一种策略是通过深入分析集束调整问题的特点，做针对性优化，减少冗余的计算。集束调整问题通常具有非常特殊的性质，合理利用这些性质，可以帮助更高效地求解 SLAM 系统的状态估计。比如，在 SLAM 系统中，通常要求解的三维点变量的数量要远大于状态变量的数量，而且通常状态约束集中在状态与状态之间、状态与三维点之间，而一般不会直接出现在三维点与三维点之间。这就导致集束调整构建的正规方程具有特别的稀疏结构，如图 1.3。应该利用这种稀疏结构，并且在优化的过程中尽可能保持这种稀疏性。另外，在线的集束调整中，通常只有较新加入的变量会发生比较大的变动，旧的变量由于经过持续的优化而变动很小。可以利用这种局部的性质，只对少部分变量进行更新，从而减少集束调整的计算量。

以 iSAM 系列（主要包括 iSAM^[3] 和 iSAM2^[1]）为例，增量式集束调整算法很好地利用了 SLAM 集束调整问题的稀疏性和局部性，实现了高效的求解。iSAM 系列的工作主要深入分析了优化求解的以下几条特点：

- 对观测的雅各比矩阵进行 QR 分解，或者对信息矩阵进行 Cholesky 分解得到平方根信息矩阵 (square root information matrix) 的过程，等同于使用消元法将因子图转化成贝叶斯网络的过程；
- 矩阵分解时产生的填充现象 (fill-in)，则等同于因子图转化过程中，多出来的变量之间的边；

- 同样，在矩阵分解时，选定不同的主元顺序（pivoting）对应于因子图转化过程中变量的消去顺序，其所产生的填充现象的程度也是不同的，如图 1.5。

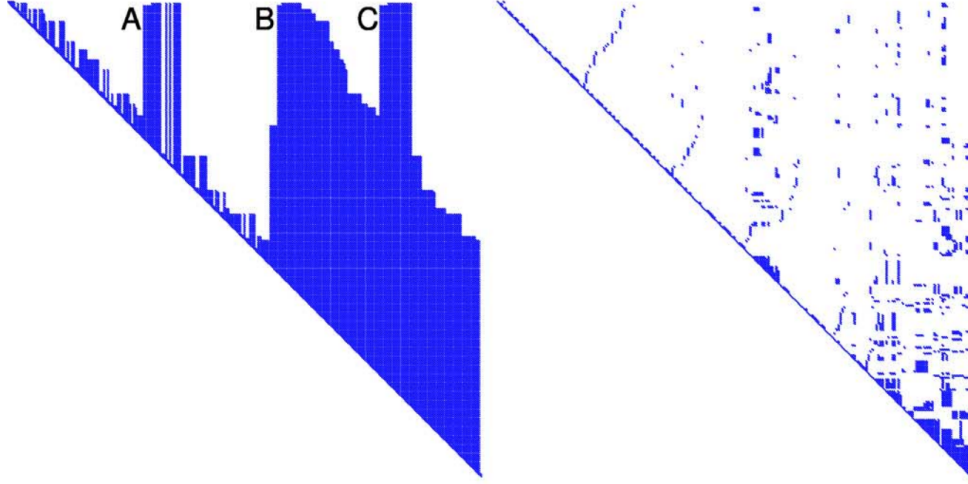


图 1.5 信息矩阵稀疏结构 [3]: 左图代表按照一般的消元顺序得到的平方根信息矩阵，右图代表先使用 COLAMD 算法选择一个较优的消元顺序，然后得到的平方根信息矩阵。可见，变量消去顺序对矩阵分解产生的填充现象有巨大影响。

因此，完全可以用分析因子图的转化过程来代替直接对矩阵分解过程的分析。具体来说，iSAM 系列算法使用一种特殊的贝叶斯树结构编码并维护了优化过程中分解得到的平方根信息矩阵，使得每次更新平方根信息矩阵时只要做很少的改动即可：

1. 减少填充现象：通过一些启发式算法如 COLAMD^[34]、CHOLMOD^[35] 算法找到次优（suboptimal）的变量消去顺序，使得因子图转化为贝叶斯置信网络的过程中尽可能地不产生新的边；
2. 使用贝叶斯树编码平方根信息矩阵：对于因子图转化而来的贝叶斯置信网络，使用特殊的贝叶斯树来编码其中的稀疏结构和变量因果关系；
3. 即时重新线性化：利用贝叶斯树中的编码的变量因果关系，每当因子图中加入新的变量或约束，则即时地对影响到的变量进行线性化，避免每次都重新构造整个平方根信息矩阵；
4. 局部变量更新策略：通过设置一个阈值 ϵ ，在求解得到变量的增量 Δ 后，根据其是否满足 $|\Delta| > \epsilon$ 来判断是否需要更新这个变量。

这样，维护一个增量求解的集束调整算法的时间复杂度就会远低于一般的完整的集束调整，从而在不损失精度或损失很少的精度情况下将 SLAM 算法的速度提升一个甚至数个数量级。

1.3.5 VISLAM 系统基本框架

前面介绍了系列经典的 VSLAM/VISLAM 基本框架，可以总结得到一个完整的 SLAM 系统一般要包括初始化、前端跟踪、后端优化（局部和全局优化）、重定位和回路闭合等模块：

- **初始化模块：**初始化模块的目标是获得初始的状态变量值。与 VSLAM 相比，VISLAM 的状态估计部分由于状态量变多，需要包括初始的旋转¹、速度、尺度、IMU 的相关参数等。有的系统并没有着重介绍初始化部分，如 MSCKF 更依赖后续的滤波算法来使状态收敛，而 VINS-Mono 则具有较为完整的初始化流程。
- **前端跟踪模块：**VISLAM 中，前端跟踪包含对视觉信息的预处理，比如提取关键点、计算描述符、进行特征匹配等。在直接法 SLAM 中，则需要计算光流跟踪等信息。通常 IMU 的频率要远高于相机获取图像的帧率，因此两种传感器之间的时间戳对齐也是 VISLAM 前端跟踪一个较为挑战的问题。一般可以采用线性插值的方法，以图像的时间戳为准，对 IMU 的读数进行插值。插值法可以得到同步后的两个图像帧之间的一组 IMU 读数组成的窗口，当接收到图像的时候，需要对图像进行预处理，同时计算 IMU 的相关积分。IMU 方面，使用 IMU 预积分通常精度要好于传统的迭代式积分。
- **后端优化模块：**后端优化模块可以采用上节提到过的包含边缘化的局部窗口优化的方法或全局优化。也可以在后端维护一个较长的状态窗口，每当有新的状态加入到窗口中时触发非线性优化求解。集束调整可以采用前面介绍的一些方法来加速：无结构的位姿图优化、关键帧优化、利用全局优化的局部性开发高效的增量优化算法等。

在基于关键帧的 VSLAM 中，通常只需要根据图像本身来考虑关键帧的筛选，比如特征是否丰富、特征分布是否合理，与其他帧能否形成足够的距离和共视关系等。在窗口维护的时候也有不同的做法：顺序滑动，或用最新的状态替换窗口中最新的状态等。离开窗口的状态则以边缘化的方式消去，其结果需要作为先验信息加入到后续优化中。

¹由于重力加速度的值可以认为是不变的，因此初始化计算重力方向即可认为是计算相对于重力的旋转

而在 VISLAM 中,除了要保证关键帧本身的图像质量和关键帧之间的视差,还要考虑 IMU 的影响。由于 IMU 积分模型的模型采用了大量的近似,其模型误差是比较大的。同时考虑到各方面的噪声以及多重积分带来的误差放大效应,移动端常用的消费级 IMU 通常只在较短的时间窗口内能保证积分的精度。故在选择关键帧时,关键帧之间的 IMU 窗口不能过长。

VISLAM 的视觉跟踪要求足够的视差,而惯导部分要求较短的时间窗口,这就形成了一些挑战,比如 AR 应用中时常会遇到低速运动情况,这时候连续帧图像差异比较小,足够的视差通常就意味着较大的 IMU 积分误差。这时候就需要算法根据实际情况合理设置关键帧筛选策略。

- **回路闭合模块:**回路闭合模块可以在检测到回路时调用全局优化,消除长时间跟踪的积累误差。但是出于性能上的考虑,全局优化的不适合频繁地调用。一种策略是在检测到回路闭合之后,将匹配到的历史关键帧加入到当前的滑动窗口中,并设置为固定的变量。这样,窗口内的集束调整就成了以历史关键帧为条件的条件状态估计。另一方面,又可以在额外维护一个全局的集束调整模块,以较低的频率运行以消除地图的误差,以及利用 Essential Graph 或者位姿图优化进一步达到节省计算,同时不会引入很大的误差的目的。

图 1.6展示了 VINS-Mobile 中对 VISLAM 框架的模块划分,具有代表性:

1.4 本文内容及结构

本章总结了 VISLAM 的研究现状,阐述了当前 SLAM 领域的研究热点和技术难点,列举并分析了主流的 VISLAM 系统的框架原理和优缺点。包括基于滤波方法和基于非线性优化方法的状态估计推导,并简要介绍了目前最高效的增量式集束调整方法的思想。接下来本文将着重介绍一个基于 IMU 预积分技术和增量式集束调整技术的 SLAM 系统的设计和实现。

第2章将详细介绍目前最新的基于 IMU 预积分的惯性导航技术,阐述 IMU 传感器的基本模型和使用方法,以及如何将惯性辅助模块集成进视觉 SLAM 系统。包括 IMU 预积分方法实现,IMU 状态估计的方法,SLAM 前端跟踪模块中 IMU 和相机之间的协同工作,并对比其和传统的 IMU 惯性导航技术的优劣;第3章将详细介绍基于增量式集束调整的 SLAM 后端优化模块。包括增量式集束调整的原理和流程,以及工程实现上的细节,并分析与其

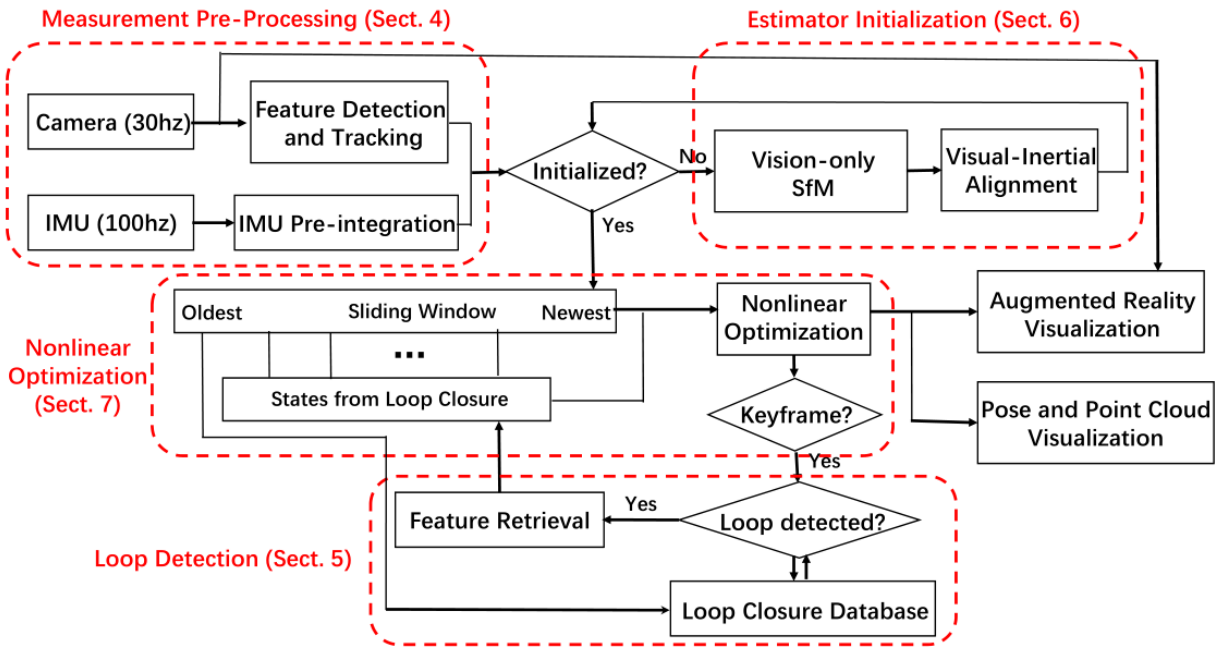


图 1.6 VINS-Mobile[4] 的模块：逻辑上被划分为初始化、前端视觉惯性跟踪、后端非线性优化以及回路闭合四个模块。

他增量式算法之间的共性和优劣；第4章通过实验结果评估求解效率，进一步验证基于 IMU 预积分的增量式集束调整的效果；第5章给出本文的结论以及后续工作的展望。

第2章 基于 IMU 预积分的惯性导航

如上一章中提到的，单目 VSLAM 系统存在一定的局限性，非常依赖相机的成像质量，并且无法绝对尺度进行观测^[36]。而在移动 AR、自动驾驶、无人机等应用中，为了与场景交互或是提供路径规划，SLAM 算法必须鲁棒地提供带尺度信息的位姿估计和运动估计。而且随时可能出现的光照、纹理质量的变化，以及相机快速运动带来的图像模糊，都时刻在挑战着相机跟踪算法的鲁棒性。

近些年来，IMU 被广泛地与 VSLAM 系统结合。不管是使用滤波方法还是非线性优化方法实现的 VISLAM 系统，对 IMU 的建模都大同小异，基本集中在两类：一类是类似于 MSCKF^[22]、OKVIS^[6] 那样传统的迭代式 IMU 积分，另一类则是使用 IMU 预积分^[7] 的方法，如 VINS^[4]。

2.1 IMU 基本模型

IMU 是测量物体三轴角速度以及加速度的装置，包含三轴陀螺仪和线加速度计。IMU 工作极其稳定，几乎不会有宕机的情况，而且通常能以远高于相机的频率（数百到上千赫兹）得到相对于自身坐标系的角速度 $\boldsymbol{\omega}$ ($\text{rad} \cdot \text{s}^{-1}$) 和加速度 \boldsymbol{a} ($\text{m} \cdot \text{s}^{-2}$)。依靠这些数据，可以通过数值积分计算得到系统的朝向 $\mathbf{R} \in \text{SO}^3$ 、速度 $\boldsymbol{v} \in \mathbb{R}^3$ 和平移 $\boldsymbol{p} \in \mathbb{R}^3$ 信息，因此基于 IMU 的跟踪结果直接就带有了场景的尺度信息。很自然地，IMU 工作并不依赖于视觉信息，可以在相机遭遇恶劣的光照、纹理以及运动模糊的情况下提供较为可靠的跟踪结果。但值得注意的是，消费级别的 IMU 传感器，其信号通常带有较大的 bias¹和噪声，并且系统的速度 \boldsymbol{v} 和平移 \boldsymbol{p} 分别来自于加速度 \boldsymbol{a} 的一次和二次数值积分，bias 和噪声也会通过积分以平方级别被放大，带入到跟踪结果中。因此仅使用 IMU 的跟踪结果是不可靠的，需要定期使用视觉跟踪信息来纠正 IMU 跟踪的误差。

鉴于其噪声特性，合理地使用 IMU 辅助视觉跟踪是一项重大挑战。不管是使用滤波的方法还是非线性优化的方法融合视觉信息和 IMU 信息，都要对 IMU 的误差进行合理的建模。将 IMU 的实际物理运动定义为传感器的输入，IMU 对实际运动进行测量得到角速度和加速度定义为传感器输出，则输出和输入的差异即为 IMU 读数的误差，常见的 IMU 误差模型可以用图 2.1 来表示。

¹指 IMU 输出值和输入值之间的偏移，可能受温度、重力等因素的影响

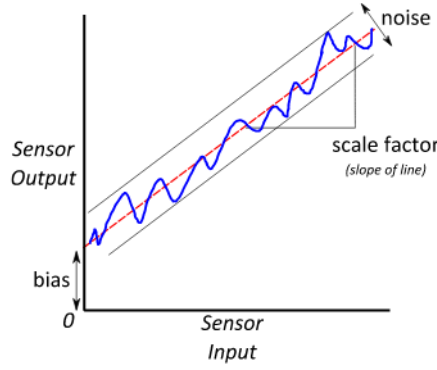


图 2.1 一般的 IMU 误差模型 [5]

可以看到，IMU 误差的成分是比较复杂的，可能包含 bias、随机游走噪声（random walk）²、尺度系数（scale factor）³、不正交性（non-orthogonality）⁴等等[5]，而不同类型误差对于运动估计的影响也是不同的。在跟踪的过程中，系统的朝向 R 可以通过角速度 ω 的数值积分得到；加速度 \mathbf{a} 会被进行一次和二次积分，分别得到速度 \mathbf{v} 和平移 \mathbf{p} 。加速度经过了二次数值积分，其误差也会通过两次积分进入平移量 \mathbf{p} 中，故相比角速度的误差，跟踪算法对于加速度的误差更为敏感。实际经验也表明，由 IMU 积分得到的朝向信息是较为准确的，而位置信息的误差则是较大的。

一般 IMU 在出厂时都会经过厂商的校准，故在实际跟踪和状态估计的过程中可以使用一个简化的 IMU 误差模型。以常用的六轴 IMU 为例，大部分的 VISLAM 系统都使用了相同的 IMU 误差模型，如 MSCKF、VINS-Mono 等，即将其误差简化为 bias 和测量噪声两个部分，并假设加速度计和陀螺仪相互统计独立，且各自的三轴之间都统计独立。使用下标 m 标记每次测量的值，不带下标的表示真实值左上标 G 来表示参考系为全局坐标系，不带左上标的表示参考系为局部坐标系。在不考虑地球自转的情况下，角速度和加速度的测量分别为：

$$\begin{aligned}\tilde{\omega} &= \omega + \mathbf{b}^g + \mathbf{n}^g, \\ \tilde{\mathbf{a}} &= {}^G\mathbf{R}^\top ({}^G\mathbf{a} - {}^G\mathbf{g}) + \mathbf{b}^a + \mathbf{n}^a\end{aligned}\tag{2.1}$$

可以看到，IMU 的的读数可以看成是真实值、偏移和加性高斯白噪声（additive Gaussian white noise）之和。而加速度的测量中又包含了重力 ${}^G\mathbf{g}$ 的影响。bias 和白噪声具备如下性质：

- bias \mathbf{b}^g 和 \mathbf{b}^a ：随时间的增长符合零均值的高斯随机游走。该噪声从 IMU 开机以来就

²指当输入固定时，IMU 输出值中包含的随机噪声，它随时间的变化可被认为是一个随机过程

³指的是输出和输入之间的线性缩放系数

⁴指由于 IMU 各轴并不完全正交所带来的误差

一直存在，不断积累，与是否从 IMU 读取数据无关。可以用如下高斯分布来描述：

$$\begin{aligned} \mathbf{b}^g &\sim \mathcal{N}(\mathbf{b}_0^g, \mathbf{Q}_g), \\ \mathbf{b}^a &\sim \mathcal{N}(\mathbf{b}_0^a, \mathbf{Q}_a) \end{aligned} \quad (2.2)$$

- 测量噪声 \mathbf{n}^g 和 \mathbf{n}^a ：零均值加性高斯白噪声，表示从 IMU 读取数据时产生的小抖动，而且这个噪声只在测量时产生，也符合高斯分布：

$$\begin{aligned} \mathbf{n}^g &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_g), \\ \mathbf{n}^a &\sim \mathcal{N}(\mathbf{0}, \mathbf{R}_a), \end{aligned} \quad (2.3)$$

为了方便后续的计算，IMU 测量误差的协方差也可以记为：

$$\Sigma = \begin{bmatrix} \mathbf{Q}_g & 0 & 0 & 0 \\ 0 & \mathbf{R}_g & 0 & 0 \\ 0 & 0 & \mathbf{Q}_a & 0 \\ 0 & 0 & 0 & \mathbf{R}_a \end{bmatrix} \quad (2.4)$$

2.1.1 离散时间噪声和连续时间噪声

上面提到的噪声模型都是在连续时间系统下面考虑的，连续时间系统的高斯协方差（或标准差）也就称为传感器的噪声强度。实际应用中，不可能每时每刻都在读取 IMU 数据，而应该考虑离散时间系统的情况。^[37] 给出了离散时间噪声和连续时间噪声的关系。可知，离散时间系统下 IMU 噪声的协方差与连续时间系统存在如下关系：

$$\begin{aligned} \mathbf{Q}_{gd} &= \mathbf{Q}_g \Delta t \\ \mathbf{Q}_{ad} &= \mathbf{Q}_a \Delta t \\ \mathbf{R}_{gd} &= \mathbf{R}_g / \Delta t \\ \mathbf{R}_{ad} &= \mathbf{R}_a / \Delta t \end{aligned} \quad (2.5)$$

其中 Δt 为采样时间间隔，并且陀螺仪、加速度计三轴的噪声统计独立。

2.1.2 IMU 状态传播

所谓的 IMU 状态传播 (IMU propagation) 指的就是根据通过将当前时刻 IMU 的读数在时间上进行积分，来得到下一时刻的 IMU 状态。这个 IMU 状态包括 IMU 的位姿、运动以及噪声等，通过积分操作，IMU 读数的信息就被传播到了下一时刻。

首先考虑理想情况，也就是 IMU 的读数不存在白噪声，偏移量也不存在随机游走的情况，此时 IMU 的读数符合：

$$\begin{aligned}\boldsymbol{\omega} &= \tilde{\boldsymbol{\omega}} - \mathbf{b}^g, \\ \mathbf{a} &= \tilde{\mathbf{a}} - \mathbf{b}^a + \mathbf{g}.\end{aligned}\tag{2.6}$$

记全局坐标系下 IMU 的状态为（为保持简洁，省略了转置符号^T）：

$$\mathbf{X}_{\text{IMU}} \doteq [{}^G\mathbf{R}, {}^G\mathbf{p}, {}^G\mathbf{v}, \mathbf{b}^g, \mathbf{b}^a].\tag{2.7}$$

具体来说，积分过程包括积分这个 IMU 状态向量的五个分量。那么，要对 IMU 状态进行积分，首先需要写出 IMU 状态关于时间的导数：

$$\begin{aligned}{}^G\dot{\mathbf{R}} &= {}^G\mathbf{R}[(\tilde{\boldsymbol{\omega}} - \mathbf{b}^g) \times], \\ {}^G\dot{\mathbf{p}} &= {}^G\mathbf{v}, \\ {}^G\dot{\mathbf{v}} &= {}^G\mathbf{R}(\tilde{\mathbf{a}} - \mathbf{b}^a + \mathbf{g}), \\ \dot{\mathbf{b}}^g &= \mathbf{0}, \\ \dot{\mathbf{b}}^a &= \mathbf{0}\end{aligned}\tag{2.8}$$

显然，由于偏移 \mathbf{b}^g 和 \mathbf{b}^a 随着时间的增长符合零均值的高斯随机游走，故它们关于时间的导数为零。积分的步骤如下：

1. 将角速度在时间上积分，得到当前 IMU 的朝向；
2. 将加速度在时间上积分，得到当前 IMU 的速度；
3. 将速度再在时间上积分，得到当前 IMU 的位置；
4. 角速度偏移和加速度偏移由于增长符合零均值随机游走，所以直接取旧的状态中的值。

可以使用任意一种数值积分算法来实现，比如欧拉法（Euler method）、龙格库塔法（Runge-Kutta methods）等。此外，不仅要状态向量进行数值积分，同时还要更新协方差矩阵。

2.2 基于 IMU 预积分技术的 VISLAM

接下来重点介绍 IMU 预积分技术^[7]。与普通基于 IMU 积分的 VISLAM 系统不同, IMU 预积分使用了更精确的相对运动模型。将 IMU 观测模型包含三个部分: 相对旋转、相对速度、相对平移, 可以认为它们是仅关于偏移量的函数。使用以下算法进行 IMU 预积分, 假设需要对 t_i 和 t_j 之间 IMU 读数进行积分 (假设相邻两帧 IMU 之间时间间隔相同, 即 $\Delta t_{i,i+1} = \Delta t_{i+1,i+2} = \dots = \Delta t_{j-1,j} = \Delta t$):

$$\begin{aligned} \mathbf{R}_j &= \mathbf{R}_i \prod_{k=i}^{j-1} \exp((\tilde{\boldsymbol{\omega}}_k - \mathbf{b}_k^g - \mathbf{n}_k^g)\Delta t) \\ \mathbf{v}_j &= \mathbf{v}_i + \mathbf{g}\Delta t_{ij} + \sum_{k=i}^{j-1} \mathbf{R}_k(\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \mathbf{n}_k^a)\Delta t \\ \mathbf{p}_j &= \mathbf{p}_i + \sum_{k=i}^{j-1} [\mathbf{v}_k\Delta t + \frac{1}{2}\mathbf{g}\Delta t^2 + \frac{1}{2}\mathbf{R}_k(\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \mathbf{n}_k^a)\Delta t^2] \end{aligned} \quad (2.9)$$

传统的 VISLAM 中, 通常将连续关键帧之间的一系列 IMU 读数进行积分, 得到相对的状态变化 $\Delta \mathbf{X}$, 然后加上前一个关键帧的状态 \mathbf{X}_i , 得到下一个关键帧的状态 \mathbf{X}_j , 同时计算出累积的协方差, 作为下一个关键帧的先验估计, 参与到视觉的优化中去。这样做的缺陷就是, 积分项 $\Delta \mathbf{X}$ 实际上会随着偏移量的更新而改变, 除非在每一轮优化迭代时重新积分计算新的 $\Delta \mathbf{X}$, 否则 IMU 积分项的误差会一直存在。

按照如下方式定义状态 \mathbf{X}_i 和 \mathbf{X}_j 之间的相对状态变化 $\Delta \mathbf{X}$:

$$\begin{aligned} \Delta \mathbf{R}_{ij} &\doteq \mathbf{R}_i^\top \mathbf{R}_j = \prod_{k=i}^{j-1} \exp((\tilde{\boldsymbol{\omega}}_k - \mathbf{b}_k^g - \mathbf{n}_k^g)\Delta t) \\ \Delta \mathbf{v}_{ij} &\doteq \mathbf{R}_i^\top (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g}\Delta t_{ij}) = \sum_{k=i}^{j-1} \Delta \mathbf{R}_{ik}(\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \mathbf{n}_k^a)\Delta t \\ \Delta \mathbf{p}_{ij} &\doteq \mathbf{R}_i^\top (\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i\Delta t_{ij} - \frac{1}{2}\mathbf{g}\Delta t_{ij}^2) \\ &= \sum_{k=i}^{j-1} [\Delta \mathbf{v}_{ik}\Delta t + \frac{1}{2}\Delta \mathbf{R}_{ik}(\tilde{\mathbf{a}}_k - \mathbf{b}_k^a - \mathbf{n}_k^a)\Delta t^2] \end{aligned} \quad (2.10)$$

每一轮优化迭代过后, 使用一阶的线性方法对 $\Delta \mathbf{R}_{ij}$, $\Delta \mathbf{p}_{ij}$, $\Delta \mathbf{v}_{ij}$ 这三个量进行更新, 这样就使得在状态变量随着优化在更新的时候, 状态的观测也在不断地向着更准确的方向更新, 使得整个观测模型更加精准, 这也是预积分工作的最主要贡献。

2.3 预积分项的误差模型

传统的迭代式 IMU 积分也可以通过不断地重新积分来不断地使观测模型变精准，但是每次重新积分的计算代价太大，一般实际应用中意义不大。而预积分的基本操作就是首先假设两帧之间的偏移量为定值，即 $\mathbf{b}_i = \mathbf{b}_{i+1} = \dots = \mathbf{b}_{j-1}$ ，然后将式(2.10)的普通 IMU 积分项认为是与预积分项-白噪声带来的扰动量之和：

$$\begin{aligned}\Delta \mathbf{R}_{ij} &\doteq \Delta \tilde{\mathbf{R}}_{ij}(\mathbf{b}_i^g) \exp(-\delta \phi_{ij}) \\ \Delta \mathbf{v}_{ij} &\doteq \Delta \tilde{\mathbf{v}}_{ij}(\mathbf{b}_i^g, \mathbf{b}_i^a) - \delta \mathbf{v}_{ij} \\ \Delta \mathbf{p}_{ij} &\doteq \Delta \tilde{\mathbf{p}}_{ij}(\mathbf{b}_i^g, \mathbf{b}_i^a) - \delta \mathbf{p}_{ij}\end{aligned}\quad (2.11)$$

最终可以得到 IMU 预积分项 $[\Delta \tilde{\mathbf{R}}_{ij}, \Delta \tilde{\mathbf{v}}_{ij}, \Delta \tilde{\mathbf{p}}_{ij}]$ 和预积分的误差项 $\mathbf{n}_{ij} \doteq [\delta \phi_{ij}, \delta \mathbf{v}_{ij}, \delta \mathbf{p}_{ij}]$ ，并且认为这个 9 维的误差也是符合均值为零的高斯分布的。记它的协方差为 Σ_{ij} ，即 $\mathbf{n}_{ij} \sim \mathcal{N}(\mathbf{0}, \Sigma_{ij})$ 。而这个协方差，显然是在预积分的过程中通过 IMU 读数的白噪声累积得来的。因此需要在预积分的过程中根据白噪声的协方差 $\mathbf{n} \doteq [\mathbf{n}^g, \mathbf{n}^a] \sim \mathcal{N}(\mathbf{0}, \mathbf{n})$ 同时也完成白噪声的传递。限于篇幅，白噪声的传递的推导，这里不再给出。

2.4 偏移量的线性修正

上面提到，预积分技术使用线性的方法将每轮迭代时偏移量增量更新到预积分项中去，作为其修正。因此需要先将偏移量写成一个偏移量初值加迭代小增量的形式：

$$\begin{aligned}\hat{\mathbf{b}}_i^g &= \bar{\mathbf{b}}_i^g + \delta \mathbf{b}_i^g \\ \hat{\mathbf{b}}_i^a &= \bar{\mathbf{b}}_i^a + \delta \mathbf{b}_i^a\end{aligned}\quad (2.12)$$

记 $\Delta \bar{\mathbf{R}}_{ij} \doteq \Delta \tilde{\mathbf{R}}_{ij}(\bar{\mathbf{b}}_i^g)$, $\Delta \bar{\mathbf{v}}_{ij} \doteq \Delta \tilde{\mathbf{v}}_{ij}(\bar{\mathbf{b}}_i^g, \bar{\mathbf{b}}_i^a)$, $\Delta \bar{\mathbf{p}}_{ij} \doteq \Delta \tilde{\mathbf{p}}_{ij}(\bar{\mathbf{b}}_i^g, \bar{\mathbf{b}}_i^a)$ ，按如下方式进行线性修正：

$$\begin{aligned}\Delta \tilde{\mathbf{R}}_{ij}(\hat{\mathbf{b}}_i^g) &\doteq \Delta \tilde{\mathbf{R}}_{ij}(\bar{\mathbf{b}}_i^g + \delta \mathbf{b}_i^g) \simeq \Delta \bar{\mathbf{R}}_{ij} \exp\left(\frac{\partial \Delta \bar{\mathbf{R}}_{ij}}{\partial \mathbf{b}_i^g} \delta \mathbf{b}_i^g\right) \\ \Delta \tilde{\mathbf{v}}_{ij}(\hat{\mathbf{b}}_i^g, \hat{\mathbf{b}}_i^a) &\doteq \Delta \tilde{\mathbf{v}}_{ij}(\bar{\mathbf{b}}_i^g + \delta \mathbf{b}_i^g, \bar{\mathbf{b}}_i^a + \delta \mathbf{b}_i^a) \simeq \Delta \bar{\mathbf{v}}_{ij} + \frac{\partial \Delta \bar{\mathbf{v}}_{ij}}{\partial \mathbf{b}_i^g} \delta \mathbf{b}_i^g + \frac{\partial \Delta \bar{\mathbf{v}}_{ij}}{\partial \mathbf{b}_i^a} \delta \mathbf{b}_i^a \\ \Delta \tilde{\mathbf{p}}_{ij}(\hat{\mathbf{b}}_i^g, \hat{\mathbf{b}}_i^a) &\doteq \Delta \tilde{\mathbf{p}}_{ij}(\bar{\mathbf{b}}_i^g + \delta \mathbf{b}_i^g, \bar{\mathbf{b}}_i^a + \delta \mathbf{b}_i^a) \simeq \Delta \bar{\mathbf{p}}_{ij} + \frac{\partial \Delta \bar{\mathbf{p}}_{ij}}{\partial \mathbf{b}_i^g} \delta \mathbf{b}_i^g + \frac{\partial \Delta \bar{\mathbf{p}}_{ij}}{\partial \mathbf{b}_i^a} \delta \mathbf{b}_i^a\end{aligned}\quad (2.13)$$

这样，在每轮优化迭代的时候，需要同时使用式(2.13)，将新的偏移量更新到 IMU 预积分结果里。不过，当必要的时候，比如偏移量的增量非常大，那么线性更新可能精度也是不够的，此时还是可以采用重新积分的方式去更新预积分项，以减少误差累积。

2.5 非线性优化

基于迭代式 IMU 积分的 VISLAM 中, IMU 项参与优化的方式一般比较直接——使用 IMU 积分得到状态先验, 再和视觉观测融合, 反复迭代。例如 OKVIS 中的非线性优化 (如图 2.2 是 OKVIS 优化的因子图)。

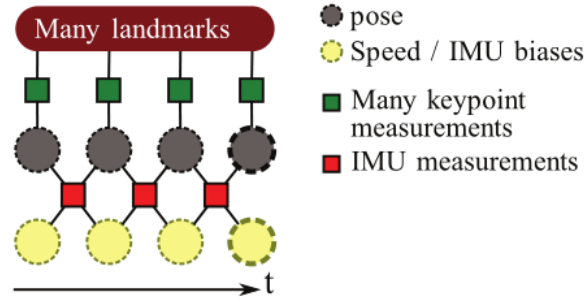


图 2.2 OKVIS 因子图示例图 [6]

和大部分的 SLAM 系统一样, IMU 预积分技术的后端优化也是基于非线性最小二乘。它的能量函数同样包括基本的视觉观测项和 IMU 预积分项 (见图 2.3), 形式上没有区别, 区别仅在 IMU 部分的能量:

$$\mathcal{X}^* = \arg \min_{\mathcal{X}} \sum_{i,j} \|\mathbf{r}_{\mathcal{I}_{ij}}\|_{ij}^2 + \sum_i \sum_l \|\mathbf{r}_{\mathcal{C}_{il}}\|_c^2 \quad (2.14)$$

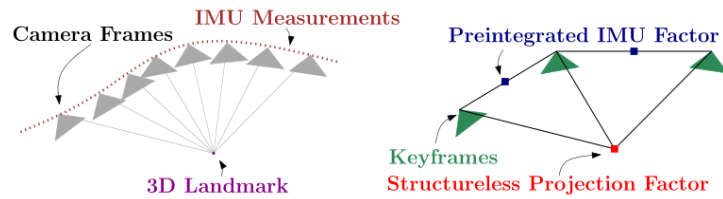


图 2.3 IMU 预积分因子图示例 [7]

其中 $\mathbf{r}_{\mathcal{C}_{il}}$ 是关于状态 i 和路标点 l 的视觉部分目标函数, 这一部分的定义和大部分 SLAM 系统一致。 $\mathbf{r}_{\mathcal{I}_{ij}}$ 则与 MSCKF、OKVIS 等系统不同, 是关于相邻状态 i 和 j 的 IMU 预积分目标函数。在 IMU 预积分技术中, 这一项被认为是仅和状态 i 的偏移量相关的函数: $\Delta \tilde{\mathbf{R}}_{ij}(\mathbf{b}_i^g)$ 、 $\Delta \tilde{\mathbf{v}}_{ij}(\mathbf{b}_i^g, \mathbf{b}_i^a)$ 和 $\Delta \tilde{\mathbf{p}}_{ij}(\mathbf{b}_i^g, \mathbf{b}_i^a)$ 。前面已经给出了 IMU 预积分项的误差模型, 因此可以很容易地写出 IMU 预积分项的目标函数:

$$\begin{aligned}
\mathbf{r}_{\Delta R} &\doteq \log \left(\left(\Delta \bar{\mathbf{R}}_{ij} \exp \left(\frac{\partial \Delta \bar{\mathbf{R}}_{ij}}{\partial \mathbf{b}_i^g} \delta \mathbf{b}_i^g \right) \right) \mathbf{R}_i^\top \mathbf{R}_j \right) \\
\mathbf{r}_{\Delta v} &\doteq \mathbf{R}_i^\top (\mathbf{v}_j - \mathbf{v}_i - \mathbf{g} \Delta t_{ij}) - \left[\Delta \bar{\mathbf{v}}_{ij} + \frac{\partial \Delta \bar{\mathbf{v}}_{ij}}{\partial \mathbf{b}_i^g} \delta \mathbf{b}_i^g + \frac{\partial \Delta \bar{\mathbf{v}}_{ij}}{\partial \mathbf{b}_i^a} \delta \mathbf{b}_i^a \right] \\
\mathbf{r}_{\Delta p} &\doteq \mathbf{R}_i^\top (\mathbf{p}_j - \mathbf{p}_i - \mathbf{v}_i \Delta t_{ij} - \frac{1}{2} \mathbf{g} \Delta t_{ij}^2) - \left[\Delta \bar{\mathbf{p}}_{ij} + \frac{\partial \Delta \bar{\mathbf{p}}_{ij}}{\partial \mathbf{b}_i^g} \delta \mathbf{b}_i^g + \frac{\partial \Delta \bar{\mathbf{p}}_{ij}}{\partial \mathbf{b}_i^a} \delta \mathbf{b}_i^a \right]
\end{aligned} \tag{2.15}$$

得到以上所示的 IMU 预积分目标函数, 即可将 IMU 部分的观测和 VSLAM 中常用重投影误差目标函数、光度误差目标函数等进行联合优化, 实现 VISLAM。

2.6 本章小结

本章详细介绍了 IMU 传感器的基本模型, 包括噪声模型和运动模型, 并给出了基于 IMU 预积分方法的 VISLAM 基本原理。如前面提到的, 由 IMU 预积分提供的观测可以和一般常用的 VSLAM 的视觉观测进行联合优化。本文提出的增量式舒尔补基数调整提供了基于 IMU 预积分的目标函数将连同重投影误差目标函数的实现。

第3章 基于增量式集束调整的高效状态估计

iSAM2 算法^[1;3] 创新性地在 SLAM 问题中使用了贝叶斯树来编码集束调整过程中的信息矩阵分解过程，达到高效地更新平方根信息矩阵的目的，显著地提升了全局优化的性能。其算法不仅适用于 SLAM 中的集束调整问题，也适用于一般的非线性最小二乘问题。也正是为了保证通用性，iSAM2 难以在状态之间的关联特点高度统一的 SLAM 问题中发挥更高的性能。

例如在图 1.3 中，正规方程的三维点部分（左上角部分）高度稀疏，呈对角块状，在求解时应该优先考虑这一部分的分解。iSAM2 算法需要依赖 COLAMD^[34] 算法来被动地检测矩阵分解的顺序，一方面需要额外的计算时间，另一方面也不一定能得到比经验更好的结果。

3.1 增量舒尔补算法

舒尔补虽然可以大幅加速集束调整的求解，但是在标准的批量式集束调整算法（batch bundle adjustment）中，每一轮迭代重新构建舒尔补方程的过程仍然需要耗费大量的时间。如前面提到的，SLAM 问题具有很好的局部性，每一轮求解一般都只有一小部分较新的变量有所更新，如果采用增量式的方法，每一轮迭代时只重新计算这些变量在舒尔补方程中对应的部分，则可以进一步减少计算量。

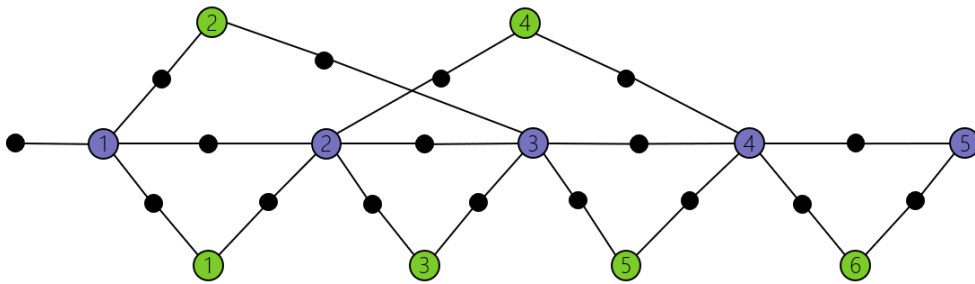


图 3.1 因子图

如图 3.1 是一个常见集束调整问题的因子图示例，其对应的正规方程具有如图 3.2 所示的稀疏结构。按照式(1.35)构建相机部分方程的舒尔补，依次迭代计算每一个三维点状态对应的舒尔补部分，并累加到图 3.3 中的高亮部分所示的舒尔补中。图 3.4 展示了计算舒尔补的第一次迭代。

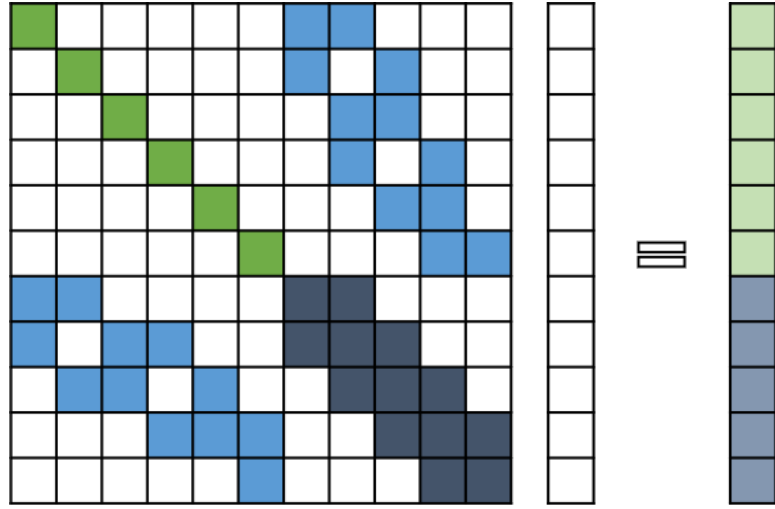


图 3.2 正规方程：左上角块对角矩阵部分对应三维点状态 1 至 6 的信息矩阵 P ，右下角部分为相机状态 1 至 5 对应的信息矩阵 C ，左下角和右上角为矩阵 W 和 W^T ；中间列向量从上至下依次对应三维点状态 1 至 6 和相机状态 1 至 5 的迭代步；右侧部分从上至下依次对应三维点状态 1 至 6 和相机状态 1 至 5。

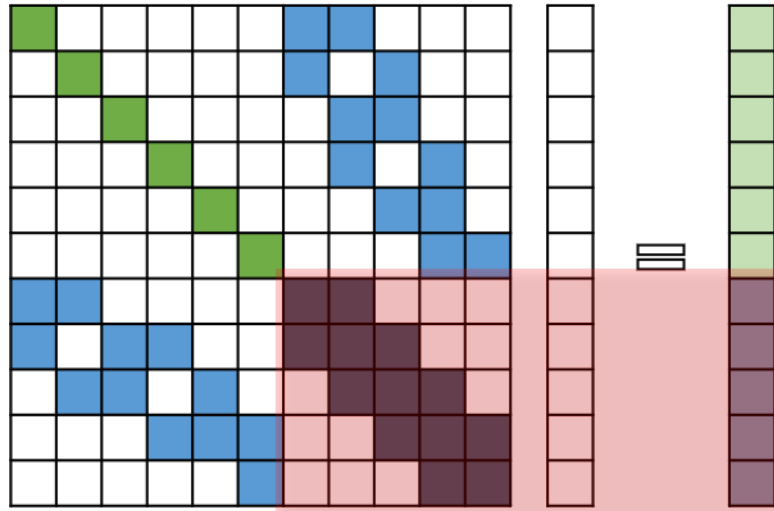


图 3.3 舒尔补：高亮部分为相机状态对应舒尔补方程，其余部分为原方程。

3.2 标记脏因子图

增量更新舒尔补的方法关键在于尽可能地复用上一次迭代的结果，只针对性地计算舒尔补方程中变化较大的变量对应的部分。以因子图 3.1 为例，假设只有相机状态 5 有较大的更新，则先将其标记为脏变量。在因子图中，只有与脏变量直接相连的因子需要标记为脏因子，如图 3.5 所示。需要注意的是，尽管相机状态 5 和三维点状态 6 的值并未发生足够大的改变，但是由于与之相连的部分因子被标记为脏因子，因此也需要更新这些因子关

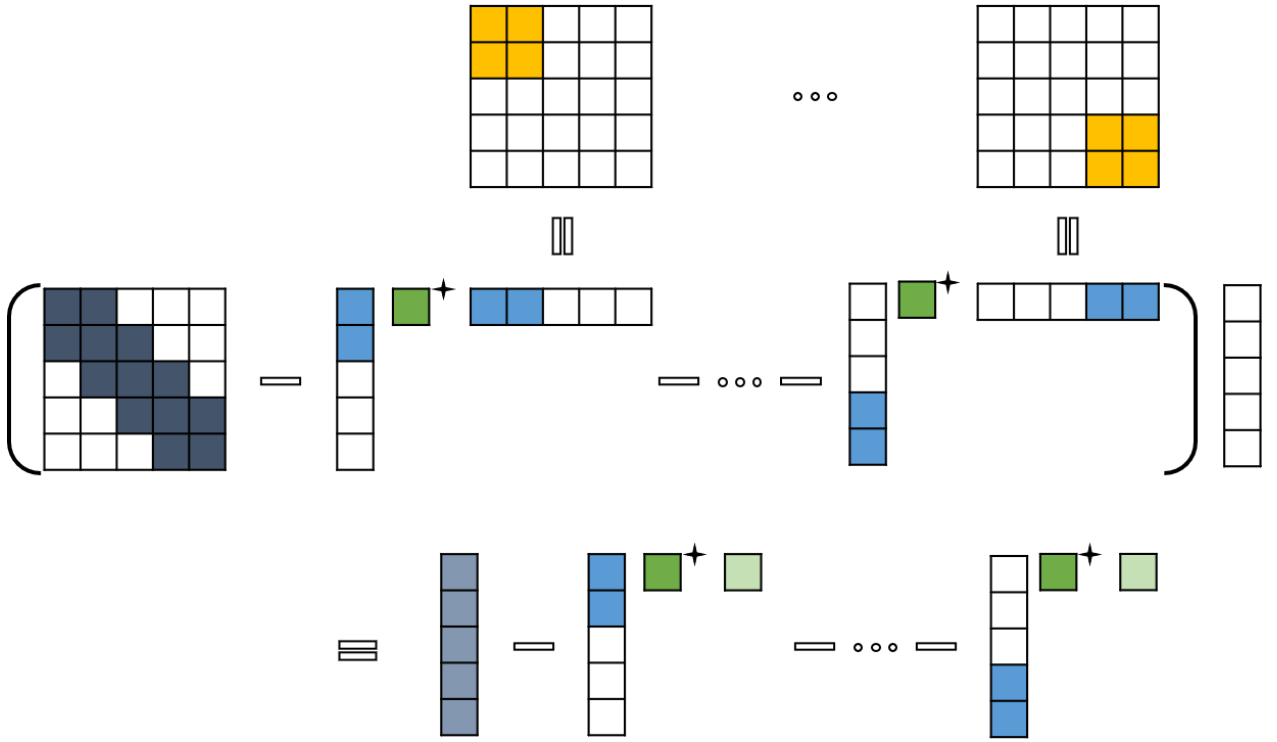


图 3.4 计算舒尔补的一次迭代：从左上角开始依次迭代计算每一个三维点对应的舒尔补部分，并累加到右下角相机部分中。

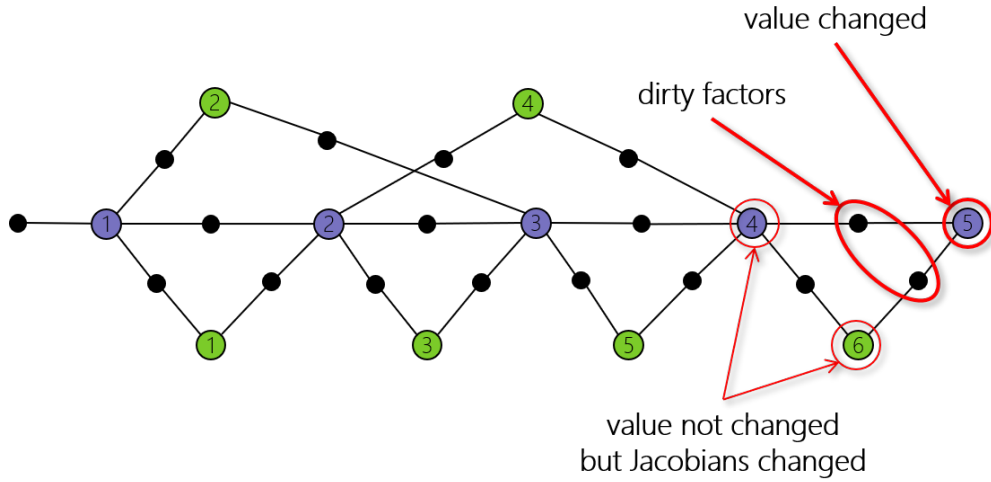


图 3.5 因子图待更新部分：相机状态 5 为脏变量；与其直接相连的因子为脏因子；相机状态 4 和三维点状态 6 为只有梯度发生了变化的脏变量。这些节点构成了一个脏子图。

于它们的雅各比矩阵。这些脏变量和脏因子共同组成了一个脏子图，对应地，图 3.6 高亮标示了舒尔补方程中需要更新的脏块。算法 2 详细说明了如何通过一个简单的宽度优先搜索标记出因子图中的脏子图。

算法 2 标记脏子图

输入: 因子图 $\tilde{\mathcal{G}} = \{\tilde{\mathcal{F}}, \tilde{\Theta}, \tilde{\mathcal{E}}\}$, 迭代步长 δ

输出: 脏子图 $\mathcal{G} = \{\mathcal{F}, \Theta, \mathcal{E}\}$

$\mathcal{F} := \{\}, \Theta := \{\}, \mathcal{E} := \{\}$

for all $\|\delta_i\|_\infty \geq \varepsilon$ **do**

$\Theta := \Theta \cup \{\theta_i\}$

{将变量 θ_i 标记为脏变量}

for all $\{f_j \in \tilde{\mathcal{F}} | \theta_i \in \Theta_j\}$ **do**

{ Θ_j 为与 f_j 相连的变量集合}

$\Theta := \Theta \cup \Theta_j$

$\mathcal{E} := \mathcal{E} \cup \{e_{ij}\}$

{记录变量 θ_i 与因子 f_j 相连的边 e_{ij} }

$\mathcal{F} := \mathcal{F} \cup \{f_j\}$

{将因子 f_j 标记为脏因子}

end for

end for

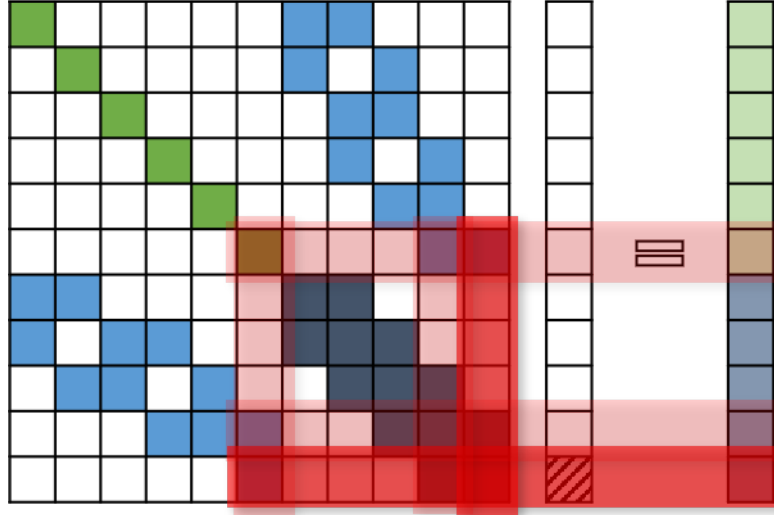


图 3.6 图 3.5 对应的舒尔补方程中的脏块: 相机状态 4、5 和三维点状态 6 构成的线性子系统。

3.3 增量更新舒尔补

舒尔补方程中被标记为脏块的矩阵构成了一个线性子系统, 对应因子图中的脏子图。如图 3.7 所展示的, 在增量更新舒尔补的过程中, 需要先从舒尔补方程中减去脏子图的贡献, 随后重新线性化所有的脏因子, 最后重新计算舒尔补。算法 3 详细描述了这个过程。需要注意的是, 在计算舒尔补的过程中会重复用到 $W_i P_{ii}^{-1} W_i^\top$, $W_i P_{ii}^{-1} p_i$, $J_j^\top J_j$, $J_j^\top f_j$, 需要将这些中间项缓存下来, 避免重复计算。

算法 3 增量更新舒尔补

输入: 脏子图 $\mathcal{G} = \{\mathcal{F}, \Theta, \mathcal{E}\}$, 上一轮迭代的 δ , H , η , S , b **输出:** 更新后的 H , η , S , b **for all** 三维点状态 $\theta_i \in \Theta$ **do** {减去三维点对舒尔补的贡献}

$$S := S + W_i P_{ii}^{-1} W_i^\top$$

$$b := b + W_i P_{ii}^{-1} p_i$$

end for**for all** 脏因子 $f_j \in \mathcal{F}$ **do** {更新脏因子对正规方程的贡献}

$$H := H - J_j^\top J_j$$

$$\eta := \eta + J_j^\top f_j$$

$$J_j := \frac{\partial f_j(\Theta_j)}{\partial \Theta_j}$$

$$f_j := f_j(\Theta_j)$$

$$H := H + J_j^\top J_j$$

$$\eta := \eta - J_j^\top f_j$$

end for**for all** 三维点状态 $\theta_i \in \Theta$ **do** {更新三维点对舒尔补的贡献}

$$S := S - W_i P_{ii}^{-1} W_i^\top$$

$$b := b - W_i P_{ii}^{-1} p_i$$

end for

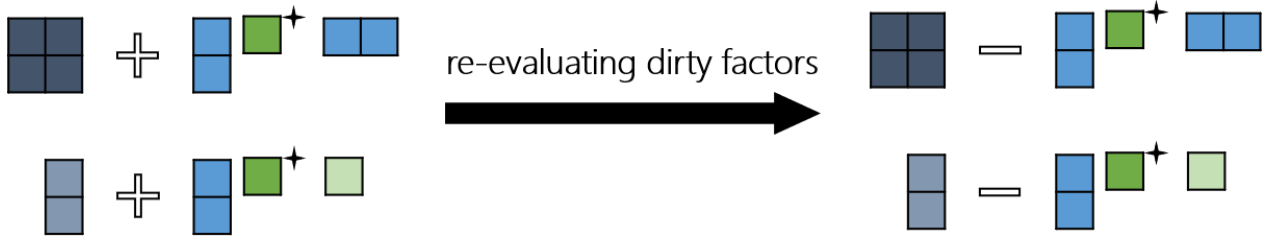


图 3.7 更新舒尔补

3.4 状态增广

除了状态变量发生较大变化的时候需要更新因子图和舒尔补方程, 集束调整过程中, 如果有新的状态或因子加入, 也需要对因子图进行增广, 并对舒尔补方程进行相应更新。实际上, 这个过程也可以认为是算法 3 的一个特例。如图 3.8, 新加入的状态通过新的因子与原因子图中相关联的部分状态共同构成了一个子图, 将这个子图标记为脏子图, 即可使用算法 3 进行更新。算法 4 给出了详细的过程, 为了提高效率, 因子图的增广和舒尔补方程的更新会被同时执行。

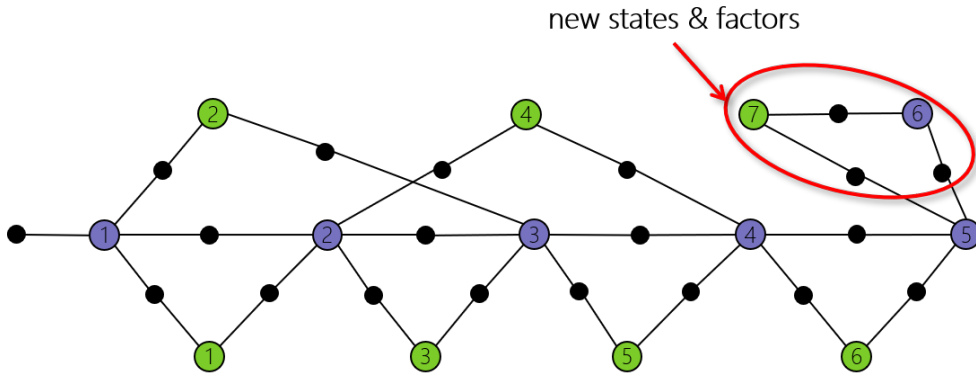


图 3.8 增广因子图

3.5 线性求解

通过增量更新得到舒尔补方程(1.36)后, 可以快速地求解相机状态的增量。这一部分可以使用任何常用的矩阵分解法, 如 QR 分解、Cholesky 分解等。也可以采用迭代式的线性求解方法, 比如预处理共轭梯度法 (preconditioned conjugate gradient, 以下简称 PCG)。通常在 SLAM 问题中, 由于变量大多已经处于收敛附近, 使用 PCG 方法并配合^[38]中提出的预处理方法, 可以显著提升线性求解的效率。求解完相机状态增量 δ_c 后, 还需要进一

算法 4 状态增广

输入: 因子图 $\tilde{\mathcal{G}} = \{\tilde{\mathcal{F}}, \tilde{\Theta}, \tilde{\mathcal{E}}\}$, 新加入的因子和状态 $\{\mathcal{F}_0, \Theta_0\}$, 上一次优化的 $H, \boldsymbol{\eta}, S, \mathbf{b}$

输出: 更新后的 $H, \boldsymbol{\eta}, S, \mathbf{b}$

for all 三维点状态 $\boldsymbol{\theta}_i \in \Theta_0$ do {增广正规方程, 如图 3.9}

增广矩阵: P, W, \mathbf{p}

end for

for all 相机状态 $\boldsymbol{\theta}_i \in \Theta_0$ do

增广矩阵: $C, S, W, \mathbf{c}, \mathbf{b}$

end for

$\Theta := \{\}$ {更新正规方程和舒尔补, 如图 3.10}

for all $\{\mathbf{f}_j \in \mathcal{F}_0 | \Theta_j \cap \tilde{\Theta} \neq \emptyset\}$ do

$\Theta := \Theta \cup (\Theta_j \cap \tilde{\Theta})$

end for

for all 三维点状态 $\boldsymbol{\theta}_i \in \Theta$ do {减去三维点对舒尔补的贡献}

$$S := S + W_i P_{ii}^{-1} W_i^\top$$

$$\mathbf{b} := \mathbf{b} + W_i P_{ii}^{-1} \mathbf{p}_i$$

end for

for all 脏因子 $\mathbf{f}_j \in \mathcal{F}_0$ do {更新新因子对正规方程的贡献}

$$J_j := \frac{\partial \mathbf{f}_j(\boldsymbol{\theta}_j)}{\partial \boldsymbol{\theta}_j}$$

$$\mathbf{f}_j := \mathbf{f}_j(\boldsymbol{\theta}_j)$$

$$H := H + J_j^\top J_j$$

$$\boldsymbol{\eta} := \boldsymbol{\eta} - J_j^\top \mathbf{f}_j$$

end for

for all 三维点状态 $\boldsymbol{\theta}_i \in \Theta$ do {更新三维点对舒尔补的贡献}

$$S := S - W_i P_{ii}^{-1} W_i^\top$$

$$\mathbf{b} := \mathbf{b} - W_i P_{ii}^{-1} \mathbf{p}_i$$

end for

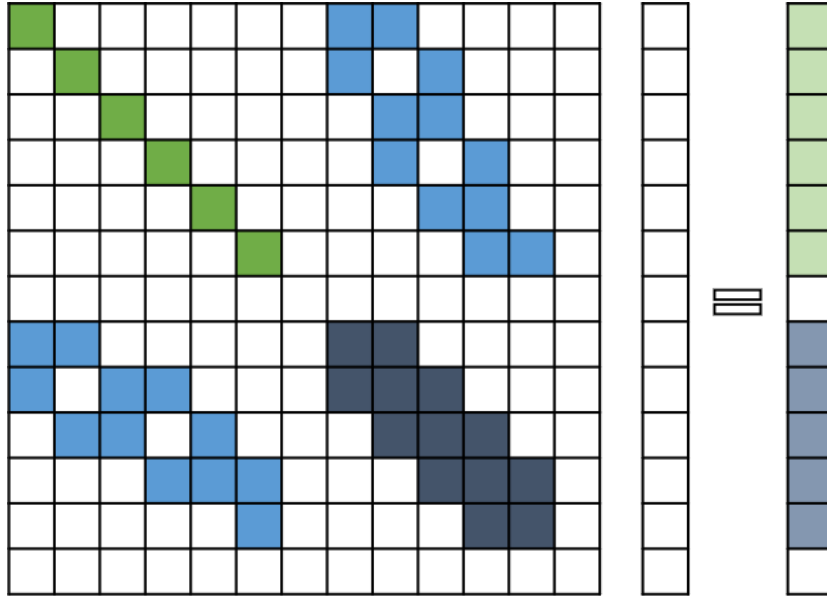


图 3.9 增广正规方程

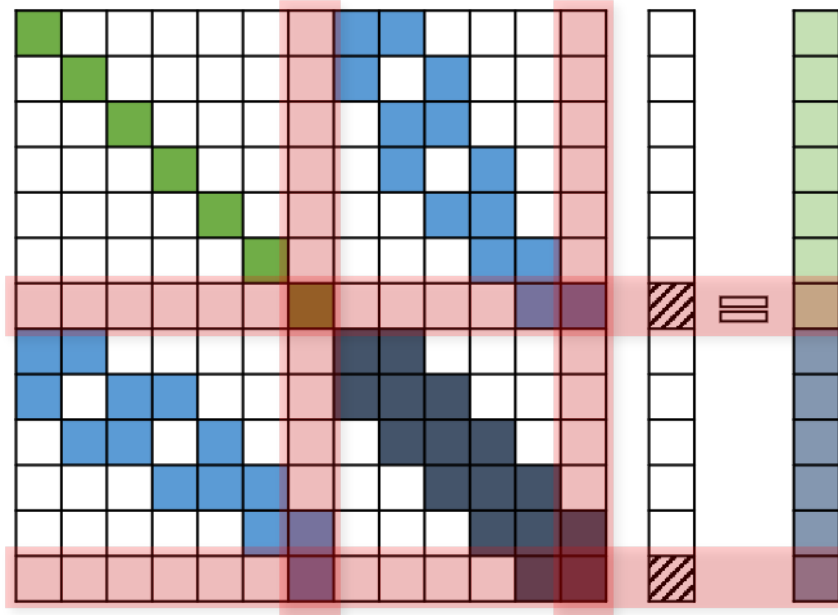


图 3.10 更新正规方程

步通过回代求解三维点状态的增量 δ_p ，如式(1.37)。最后得到的总的状态增量 δ ，通过阈值判断其每个分量 δ_i 是否发生了足够大的更新并将发生较大更新的分量更新到状态中：

$$\mathbf{x}_i \leftarrow \mathbf{x}_i + \delta_i, \quad \forall \|\delta_i\|_\infty \geq \varepsilon \quad (3.1)$$

进一步地，如果还需要进行下一轮迭代，则将更新量 δ 传递给算法 2，计算新一轮的迭代。

整个增量舒尔补和线性求解的过程均可以基于块状稀疏矩阵实现。相比与稠密矩阵或

一般的稀疏矩阵算法，块状稀疏矩阵的计算过程对 cache 更友好，也更适合使用 CPU 浮点指令集加速。得益于块状稀疏矩阵的高效，整个增量式集束调整的效率相比批量式集束调整可以有非常大的提升。

3.6 本章小结

本章详细介绍了基于增量式舒尔补的集束调整优化算法，一组增量式集束调整优化的流程总结如下：

1. 根据因子图，采用传统方法构建舒尔补并使用 Cholesky 分解或 PCG 方法求解首轮迭代；
2. 使用算法 2 标记因子图和舒尔补方程需要更新的部分；
3. 使用算法 3 更新舒尔补；
4. 使用 Cholesky 分解或 PCG 方法求解舒尔补方程；
5. 重复2至4直到收敛；
6. 当新增了因子或状态时，使用算法 4 进行增广，并重新从2开始新一组优化

第4章 实验和结论

本章将通过模拟数据和真实数据，测试本文提出的增量式集束调整算法和传统批量式集束调整算法的速度和精度。Ceres-Solver 是 Google 为了求解 SfM 问题而推出的非线性最小二乘求解器，经过多年的发展，Ceres-Solver 已经广泛地应用在多个 SLAM 系统中，包括 OKVIS、VINS-Mono 等，并已成为开源社区中最为流行的通用非线性最小二乘求解器之一。作为一个通用的求解器，Ceres-Solver 提供了稠密版本和稀疏版本的多种批量式数值优化算法实现（包括莱文贝格-马夸特法、Dog-Leg 法等），具有极高的数值稳定性和较高的效率。

4.1 测试设置

测试数据包含模拟生成的相机-IMU-三维点的数据集 Cubic 和包含相机-三维点的 VSLAM 真实数据集 Cathedral^[39] 和 Venice^[12]。其中 Cubic 数据集为模拟生成的带高斯噪声的 VSLAM 数据集，包含 IMU 预积分的结果和协方差、相机状态的初值、视觉观测。Cathedral 和 Venice 为真实数据集，包含相机状态的初值、三维点状态的初值和视觉观测。这样的设置省去了 SLAM 前端视觉匹配、IMU 积分的过程，保证所有算法求解的问题相同。几组数据具体包含的内容如表 4.1 所示，由于 Ceres-Solver 求解完整的 Venice 数据集失败，顾只取了该数据集前 10 个相机状态，大约包含 8000 个三维点。

表 4.1 测试数据集明细

数据集	相机数量	三维点数量	IMU 目标函数	重投影误差目标函数
Cubic	50	163	49	789
Venice	11	8366	0	20271
Cathedral	92	57957	0	422163

所有测试均基于最新的 Arch Linux 操作系统，内核版本为 4.19.12，搭载主频为 3.6GHz 的 Intel Core i7-7700 处理器，内存为 8GB。运行测试期间关闭其他所有应用程序，每一组测试的结果均取 10 次运行后的平均值。

4.2 结果对比

为了保证公平,所有求解器尽可能地设置成相同的算法,并将算法运行时间 t_{solve} 定义为总运行时间减去冷启动所用时间。所有优化器均使用稀疏求解策略,本文的优化策略非线性部分为 Dog-Leg 法,线性部分为增量舒尔补-正规 Cholesky 分解/增量舒尔补-正规 PCG 迭代法。Ceres-Solver 的优化策略非线性部分也设置为 Dog-Leg 法,由于 Ceres-Solver 的限制,线性部分只能设置为舒尔补-雅各比 QR 分解法。将所有优化算法的迭代次数设置为相同,通过运行时间 T_{solve} 、迭代次数 Iter 和优化结束时的重投影误差 Err_{reproj} 和相对位姿约束误差 Err_{IMU} 来对比效率和精度。

表 4.2 测试结果: 运行时间和迭代次数

$T_{\text{solve}}/\text{Iter.}$	Cubic	Venice	Cathedral
基准 Ceres: Schur-QR	69ms/5 100.0%	125ms/6 100.0%	3860ms/5 100.0%
本文算法: Schur-Cholesky	17ms/5 24.6%	78ms/6 62.4%	2334ms/5 60.5%
本文算法: Schur-PCG	18ms/5 26.1%	80ms/6 64.0%	2423ms/5 62.8%

表 4.3 测试结果: 收敛时的误差

$Err_{\text{reproj}}/Err_{\text{reproj}} + Err_{\text{IMU}}$ (优化前误差)	Cubic (2.514×10^3)	Venice (5.641×10^{-1})	Cathedral (7.475×10^0)
基准 Ceres: Schur-QR	5.153×10^2	2.996×10^{-2}	4.153×10^{-1}
本文算法: Schur-Cholesky	4.982×10^2	3.166×10^{-2}	4.246×10^{-1}
本文算法: Schur-PCG	5.004×10^2	2.996×10^{-2}	5.524×10^{-1}

4.3 本章小结

本章通过在模拟 VISLAM 数据集 Cubic 和开源 VSLAM 数据集上测试对比了本文提出的增量式舒尔补集束调整优化算法和流行的开源 Ceres-Solver 的求解效率和精度。可以看到在各个数据集上,相比开源的 Ceres-Solver,本文提出的基于增量舒尔补的集束调整方法均在保证一定精度的前提下将求解时间减少了 75.4% 至 36.0% 不等,显著提升了效率。

值得注意的是,本文提出的集束调整算法仍然具有较高的通用性和可扩展性:除了集束调整问题,也可以求解一般的非线性最小二乘问题;除了提供的 Cholesky 分解法和 PCG 方法,用户也可以自定义新的线性求解方法,以供不同场景使用。

第5章 总结与展望

5.1 总结

在 SLAM 算法中，状态估计算法的速度和精度一定程度上决定了 SLAM 应用的可用性。传统的批量式集束调整求解器为了保证通用性，牺牲了一定的效率，而现有的基于增量式集束调整的求解器虽然效率很高，但由于耦合程度过高，难以在其他 SLAM 系统中使用。本文基于以上分析，提出了基于增量式集束调整的通用求解器，在保证效率和精度的基础上提供了很好的通用性，可以很容易地集成到其他 SLAM 系统中。

在非线性优化部分，提供了 IMU 预积分的目标函数接口和重投影误差目标函数接口，同时保留扩展性，用户可以直接使用内置的目标函数，也可以自定义目标函数；在线性部分，使用了增量舒尔补的方法大幅提升了线性化过程的效率；并基于块状稀疏矩阵，提供了 Cholesky 分解法、预处理共轭梯度法等常用的线性求解接口，同时用户也可以定制专用的线性求解器。最后，在模拟数据和 SLAM++ 提供的集束调整数据集上进行了测试，通过实验验证了求解器的效率和精度。

5.2 展望

增量式集束调整利用了 SLAM 问题的稀疏性和局部性，极大地提升了效率。但是仍有一些情况会降低增量式集束调整的效率：

1. 在发生回路闭合的时候，通常大量的状态变量都需要进行更新，增量式集束调整会退化到普通的批量式集束调整方法，效率严重下降；
2. 状态边缘化通常会影响问题的稀疏性，频繁的边缘化操作则会导致系统变得十分稠密，不利于基于块状稀疏矩阵的求解算法。

如何在集束调整过程中避免以上问题，或是在遇到以上问题的时候保持计算效率和精度，仍是一个很大的挑战。

参考文献

- [1] Michael Kaess, Hordur Johannsson, Richard Roberts, Viorela Ila, John J Leonard, Frank Dellaert. isam2: Incremental smoothing and mapping using the bayes tree[J]. The International Journal of Robotics Research, 2012, 31(2):216–235.
- [2] O Tingleff, K Madsen, HB Nielsen. Methods for non-linear least squares problems[J]. Lecture Note in Computer Science 02611 Optimization and Data Fitting, 2004.
- [3] Michael Kaess, Ananth Ranganathan, Frank Dellaert. isam: Incremental smoothing and mapping[J]. IEEE Transactions on Robotics, 2008, 24(6):1365–1378.
- [4] Peiliang Li, Tong Qin, Botao Hu, Fengyuan Zhu, Shaojie Shen. Monocular visual-inertial state estimation for mobile augmented reality[C]. In 2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). IEEE, 2017, 11–21.
- [5] novatel. Imu errors and their effects[EB/OL]. <https://www.novatel.com/assets/Documents/Bulletins/APN064.pdf>, 2014.
- [6] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, Paul Furgale. Keyframe-based visual-inertial odometry using nonlinear optimization[J]. The International Journal of Robotics Research, 2015, 34(3):314–334.
- [7] Christian Forster, Luca Carlone, Frank Dellaert, Davide Scaramuzza. On-manifold preintegration for real-time visual-inertial odometry[J]. IEEE Transactions on Robotics, 2017, 33(1):1–21.
- [8] Richard Hartley, Andrew Zisserman. Multiple view geometry in computer vision[M]. Cambridge university press, 2003.
- [9] Yi Ma, Stefano Soatto, Jana Kosecka, S Shankar Sastry. An invitation to 3-d vision: from images to geometric models[M], volume 26. Springer Science & Business Media, 2012.
- [10] Ji Zhang, Sanjiv Singh. Visual-lidar odometry and mapping: Low-drift, robust, and fast[C]. In Robotics and Automation (ICRA), 2015 IEEE International Conference on. IEEE, 2015, 2174–2181.
- [11] Sameer Agarwal, Keir Mierle, Others. Ceres solver[EB/OL]. <http://ceres-solver.org>, 2012.
- [12] Rainer Kümmerle, Giorgio Grisetti, Hauke Strasdat, Kurt Konolige, Wolfram Burgard. g 2 o: A general framework for graph optimization[C]. In Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011, 3607–3613.
- [13] Viorela Ila, Lukas Polok, Marek Solony, Klemen Istenic. Fast incremental bundle adjustment with covariance recovery[C]. In International Conference on 3D Vision. 2017, 4321–4330.

- [14] Bill Triggs, Philip F McLauchlan, Richard I Hartley, Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis[C]. In International workshop on vision algorithms. Springer, 1999, 298–372.
- [15] Feng Lu, Evangelos Milios. Globally consistent range scan alignment for environment mapping[J]. Autonomous robots, 1997, 4(4):333–349.
- [16] Feng Lu, Evangelos Milios. Robot pose estimation in unknown environments by matching 2d range scans[J]. Journal of Intelligent and Robotic systems, 1997, 18(3):249–275.
- [17] Sebastian Thrun, Michael Montemerlo. The graph slam algorithm with applications to large-scale mapping of urban structures[J]. The International Journal of Robotics Research, 2006, 25(5-6):403–429.
- [18] Jorge Nocedal, Stephen J. Wright. Numerical Optimization[M]. New York, NY, USA: Springer, 2006, 2nd edition.
- [19] Zhengyou Zhang. Parameter estimation techniques: A tutorial with application to conic fitting[D]. [博士学位论文], INRIA, 1995.
- [20] Timothy D Barfoot. State Estimation for Robotics[M]. Cambridge University Press, 2017.
- [21] Andrew J Davison, Ian D Reid, Nicholas D Molton, Olivier Stasse. Monoslam: Real-time single camera slam[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2007, (6):1052–1067.
- [22] Anastasios I Mourikis, Stergios I Roumeliotis. A multi-state constraint kalman filter for vision-aided inertial navigation[C]. In Proceedings 2007 IEEE International Conference on Robotics and Automation. IEEE, 2007, 3565–3572.
- [23] Mingyang Li, Anastasios I Mourikis. Improving the accuracy of ekf-based visual-inertial odometry[C]. In Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012, 828–835.
- [24] Guoquan P Huang, Anastasios I Mourikis, Stergios I Roumeliotis. Analysis and improvement of the consistency of extended kalman filter based slam[C]. In Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on. IEEE, 2008, 473–479.
- [25] Guoquan P Huang, Anastasios I Mourikis, Stergios I Roumeliotis. An observability-constrained sliding window filter for slam[C]. In Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on. IEEE, 2011, 65–72.
- [26] Guoquan P Huang, Anastasios I Mourikis, Stergios I Roumeliotis. A quadratic-complexity observability-constrained unscented kalman filter for slam[J]. IEEE Transactions on Robotics, 2013, 29(5):1226–1243.
- [27] Axel Barrau, Silvere Bonnabel. An ekf-slam algorithm with consistency properties[J]. arXiv preprint arXiv:1510.06263, 2015.

- [28] Georg Klein, David Murray. Parallel tracking and mapping for small ar workspaces[C]. In Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on. IEEE, 2007, 225–234.
- [29] Raul Mur-Artal, Jose Maria Martinez Montiel, Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system[J]. IEEE Transactions on Robotics, 2015, 31(5):1147–1163.
- [30] Raúl Mur-Artal, Juan D. Tardós. ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-D cameras[J]. IEEE Transactions on Robotics, 2017, 33(5):1255–1262.
- [31] Tong Qin, Peiliang Li, Shaojie Shen. Vins-mono: A robust and versatile monocular visual-inertial state estimator[J]. IEEE Transactions on Robotics, 2018, 34(4):1004–1020.
- [32] Jakob Engel, Vladlen Koltun, Daniel Cremers. Direct sparse odometry[J]. IEEE transactions on pattern analysis and machine intelligence, 2018, 40(3):611–625.
- [33] Hailin Jin, Paolo Favaro, Stefano Soatto. A semi-direct approach to structure from motion[J]. The Visual Computer, 2003, 19(6):377–394.
- [34] Timothy A Davis, John R Gilbert, Stefan I Larimore, Esmond G Ng. Algorithm 836: Colamd, a column approximate minimum degree ordering algorithm[J]. ACM Transactions on Mathematical Software (TOMS), 2004, 30(3):377–380.
- [35] Yanqing Chen, Timothy A Davis, William W Hager, Sivasankaran Rajamanickam. Algorithm 887: Cholmod, supernodal sparse cholesky factorization and update/downdate[J]. ACM Transactions on Mathematical Software (TOMS), 2008, 35(3):22.
- [36] Eagle S Jones, Stefano Soatto. Visual-inertial navigation, mapping and localization: A scalable real-time causal approach[J]. The International Journal of Robotics Research, 2011, 30(4):407–430.
- [37] MWA Smith, AP Roberts. An exact equivalence between the discrete-and continuous-time formulations of the kalman filter[J]. Mathematics and Computers in Simulation, 1978, 20(2):102–109.
- [38] Yekeun Jeong, David Nister, Drew Steedly, Richard Szeliski, In-So Kweon. Pushing the envelope of modern methods for bundle adjustment[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 34(8):1605–1617.
- [39] Hansung Kim, Adrian Hilton. Influence of colour and feature geometry on multi-modal 3d point clouds data registration[C]. In 3D Vision (3DV), 2014 2nd International Conference on. IEEE, 2014, volume 1, 202–209.