Bundle Adjustment: A tutorial

Stergios I. Roumeliotis

MARS LAB

Multiple Autonomous
Robotic Systems Laboratory

Technical Report
Number -2014-002
May 2014
Repository Revision:
Last Update:

Dept. of Computer Science & Engineering
University of Minnesota
4-192 EE/CS Building
200 Union St. S.E.
Minneapolis, MN 55455
Tel: (612) 625-2217
Fax: (612) 625-0572
URL: http://mars.cs.umn.edu

Bundle Adjustment: A tutorial

Stergios I. Roumeliotis

Multiple Autonomous Robotic Systems Laboratory, TR-2014-002

May 2014

I. Cost function derivation

Consider the state vector that will be estimated:

$$\mathbf{x} = \begin{bmatrix} \mathbf{r}_0^T & \mathbf{r}_1^T & \cdots & \mathbf{r}_i^T & \cdots & \mathbf{r}_N^T & \mathbf{f}_1^T & \mathbf{f}_2^T & \cdots & \mathbf{f}_j^T & \cdots & \mathbf{f}_M^T \end{bmatrix}^T \tag{1}$$

The proprioceptive model for the robot is given by

$$\mathbf{r}_{i+1} = \mathbf{f}(\mathbf{r}_i, \mathbf{u}_i, \mathbf{w}_i), \quad \mathbf{r}_{i+1} \in \mathbb{R}^n, \quad i = 0, 1, ..., N-1 \tag{2}$$

where $\mathbf{w}_i$ is zero-mean, white Gaussian noise with covariance $\mathbf{Q}_i$.

The exteroceptive model is given by

$$\mathbf{z}_{ij} = \mathbf{h}(\mathbf{r}_i, \mathbf{f}_j) + \mathbf{n}_{ij}, \quad \mathbf{z}_{ij} \in \mathbb{R}^m, \quad i = 1, ..., N, \quad j = 1, ..., M \tag{3}$$

where $\mathbf{n}_{ij}$ is zero-mean, white Gaussian noise with covariance $\mathbf{R}_{ij}$.

The prior is given by:

$$\mathbf{z}^o = \begin{bmatrix} \vdots \\ \widehat{\mathbf{r}}_i \\ \vdots \\ \widehat{\mathbf{f}}_j \\ \vdots \end{bmatrix} + \mathbf{n}_s \tag{4}$$

where $i \in \mathbf{S}_i, j \in \mathbf{S}_j$ for some $\mathbf{S}$; $\mathbf{n}_s \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \ddots & & & & \\ & \mathbf{P}_{r_i r_j} & & & \\ & & \ddots & & \\ & & & \mathbf{P}_{f_j f_j} & \\ & & & & \ddots \end{bmatrix} \right)$

Our objective is to find the estimate of robot poses, $\mathbf{r}_0, \mathbf{r}_1, ..., \mathbf{r}_N$ and the map $\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_M$, given measurements $\mathbf{z}_{00}, ..., \mathbf{z}_{NM}$, that maximize the posterior pdf $p(\mathbf{r}_N, ..., \mathbf{r}_1, \mathbf{r}_0, \mathbf{f}_M, ..., \mathbf{f}_2, \mathbf{f}_1 | \mathbf{z}_{NM}, \ldots, \mathbf{z}_{11})$. The posterior pdf can be written as

$$
\begin{aligned}
& p(\mathbf{r}_N,...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1|\mathbf{z}_{NM},\ldots,\mathbf{z}_{11},\mathbf{z}^o) && (5)\\
= \; & \frac{1}{p(\mathbf{z}_{NM},\ldots,\mathbf{z}_{11},\mathbf{z}^o)} p(\mathbf{z}_{NM},\ldots,\mathbf{z}_{11},\mathbf{z}^o|\mathbf{r}_N,...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1) && (6)\\
= \; & \frac{1}{c} p(\mathbf{z}_{NM}|\mathbf{r}_N,...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)\cdots p(\mathbf{z}_{ij}|\mathbf{r}_i,...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)\cdots p(\mathbf{z}_{11}|\mathbf{r}_1,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1) \\
& p(\mathbf{r}_N|\mathbf{r}_{N-1},...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)\cdots p(\mathbf{r}_1|\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)p(\mathbf{r}_0|\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)p(\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)p(\mathbf{z^o}|\mathbf{x}) && (7)\\
= \; & \frac{1}{c} p(\mathbf{z}_{NM}|\mathbf{r}_N,\mathbf{f}_M)\cdots p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j)\cdots p(\mathbf{z}_{11}|\mathbf{r}_1,\mathbf{f}_1) \\
& p(\mathbf{r}_N|\mathbf{r}_{N-1})\cdots p(\mathbf{r}_1|\mathbf{r}_0)p(\mathbf{r}_0)p(\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)p(\mathbf{z^o}|\mathbf{x}) && (8)\\
& && (9)
\end{aligned}
$$

where $c = p(\mathbf{z}_{NM},\ldots,\mathbf{z}_{11})$ is the normalization constant, and we used Bayes rule, conditional independence of measurements, and Markov assumption for the robot motion. Therefore the MAP estimator can be formulated as

$$\arg\max \quad p(\mathbf{r}_N,...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1|\mathbf{z}_{NM},\ldots,\mathbf{z}_{11},\mathbf{z}^o) \tag{10}$$

Using monotonicity of logarithmic function, and substitute (9) into (10), we get

$$
\begin{aligned}
& \arg\max \quad \log\; p(\mathbf{r}_N,...,\mathbf{r}_1,\mathbf{r}_0,\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1|\mathbf{z}_{NM},\ldots,\mathbf{z}_{11},\mathbf{z}^o) && (11)\\
(9) \;\Leftrightarrow\; & \arg\max \quad \log\; [\frac{1}{c}\prod_{i=1,j=1}^{N,M} p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j)\prod_{i=0}^{N} p(\mathbf{r}_{i+1}|\mathbf{r}_i)p(\mathbf{r}_0)p(\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1)p(\mathbf{z}^o|\mathbf{x})] && (12)\\
\Leftrightarrow\; & \arg\max \quad \log\frac{1}{c} + \sum_{i=1,j=1}^{N,M}\log p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) + \sum_{i=0}^{N}\log p(\mathbf{r}_{i+1}|\mathbf{r}_i) \\
& \qquad + \log p(\mathbf{r}_0) + \log p(\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1) + \log p(\mathbf{z}^o|\mathbf{x}) && (13)
\end{aligned}
$$

After getting rid of constant (independent of $\mathbf{x}$ or in the other ways $\mathbf{r}_i,\mathbf{f}_j \;\; \forall i,j$ from (1)) in (12), we define the cost function as follow: Then:

$$\mathbb{C}(\mathbf{x}) = -\sum_{i=1,j=1}^{N,M}\log p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) - \sum_{i=0}^{N}\log p(\mathbf{r}_{i+1}|\mathbf{r}_i) - \log p(\mathbf{r}_0) - \log p(\mathbf{z}^o|\mathbf{x}) \tag{14}$$

From (13) we have:

$$
\begin{aligned}
\mathbf{x}^{MAP} &= \arg\max \quad \log\frac{1}{c} + \sum_{i=1,j=1}^{N,M}\log p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) + \sum_{i=0}^{N}\log p(\mathbf{r}_{i+1}|\mathbf{r}_i) + \log p(\mathbf{r}_0) + \log p(\mathbf{f}_M,...,\mathbf{f}_2,\mathbf{f}_1) + \log p(\mathbf{z}^o|\mathbf{x}) && (15)\\
&= \arg\min \mathbb{C}(\mathbf{x}) && (16)
\end{aligned}
$$

We define the best estimate is the one that minimize our cost function. We will find $\mathbf{x}^{MAP}$ using the Gauss-Newton method by looping many iterations until $\mathbf{x}^{(k)}$ converges.

These next subsections describe the contribution of each factor into the cost function.

A. Exteroceptive:

$$\mathbf{z}_{ij} = \mathbf{h}(\mathbf{r}_i,\mathbf{f}_j) + \mathbf{n}_{ij}, \quad \mathbf{n}_{ij} \sim \mathcal{N}(\mathbf{0},\mathbf{R}_{ij}) \tag{17}$$

Linearization around arbitrarily point $\hat{\mathbf{r}}_i^*,\hat{\mathbf{f}}_j^*$ we have:

$$\mathbf{z}_{ij} \simeq \mathbf{h}(\hat{\mathbf{r}}_i^*,\hat{\mathbf{f}}_j^*) + \mathbf{H}_{ij}^{r\;(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^*) + \mathbf{H}_{ij}^{f\;(k)}(\mathbf{f}_j - \hat{\mathbf{f}}_j^*) + \mathbf{n}_{ij} \tag{18}$$

where $\mathbf{H}^{r}_{ij}{}^{(k)} = \frac{\partial \mathbf{h}(\mathbf{r}_i,\mathbf{f}_j)}{\partial \mathbf{r}_i}\Big|_{\mathbf{r}_i=\hat{\mathbf{r}}_i^*}$, $\mathbf{H}^{f}_{ij}{}^{(k)} = \frac{\partial \mathbf{h}(\mathbf{r}_i,\mathbf{f}_j)}{\partial \mathbf{f}_j}\Big|_{\mathbf{f}_j=\hat{\mathbf{f}}_j^*}$

Given the estimated value of the state vector $\hat{\mathbf{r}}_i^{(k)}, \hat{\mathbf{f}}_j^{(k)}$ at iteration k of Gauss-Newton and choose to linearize $\mathbf{h}(\mathbf{r}_i,\mathbf{f}_j)$ at exactly theses points we can get the first order statistics of $\mathbf{z}_{ij}$:

$$\mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j] \simeq \mathbf{h}(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)}) + \mathbf{H}^{r}_{ij}{}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) + \mathbf{H}^{f}_{ij}{}^{(k)}(\mathbf{f}_j - \hat{\mathbf{f}}_j^{(k)}) \tag{19}$$

Denote the quantity $\hat{\mathbf{z}}_{ij}^{(k)}$ as followed:

$$\hat{\mathbf{z}}_{ij}^{(k)} = \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i = \hat{\mathbf{r}}_i^{(k)}, \mathbf{f}_j = \hat{\mathbf{f}}_j^{(k)}] \simeq \mathbf{h}(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)}) + \mathbf{H}^{r}_{ij}{}^{(k)}(\hat{\mathbf{r}}_i^{(k)} - \hat{\mathbf{r}}_i^{(k)}) + \mathbf{H}^{f}_{ij}{}^{(k)}(\hat{\mathbf{f}}_j^{(k)} - \hat{\mathbf{f}}_j^{(k)}) = \mathbf{h}(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)}) \tag{20}$$

This is worth noticing that $\hat{\mathbf{z}}_{ij}^{(k)} \neq \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j]$ in general. The equality happens only when we linearize around the point $(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)})$ and we conditioned the realization $\mathbf{r}_i = \hat{\mathbf{r}}_i^{(k)}, \mathbf{f}_j = \hat{\mathbf{f}}_j^{(k)}$.

, then the second order statistics of $\mathbf{z}_{ij}$ is obtained by:

$$\mathbb{E}[(\mathbf{z}_{ij} - \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j])(\mathbf{z}_{ij} - \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j])^T|\mathbf{r}_i,\mathbf{f}_j] \simeq \mathbb{E}[\mathbf{n}_{ij}\mathbf{n}_{ij}^T] = \mathbf{R}_{ij} \tag{21}$$

due to the normal distribution of $\mathbf{n}_{ij}$ follows from (17).

Then, from (19), (21), notice that at k^{th} iteration $\mathbf{z}_{ij} \sim \mathcal{N}(\mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j],\mathbf{R}_{ij})$.

$$p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) = \frac{1}{(2\pi)^{m/2} det(\mathbf{R}_{ij})^{1/2}} e^{-\frac{1}{2}\left(\mathbf{z}_{ij} - \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j]\right)^T \mathbf{R}_{ij}^{-1}\left(\mathbf{z}_{ij} - \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j]\right)} \tag{22}$$

Denote $\delta\mathbf{z}_{ij}{}^{(k)} = \mathbf{z}_{ij} - \hat{\mathbf{z}}_{ij}^{(k)} = \mathbf{z}_{ij} - \mathbf{h}(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)})$ (residual), from (19) we have:

$$\begin{aligned}
\mathbf{z}_{ij} - \mathbb{E}[\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j] &= \mathbf{z}_{ij} - \mathbf{h}(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)}) - \mathbf{H}^{r}_{ij}{}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) - \mathbf{H}^{f}_{ij}{}^{(k)}(\mathbf{f}_j - \hat{\mathbf{f}}_j^{(k)}) && (23)\\
&= \delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}^{r}_{ij}{}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) - \mathbf{H}^{f}_{ij}{}^{(k)}(\mathbf{f}_j - \hat{\mathbf{f}}_j^{(k)}) && (24)\\
&= \delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}^{r}_{ij}{}^{(k)}\tilde{\mathbf{r}}_i^{(k)} - \mathbf{H}^{f}_{ij}{}^{(k)}\tilde{\mathbf{f}}_j^{(k)} && (25)\\
&= \delta\mathbf{z}_{ij}{}^{(k)} - \underbrace{\begin{bmatrix}\mathbf{0} & \mathbf{0}\cdots\mathbf{0} & \mathbf{H}^{r}_{ij}{}^{(k)} & \mathbf{0}\cdots\mathbf{0} & | & \mathbf{0}\cdots\mathbf{0} & \mathbf{H}^{r}_{ij}{}^{(k)} & \mathbf{0}\cdots\mathbf{0}\end{bmatrix}}_{\mathbf{H}_{ij}^{(k)}} \underbrace{\begin{bmatrix}\tilde{\mathbf{r}}_0^{(k)}\\ \tilde{\mathbf{r}}_1^{(k)}\\ \vdots\\ \tilde{\mathbf{r}}_{i-1}^{(k)}\\ \tilde{\mathbf{r}}_i^{(k)}\\ \tilde{\mathbf{r}}_{i+1}^{(k)}\\ \vdots\\ \tilde{\mathbf{r}}_N^{(k)}\\ --\\ \tilde{\mathbf{f}}_1^{(k)}\\ \vdots\\ \tilde{\mathbf{f}}_{j-1}^{(k)}\\ \tilde{\mathbf{f}}_j^{(k)}\\ \tilde{\mathbf{f}}_{j+1}^{(k)}\\ \vdots\\ \tilde{\mathbf{f}}_M^{(k)}\end{bmatrix}}_{\tilde{\mathbf{x}}^{(k)}} && (26)\\
&= \delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)} && (27)
\end{aligned}$$

Substitute (27) to (22) we get:

$$p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) = \frac{1}{(2\pi)^{m/2}det(\mathbf{R}_{ij})^{1/2}} e^{-\frac{1}{2}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right]^T \mathbf{R}_{ij}^{-1}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right]} \tag{28}$$

Taking log of both sides we get:

$$\log p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) = \log(\frac{1}{(2\pi)^{m/2}det(\mathbf{R}_{ij})^{1/2}}) - \frac{1}{2}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right]^T \mathbf{R}_{ij}^{-1}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right] \tag{29}$$

$$\Leftrightarrow -\log p(\mathbf{z}_{ij}|\mathbf{r}_i,\mathbf{f}_j) = -\log(\frac{1}{(2\pi)^{m/2}det(\mathbf{R}_{ij})^{1/2}}) + \frac{1}{2}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right]^T \mathbf{R}_{ij}^{-1}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right] \tag{30}$$

After getting rid of the constant $-\log(\frac{1}{(2\pi)^{m/2}det(\mathbf{R}_{ij})^{1/2}})$ again, only the term $+\frac{1}{2}\left[\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right]^T \mathbf{R}_{ij}^{-1}\left[\delta\mathbf{z}_{\mathbf{ij}}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right]$ or $\frac{1}{2}\left\|\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right\|_{\mathbf{R}_{ij}}^2$ will contribute to the cost function (14).

Here, we can consider a less rigorous way to obtain the contributor factor of $p(\mathbf{z}_{ij},\mathbf{r}_{ij},\mathbf{f}_j)$ to the cost function $\mathbb{C}(\mathbf{x})$.

From (17) with $\hat{\mathbf{r}}_i^* = \hat{\mathbf{r}}_i^{(k)}$ and $\hat{\mathbf{f}}_j^* = \hat{\mathbf{f}}_j^{(k)}$:

$$\mathbf{z}_{ij} \simeq \mathbf{h}(\hat{\mathbf{r}}_i^{(k)},\hat{\mathbf{f}}_j^{(k)}) + \mathbf{H}^r_{ij}{}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) + \mathbf{H}^f_{ij}{}^{(k)} + \mathbf{n}_{ij} \tag{31}$$

$$\Leftrightarrow \quad \mathbf{n}_{ij} = \delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)} \tag{32}$$

$$\mathbf{n}_{ij} \sim \mathcal{N}(\mathbf{0},\mathbf{R}_{ij}) \tag{33}$$

$$\Rightarrow \quad \delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)} \sim \mathcal{N}(\mathbf{0},\mathbf{R}_{ij}) \tag{34}$$

$$\Rightarrow \text{contribution factor: } \left\|\delta\mathbf{z}_{ij}{}^{(k)} - \mathbf{H}_{ij}^{(k)}\tilde{\mathbf{x}}^{(k)}\right\|_{\mathbf{R}_{ij}}^2$$

B. *Proprioceptive:*

$$\mathbf{r}_{i+1} = \mathbf{f}(\mathbf{r}_i,\mathbf{u}_i,\mathbf{w}_i), \quad \mathbf{w}_i \sim \mathcal{N}(\mathbf{0},\mathbf{Q}_i), \quad \mathbf{r}_{i+1} \in \mathbb{R}^n, \quad i = 0,1,...,N-1 \tag{35}$$

Linearize around arbitrary point $\hat{\mathbf{r}}_i^*$ and $\hat{\mathbf{w}}_i^*$ for the noise we can get:

$$\mathbf{r}_{i+1} \simeq \mathbf{f}(\hat{\mathbf{r}}_i^*,\mathbf{u}_i,\hat{\mathbf{w}}_i^*) + \underbrace{\frac{\partial\mathbf{f}}{\partial\mathbf{r}_i}\bigg|_{\mathbf{r}_i=\hat{\mathbf{r}}_i^*}}_{\mathbf{\Phi}_{i+1,i}}(\mathbf{r}_i - \hat{\mathbf{r}}_i^*) + \underbrace{\frac{\partial\mathbf{f}}{\partial\mathbf{w}_i}\bigg|_{\mathbf{w}_i=\hat{\mathbf{w}}_i^*}}_{\mathbf{G}_i}(\mathbf{w}_i - \hat{\mathbf{w}}_i^*) \tag{36}$$

Given the estimate of robot pose and noise at time step i from k^{th} Gauss-Newton iteration ($\hat{\mathbf{r}}_i^{(k)}$, $\hat{\mathbf{w}}_i^{(k)}$) and choose to linearize $\mathbf{f}(\mathbf{r}_i,\mathbf{u}_i,\mathbf{w}_i)$ at $(\hat{\mathbf{r}}_i^{(k)},\mathbf{u}_i,\hat{\mathbf{w}}_i^{(k)})$, we have the following:

$$\mathbf{r}_{i+1} \simeq \mathbf{f}(\hat{\mathbf{r}}_i^{(k)},\mathbf{u}_i,\hat{\mathbf{w}}_i^{(k)}) + \underbrace{\frac{\partial\mathbf{f}}{\partial\mathbf{r}_i}\bigg|_{\mathbf{r}_i=\hat{\mathbf{r}}_i^{(k)}}}_{\mathbf{\Phi}_{i+1,i}^{(k)}}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) + \underbrace{\frac{\partial\mathbf{f}}{\partial\mathbf{w}_i}\bigg|_{\mathbf{w}_i=\hat{\mathbf{w}}_i^{(k)}}}_{\mathbf{G}_i^{(k)}}(\mathbf{w}_i - \hat{\mathbf{w}}_i^{(k)}) \tag{37}$$

$$\Leftrightarrow \quad \mathbf{r}_{i+1} \simeq \underbrace{\mathbf{f}(\hat{\mathbf{r}}_i^{(k)},\mathbf{u}_i,\hat{\mathbf{w}}_i^{(k)})}_{\hat{\mathbf{r}}_{i+1}^{(k)}} + \mathbf{\Phi}_{i+1,i}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) + \mathbf{G}_i^{(k)}(\mathbf{w}_i - \hat{\mathbf{w}}_i^{(k)}) \tag{38}$$

From the linearized equation at k^{th} iteration, we can get the first order statistic of $\mathbf{r}_i^{(k)}$:

$$\mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i] = \mathbf{f}(\hat{\mathbf{r}}_i^{(k)}, \mathbf{u}_i, \hat{\mathbf{w}}_i^{(k)}) + \boldsymbol{\Phi}_{i+1,i}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) - \mathbf{G}_i^{(k)}\hat{\mathbf{w}}_i^{(k)} \tag{39}$$

$$\Rightarrow \quad \mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i] = \mathbf{r}_{i+1} - \mathbf{f}(\hat{\mathbf{r}}_i^{(k)}, \mathbf{u}_i, \hat{\mathbf{w}}_i^{(k)}) - \boldsymbol{\Phi}_{i+1,i}^{(k)}(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)}) + \mathbf{G}_i^{(k)}\hat{\mathbf{w}}_i^{(k)} \tag{40}$$

$$\Leftrightarrow \quad \mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i] = \underbrace{\mathbf{r}_{i+1} - \hat{\mathbf{r}}_{i+1}^{(k)}}_{\tilde{\mathbf{r}}_{i+1}} - \boldsymbol{\Phi}_{i+1,i}^{(k)}\underbrace{(\mathbf{r}_i - \hat{\mathbf{r}}_i^{(k)})}_{\tilde{\mathbf{r}}_i} + \underbrace{\mathbf{G}_i^{(k)}\hat{\mathbf{w}}_i^{(k)}}_{\delta\mathbf{z}_i^{(k)}} \tag{41}$$

$$\Leftrightarrow \quad \mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i] = \tilde{\mathbf{r}}_{i+1} - \boldsymbol{\Phi}_{i+1,i}^{(k)}\tilde{\mathbf{r}}_i + \delta\mathbf{z}_i^{(k)} \tag{42}$$

$$\Leftrightarrow \quad \mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i] = \delta\mathbf{z}_i^{(k)} + \underbrace{\begin{bmatrix}\mathbf{0} & \mathbf{0}\cdots\mathbf{0} & -\boldsymbol{\Phi}_{i+1,i}^{(k)} & \mathbf{I} & \mathbf{0}\cdots\mathbf{0} & | & \mathbf{0}\cdots\mathbf{0}\end{bmatrix}}_{\bar{\mathbf{H}}_{i+1,i}^{(k)}} \underbrace{\begin{bmatrix}\tilde{\mathbf{r}}_0^{(k)} \\ \tilde{\mathbf{r}}_1^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_{i-1}^{(k)} \\ \tilde{\mathbf{r}}_i^{(k)} \\ \tilde{\mathbf{r}}_{i+1}^{(k)} \\ \tilde{\mathbf{r}}_{i+2}^{(k)} \\ \vdots \\ \tilde{\mathbf{r}}_N^{(k)} \\ -- \\ \tilde{\mathbf{f}}_1^{(k)} \\ \vdots \\ \tilde{\mathbf{f}}_M^{(k)}\end{bmatrix}}_{\tilde{\mathbf{x}}^{(k)}} \tag{43}$$

$$\Leftrightarrow \quad \mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i] = \delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_i^{(k)}\tilde{\mathbf{x}}^{(k)} \tag{44}$$

This is very important to notice that $\hat{\mathbf{r}}_{i+1}^{(k)} = \mathbb{E}[\mathbf{r}_{i+1}|\hat{\mathbf{r}}_i^{(k)}, \hat{\mathbf{w}}_i^{(k)}] \neq \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i]$.

Then second statistics of $\mathbf{r}_i^{(k)}$:

$$\mathbb{E}[(\mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i])(\mathbf{r}_{i+1} - \mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i])^T|\mathbf{r}_i] \simeq \mathbf{G}_i^{(k)}\mathbb{E}[\mathbf{w}_i\mathbf{w}_i^T]\mathbf{G}_i^{(k)^T} = \mathbf{G}_i^{(k)}\mathbf{Q}_i\mathbf{G}_i^{(k)^T} \tag{45}$$

Notice that we have updated our information on the noise in this particular process of pose $(i+1)^{th}$; however, the statistics of the noise does not change and remain as (35).

Denote $\mathbf{Q'}_i^{(k)} = \mathbf{G}_i^{(k)}\mathbf{Q}_i\mathbf{G}_i^{(k)^T}$ we have:

$$p(\mathbf{r}_{i+1}|\mathbf{r}_i) = \frac{1}{(2\pi)^{n/2}det(\mathbf{Q'}_i^{(k)})^{1/2}} e^{-\frac{1}{2}(\mathbf{r}_{i+1}-\mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i])^T(\mathbf{Q'}_i^{(k)})^{-1}(\mathbf{r}_{i+1}-\mathbb{E}[\mathbf{r}_{i+1}|\mathbf{r}_i])} \tag{46}$$

$$= \frac{1}{(2\pi)^{n/2}det(\mathbf{Q'}_i^{(k)})^{1/2}} e^{-\frac{1}{2}\left(\delta\mathbf{z}_i^{(k)}+\bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right)^T(\mathbf{Q'}_i^{(k)})^{-1}\left(\delta\mathbf{z}_i^{(k)}+\bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right)} \tag{47}$$

Taking the logarithm of both sides of (47):

$$\log p(\mathbf{r}_{i+1}|\mathbf{r}_i) = \log[\frac{1}{(2\pi)^{n/2}det(\mathbf{Q'}_i^{(k)})^{1/2}}] - \frac{1}{2}\left(\delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_{i+1,1}^{(k)}\tilde{\mathbf{x}}^{(k)}\right)^T (\mathbf{Q'}_i^{(k)})^{-1}\left(\delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right) \tag{48}$$

$$\Leftrightarrow -\log p(\mathbf{r}_{i+1}|\mathbf{r}_i) = -\log[\frac{1}{(2\pi)^{n/2}det(\mathbf{Q'}_i^{(k)})^{1/2}}] + \frac{1}{2}\left(\delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right)^T (\mathbf{Q'}_i^{(k)})^{-1}\left(\delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right) \tag{49}$$

Neglecting the constant term $-\log[\frac{1}{(2\pi)^{n/2}det(\mathbf{Q'}_i^{(k)})^{1/2}}]$, we left with the term $\frac{1}{2}\left(\delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right)^T (\mathbf{Q'}_i^{(k)})^{-1}\left(\delta\mathbf{z}_i^{(k)} + \bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right)$ or $\frac{1}{2}\left\|\delta\mathbf{z}_i + \bar{\mathbf{H}}_{i+1,i}^{(k)}\tilde{\mathbf{x}}^{(k)}\right\|^2_{\mathbf{Q'}_i}$ that will contribute to the cost function (14).

C. Prior:

We may be given the prior estimates for both the "robot" and the features. For simplicity, now we will assume that we are only given priors for few of the states and the errors in them are additive (for the more complicated case like quaternion is dealt later)

$$\mathbf{z}^o = \begin{bmatrix} \vdots \\ \widehat{\mathbf{r}}_i \\ \vdots \\ \widehat{\mathbf{f}}_j \\ \vdots \end{bmatrix} + \mathbf{n}_s, \quad \mathbf{n}_s \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \ddots & & & & \\ & \mathbf{P}_{r_i r_j} & & & \\ & & \ddots & & \\ & & & \mathbf{P}_{f_j f_j} & \\ & & & & \ddots \end{bmatrix} \right) \tag{50}$$

Using the sample approach (as we did in Exteroceptive) to find the quadratic contribution to cost function

$$\mathbf{r}_i = \widehat{\mathbf{r}}_i + \tilde{\mathbf{r}}_i \Rightarrow \tilde{\mathbf{r}}_i = \mathbf{I} \cdot \mathbf{r}_i - \widehat{\mathbf{r}}_i, \quad \tilde{\mathbf{r}}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{\mathbf{r}_i \mathbf{r}_i}) \tag{51}$$

$$\mathbf{f}_j = \widehat{\mathbf{f}}_j + \tilde{\mathbf{f}}_j \Rightarrow \tilde{\mathbf{f}}_j = \mathbf{I} \cdot \mathbf{f}_j - \widehat{\mathbf{f}}_j, \quad \tilde{\mathbf{f}}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_{\mathbf{f}_j \mathbf{f}_j}) \tag{52}$$

for $\mathbf{i} \in \mathbf{S}_i$, $\mathbf{j} \in \mathbf{S}_j$.

From (51)

$$\mathbf{0} - \underbrace{\begin{bmatrix} \mathbf{0} \cdots \mathbf{I} \cdots \mathbf{0} & | & \mathbf{0} \cdots \mathbf{0} \end{bmatrix}}_{\mathbf{H}_i^{or}} \begin{bmatrix} \tilde{\mathbf{r}}_0 \\ \vdots \\ \tilde{\mathbf{r}}_i \\ \vdots \\ \tilde{\mathbf{r}}_N \\ -- \\ \tilde{\mathbf{f}}_1 \\ \vdots \\ \tilde{\mathbf{f}}_M \end{bmatrix} = -\tilde{\mathbf{r}}_i \tag{53}$$

$$\Rightarrow \quad \mathbf{0} - \mathbf{H}_i^{or} \tilde{\mathbf{x}} = -\tilde{\mathbf{r}}_i \tag{54}$$

contribution to quadratic cost function $\|\mathbf{0} - \mathbf{H}_i^{or} \tilde{\mathbf{x}}\|^2_{\mathbf{P}_{\mathbf{r}_i \mathbf{r}_i}}$.

From (52)

$$\mathbf{0} - \underbrace{\begin{bmatrix} \mathbf{0} \cdots \mathbf{0} & | & \mathbf{0} \cdots \mathbf{I} \cdots \mathbf{0} \end{bmatrix}}_{\mathbf{H}_j^{of}} \begin{bmatrix} \tilde{\mathbf{r}}_0 \\ \vdots \\ \tilde{\mathbf{r}}_N \\ -- \\ \tilde{\mathbf{f}}_1 \\ \vdots \\ \tilde{\mathbf{f}}_j \\ \vdots \\ \tilde{\mathbf{f}}_M \end{bmatrix} = -\tilde{\mathbf{f}}_j \tag{55}$$

$$\Rightarrow \quad \mathbf{0} - \mathbf{H}_j^{of} \tilde{\mathbf{x}} = -\tilde{\mathbf{f}}_i \tag{56}$$

contribution to quadratic cost function $\left\|\mathbf{0} - \mathbf{H}_j^{of} \tilde{\mathbf{x}}\right\|^2_{\mathbf{P}_{\mathbf{f}_j \mathbf{f}_j}}$.

Note: In most cases in practice we only have prior for only part of the initial robot pose which often are deterministically defined. (e.g. set position and yaw which are unobservable deterministically to zero and give initial estimate to the others from IMU etc.)

D. Summary of cost function:

$$\mathbb{C}(\mathbf{x}) = \mathbb{C}(\tilde{\mathbf{x}}) = \frac{1}{2}\sum_i \|\mathbf{0} - \mathbf{H}_i^{or}\tilde{\mathbf{x}}\|^2_{\mathbf{P}_{r_i r_i}} + \frac{1}{2}\sum_i \left\|\mathbf{0} - \mathbf{H}_j^{of}\tilde{\mathbf{x}}\right\|^2_{\mathbf{Q}'_i} + \frac{1}{2}\sum_{ij} \left\|\boldsymbol{\delta}\mathbf{z}_{ij} - \mathbf{H}_{ij}\tilde{\mathbf{x}}\right\|^2_{\mathbf{R}_{ij}} + \frac{1}{2}\sum_{i=0}^{N-1} \left\|\boldsymbol{\delta}\mathbf{z}_i + \bar{\mathbf{H}}_{i+1,i}\tilde{\mathbf{x}}\right\|^2_{\mathbf{Q}'_i} \quad (57)$$

• Here for simplicity, we drop the superscript (k) the current iteration of Gauss-Newton.

• We can easily recognize that every quadratic term in the cost function $\mathbb{C}(\mathbf{x})$ has the form: $\|\mathbf{y}_k - \mathbf{C}_k\mathbf{x}\|^2_{\mathbf{R}_k}$, where k spans all possible values in (57) and $\mathbf{y}_k$, $\mathbf{C}_k$, $\mathbf{R}_k$ are adjusted accordingly.

$$(57) \Rightarrow \mathbb{C}(\mathbf{x}) = \frac{1}{2}\sum_{k=1}^{K} \|\mathbf{y}_k - \mathbf{C}_k\mathbf{x}\|^2_{\mathbf{R}_k} \quad (58)$$

$$(58) \Leftrightarrow \mathbb{C}(\mathbf{x}) = \frac{1}{2}\sum_{k=1}^{K} (\mathbf{y}_k - \mathbf{C}_k\mathbf{x})^T \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{C}_k\mathbf{x}) \quad (59)$$

$$= \frac{1}{2}\sum_{k=1}^{K} [\mathbf{R}_k^{-\frac{1}{2}}(\mathbf{y}_k - \mathbf{C}_k\mathbf{x})]^T [\mathbf{R}_k^{-\frac{1}{2}}(\mathbf{y}_k - \mathbf{C}_k\mathbf{x})] \quad (60)$$

$$= \frac{1}{2}\sum_{k=1}^{K} \mathbf{u}_k^T\mathbf{u}_k \quad (61)$$

$$= \frac{1}{2}\begin{bmatrix}\mathbf{u}_1^T & \mathbf{u}_2^T & \cdots & \mathbf{u}_K^T\end{bmatrix}\begin{bmatrix}\mathbf{u}_1 \\ \mathbf{u}_2 \\ \vdots \\ \mathbf{u}_K\end{bmatrix} \quad (62)$$

$$(63)$$

where

$$\begin{bmatrix}\mathbf{u}_1 \\ \vdots \\ \mathbf{u}_K\end{bmatrix} = \begin{bmatrix}\mathbf{R}_1^{-\frac{1}{2}}(\mathbf{y}_1 - \mathbf{C}_1\mathbf{x}) \\ \vdots \\ \mathbf{R}_K^{-\frac{1}{2}}(\mathbf{y}_K - \mathbf{C}_K\mathbf{x})\end{bmatrix} \quad (64)$$

$$= \begin{bmatrix}\mathbf{R}_1^{-\frac{1}{2}} & & & \\ & \mathbf{R}_2^{-\frac{1}{2}} & & \\ & & \ddots & \\ & & & \mathbf{R}_K^{-\frac{1}{2}}\end{bmatrix}\begin{bmatrix}(\mathbf{y}_1 - \mathbf{C}_1\mathbf{x}) \\ (\mathbf{y}_2 - \mathbf{C}_2\mathbf{x}) \\ \vdots \\ (\mathbf{y}_K - \mathbf{C}_K\mathbf{x})\end{bmatrix} \quad (65)$$

$$= \underbrace{\begin{bmatrix}\mathbf{R}_1^{-\frac{1}{2}} & & & \\ & \mathbf{R}_2^{-\frac{1}{2}} & & \\ & & \ddots & \\ & & & \mathbf{R}_K^{-\frac{1}{2}}\end{bmatrix}}_{\mathbf{R}^{-\frac{1}{2}}}\left[\underbrace{\begin{bmatrix}\mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_K\end{bmatrix}}_{\mathbf{y}} - \underbrace{\begin{bmatrix}\mathbf{C}_1 \\ \mathbf{C}_2 \\ \vdots \\ \mathbf{C}_K\end{bmatrix}}_{\mathbf{C}}\mathbf{x}\right] \quad (66)$$

$$= \mathbf{R}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{C}\mathbf{x}) \quad (67)$$

Then the total cost function can be written as:

$$\mathbb{C}(\mathbf{x}) = \frac{1}{2}\left[\mathbf{R}^{-\frac{1}{2}}(\mathbf{y}-\mathbf{Cx})\right]^T \left[\mathbf{R}^{-\frac{1}{2}}(\mathbf{y}-\mathbf{Cx})\right] \tag{68}$$

$$= \frac{1}{2}(\mathbf{y}-\mathbf{Cx})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Cx}) \tag{69}$$

$$= \frac{1}{2}\|\mathbf{y}-\mathbf{Cx}\|^2_{\mathbf{R}} \tag{70}$$

To minimize $\mathbb{C}(\mathbf{x})$, we take the derivative with respect to $\mathbf{x}$ and set it to zero:

$$\frac{\partial \mathbb{C}(\mathbf{x})}{\partial \mathbf{x}} = \frac{1}{2}\frac{\partial}{\partial \mathbf{x}}(\mathbf{y}^T\mathbf{R}^{-1}\mathbf{y} - 2\mathbf{y}^T\mathbf{R}^{-1}\mathbf{Cx} + \mathbf{x}^T\mathbf{C}^T\mathbf{R}^{-1}\mathbf{Cx}) \tag{71}$$

$$= \frac{1}{2}(-2\mathbf{y}^T\mathbf{R}^{-1}\mathbf{Cx} + 2\mathbf{x}^T\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}) \tag{72}$$

$$= -\mathbf{y}^T\mathbf{R}^{-1}\mathbf{Cx} + \mathbf{x}^T\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C} \tag{73}$$

$$\frac{\partial \mathbb{C}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0} \tag{74}$$

$$\Leftrightarrow \underbrace{(\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C})}_{Hessian}\mathbf{x} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{y} \tag{75}$$

(75) is the so-called normal equation.

E. The Structure of Hessian:

The structure of the Hessian in (75) is:

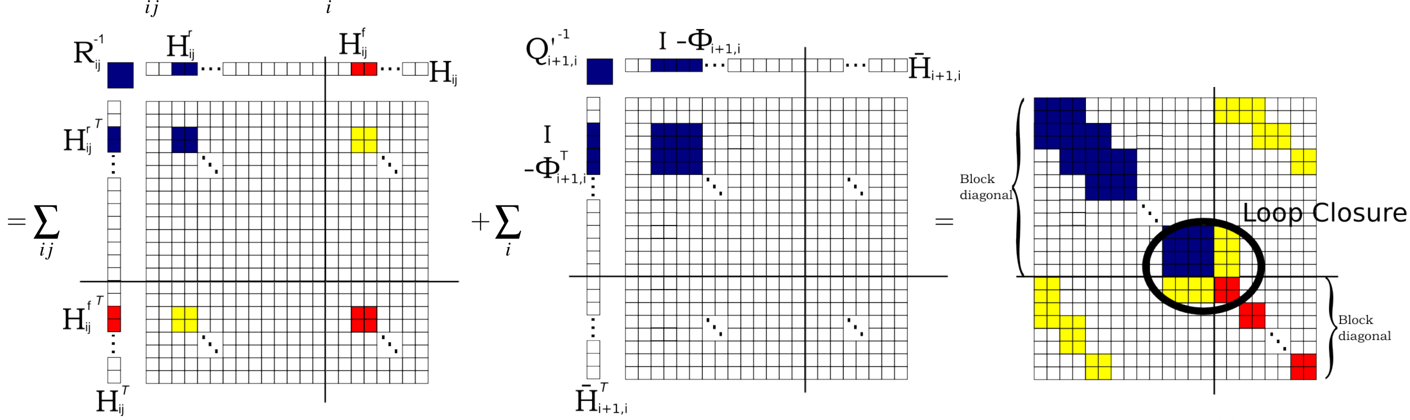
$$\mathcal{H} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{C} \tag{76}$$

$$= \begin{bmatrix}\mathbf{C}_1^T & \cdots & \mathbf{C}_K^T\end{bmatrix}\begin{bmatrix}\mathbf{R}_1^{-1} & & \\ & \ddots & \\ & & \mathbf{R}_K^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{C}_1 \\ \vdots \\ \mathbf{C}_K\end{bmatrix} \tag{77}$$

$$= \sum_{k=1}^{K}\mathbf{C}_k^T\mathbf{R}_k^{-1}\mathbf{C}_k \tag{78}$$

which depending on the contraints to which each of the $\mathbf{C}_k$ corresponds to. (from here on we ignore the minor terms and concentrate on the proprioceptive and exteroceptive measurement.) Note that the prior can be thought as an exteroceptive measurement providing information for individual states of the robot of the feature.

$$(78) \Rightarrow \mathcal{H} = \sum_{ij}\mathbf{H}_{ij}^T\mathbf{R}_{ij}^{-1}\mathbf{H}_{ij} + \sum_i \bar{\mathbf{H}}_{i+1,i}^T\mathbf{Q}'^{-1}_i\bar{\mathbf{H}}_{i+1,i} \tag{79}$$



(80)

F. General Case of Error model:

$$\boldsymbol{\delta\xi} = \boldsymbol{\Psi}(\boldsymbol{\xi}, \hat{\boldsymbol{\xi}}) \simeq \boldsymbol{\Psi}(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\xi}}) + \frac{\partial \boldsymbol{\Psi}}{\partial \boldsymbol{\xi}} \tilde{\boldsymbol{\xi}} \tag{81}$$

Example: In the first iteration, we have: $\mathbf{q} = \boldsymbol{\delta q} \otimes \hat{\mathbf{q}} \Rightarrow$ constant contributors:

$$\boldsymbol{\delta q} = \mathbf{q} \otimes \hat{\mathbf{q}}^{-1} = \mathbf{R}(\hat{\mathbf{q}}^{-1})\mathbf{q} = \mathbf{R}(\hat{\mathbf{q}}^{-1})\hat{\mathbf{q}} + \mathbf{R}(\hat{\mathbf{q}}^{-1})(\mathbf{q} - \hat{\mathbf{q}}) \tag{82}$$

$$\boldsymbol{\delta q} = \mathbf{q} = \mathbf{q} \otimes \hat{\mathbf{q}}^{-1} \simeq \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta} \\ 1 \end{bmatrix} \Rightarrow \mathbf{q} = \boldsymbol{\delta q} \otimes \mathbf{q} \tag{83}$$

$$^{\mathbf{I}}_{\mathbf{G}}\mathbf{q} = {}^{\mathbf{I}}_{\hat{\mathbf{I}}_1}\boldsymbol{\delta}\mathbf{q} \otimes {}^{\hat{\mathbf{I}}_1}_{\mathbf{G}}\mathbf{q} = \boldsymbol{\delta}\mathbf{q}_1 \otimes \hat{\mathbf{q}}_1 \tag{84}$$

$$\boldsymbol{\delta}\bar{\mathbf{q}}_1 = \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}_1 \\ 1 \end{bmatrix} \;;\; \boldsymbol{\delta\theta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_1) \tag{85}$$

for small angle approximation

$$\boldsymbol{\delta\theta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}) \tag{86}$$

$$^{\mathbf{I}}_{\mathbf{G}}\mathbf{q} = {}^{\mathbf{I}}_{\hat{\mathbf{I}}_2}\boldsymbol{\delta}\mathbf{q} \otimes {}^{\hat{\mathbf{I}}_2}_{\mathbf{G}}\mathbf{q} = \boldsymbol{\delta}\mathbf{q}_2 \otimes \hat{\mathbf{q}}_2 \tag{87}$$

$$\boldsymbol{\delta}\bar{\mathbf{q}}_2 = \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}_2 \\ 1 \end{bmatrix} \tag{88}$$

for small angle approximation we want to find the distribution of $\boldsymbol{\delta\theta}_2$

$$^{\mathbf{I}}_{\mathbf{G}} = \boldsymbol{\delta}\mathbf{q}_1 \otimes \hat{\mathbf{q}}_1 = \boldsymbol{\delta}\mathbf{q}_2 \otimes \hat{\mathbf{q}}_2 \tag{89}$$

$$\Leftrightarrow \boldsymbol{\delta}\mathbf{q}_1 = \boldsymbol{\delta\delta}_2 \otimes \underbrace{\hat{\mathbf{q}}_2 \otimes \hat{\mathbf{q}}_1^{-1}}_{\begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}_{12} \\ 1 \end{bmatrix}} \tag{90}$$

$$\Leftrightarrow \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}_2 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} \frac{1}{2}\boldsymbol{\delta\theta}_{12} \\ 1 \end{bmatrix} \tag{91}$$

$$\Leftrightarrow \boldsymbol{\delta\theta}_1 = \boldsymbol{\delta\theta}_2 + \boldsymbol{\delta\theta}_{12} \quad \boldsymbol{\delta\theta}_2 + \boldsymbol{\delta\theta}_{12} = \boldsymbol{\delta\theta}_1 \sim \mathcal{N}(\mathbf{0}, \mathbf{P}_1) \tag{92}$$

$$\Rightarrow combination\ factor\ to\ cost\ function:$$

$$\|\boldsymbol{\delta\theta}_2 + \underbrace{\boldsymbol{\delta\theta}_{12}}_{Deterministic\ quantity} \|^2_{P_1} \tag{93}$$

II. State Marginalization using the Hessian

A. Solution in batch form:

$$\mathbb{C}(\mathbf{x}) = \|\mathbf{y} - \mathbf{C}\mathbf{x}\|^2_{\mathbf{R}} = (\mathbf{y} - \mathbf{C}\mathbf{x})^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{C}\mathbf{x}) \tag{94}$$

with $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}$; optimal solution:

$$\frac{\partial \mathbf{C}}{\partial \mathbf{x}} = 0 \tag{95}$$

$$\Rightarrow \left(\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}\right) \mathbf{x} = \mathbf{C}^T \mathbf{R}^{-1} \mathbf{y} \tag{96}$$

$$\Leftrightarrow \mathbf{x} = \left(\mathbf{C}^T \mathbf{R}^{-1} \mathbf{C}\right)^{-1} \mathbf{C}^T \mathbf{R}^{-1} \mathbf{y} \tag{97}$$

B. Schur complement:

In this subsection we try to solve the normal equation to get the optimal estimate for the state vector $\mathbf{x}$, by solving partially parts of the vector state instead of solving the in batch form as in (75)

1) First approach:

$$\arg\min_{\mathbf{x}} \mathbb{C}(\mathbf{x}) \Leftrightarrow \arg\min_{\mathbf{x}_2}\left(\arg\min_{\mathbf{x}_1}\left(\mathbf{C}\left(\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\right)\right)\right) \tag{98}$$

In this approach we seek for a solution of $\mathbf{x}_1$ based on $\mathbf{x}_2$ and then substitute to the original normal equation (75) to solve for $\mathbf{x}_2$, then substitute back to get $\mathbf{x}_1$. The procedure is described as followed:

$$\mathbb{C}(\mathbf{x}) = \|\mathbf{y} - \begin{bmatrix}\mathbf{C}_1 & \mathbf{C}_2\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}\|^2_{\mathbf{R}} = \|\underbrace{\mathbf{y}-\mathbf{C}_2\mathbf{x}_2}_{\mathbf{y}'} - \mathbf{C}_1\mathbf{x}_1\|^2_{\mathbf{R}} = \|\mathbf{y}' - \mathbf{C}_1\mathbf{x}_1\|^2_{\mathbf{R}} \tag{99}$$

$$\arg\min \mathbb{C}(\mathbf{x}_1) \Leftrightarrow \frac{\partial \mathbb{C}(\mathbf{x}_1)}{\partial \mathbf{x}_1} = 0 \tag{100}$$

$$\Rightarrow \left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)\mathbf{x}_1 = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{y}' \tag{101}$$

$$\Leftrightarrow \mathbf{x}_1 = \left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{y}' \tag{102}$$

Substitute (102) into (99) we have:

$$\begin{aligned}\mathbb{C}(\mathbf{x}_2) &= \|\mathbf{y} - \mathbf{C}_2\mathbf{x}_2 - \mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{C}_2\mathbf{x}_2)\|^2_{\mathbf{R}} &(103)\\ &= \|\underbrace{\left(\mathbf{I} - \mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)}_{\mathbf{D}}(\mathbf{y}-\mathbf{C}_2\mathbf{x}_2)\|^2_{\mathbf{R}} &(104)\\ &= (\mathbf{y}-\mathbf{C}_2\mathbf{x}_2)^T\underbrace{\mathbf{D}^T\mathbf{R}^{-1}\mathbf{D}}_{\mathbf{\Theta}^{-1}}(\mathbf{y}-\mathbf{C}_2\mathbf{x}_2) = \|\mathbf{y}-\mathbf{C}_2\mathbf{x}_2\|^2_{\mathbf{\Theta}} &(105)\end{aligned}$$

where

$$\mathbf{\Theta}^{-1} = \left(\mathbf{I} - \mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)^T\mathbf{R}^{-1}\left(\mathbf{I} - \mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right) \tag{106}$$

$$\begin{aligned}\Leftrightarrow \mathbf{\Theta}^{-1} = \mathbf{R}^{-1} &- \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\\ &+ \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\underbrace{\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}}_{\mathbf{I}}\mathbf{C}_1^T\mathbf{R}^{-1} \qquad (107)\end{aligned}$$

$$\Leftrightarrow \mathbf{\Theta}^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1} \tag{108}$$

substituting into (102) we have:

$$\mathbb{C}(\mathbf{x}) = \|\mathbf{y}-\mathbf{C}_2\mathbf{x}_2\|^2_{\mathbf{\Theta}} \tag{109}$$

$$\Rightarrow \mathbf{C}_2^T\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\mathbf{C}_2\mathbf{x}_2 = \mathbf{C}_2^T\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\mathbf{y} \tag{110}$$

(102) can be written as:

$$\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)\mathbf{x}_1 = \mathbf{C}_1^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{C}_2\mathbf{x}_2) \tag{111}$$

$$\Leftrightarrow \left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)\mathbf{x}_1 + \left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2\right)\mathbf{x}_2 = \mathbf{C}_1\mathbf{R}^{-1}\mathbf{y} \tag{112}$$

From (112) and (110):

$$\begin{bmatrix}\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2\\ \mathbf{0} & \mathbf{C}_2^T\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1^{-1}\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} = \begin{bmatrix}\mathbf{C}_1^T\\ \mathbf{C}_2^T\left(\mathbf{I} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\right)\end{bmatrix}\mathbf{R}^{-1}\mathbf{y} \tag{113}$$

2) Second approach: A different way to get (113).(i.e. to split the problems into 2 parts where to first solve for x_2 and then back substitute to find x_1) is the following (based on block Gaussian elimination):

From (75) we have:

$$\left(\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}\right)\mathbf{x} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{y} \qquad (\textit{where} \quad \mathbf{C} = [\mathbf{C}_1 \;\; \mathbf{C}_2]) \tag{114}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2 \\ \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1 & \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_2^T \end{bmatrix} \mathbf{R}^{-1}\mathbf{y} \tag{115}$$

Multiply both sides the lower triangular block matrix we have:

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\left(\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1\right)\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2 \\ \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1 & \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\left(\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1\right)\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_2^T \end{bmatrix} \mathbf{R}^{-1}\mathbf{y} \tag{116}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2 \\ \mathbf{0} & \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_2 - \left(\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1\right)\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2\right) \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T \\ \mathbf{C}_2^T\left(\mathbf{I} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\right) \end{bmatrix} \mathbf{R}^{-1}\mathbf{y} \tag{117}$$

We can obviously see that the last equation is the same as (113)

What we accomplished before is to ”marginalize” $\mathbf{x}_1$ and only have a square system of equations for $\mathbf{x}_2$ which is of lower dimension and which we can solve first and then use its solution in (117) to efficiently solve for $\mathbf{x}_1$. That is if $\mathbf{x}_2$ is the solution of (117) (or (113) equivalently). Then we substitute to (102) to get $\mathbf{x}_1$.

In this process of reducing the problem to a smaller one, we employed the Hessian and applied block Gaussian elimination.

In the next section we will show how we can accomplish the same without forming the normal equation but instead by directly operate on the Jacobian.

III. State Marginalization using the Jacobian

In this section, we give an alternate solution for solving an optimal $\mathbf{x}$ to minimize the cost function (14):

$$\arg\min \|\mathbf{y} - \mathbf{C}\mathbf{x}\|_{\mathbf{R}}^2 \Leftrightarrow \arg\max p(\mathbf{y}|\mathbf{x}) \tag{118}$$

where

$$\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{R}), \quad \mathbf{y} \in \mathbb{R}^M, \mathbf{x} \in \mathbb{R}^N, \mathbf{C}_{M\times N}, M > N \tag{119}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}, \quad \mathbf{x}_1 \in \mathbb{R}^{N_1}, \mathbf{x}_2 \in \mathbb{R}^{N_2}, N = N_1 + N_2 \tag{120}$$

$$(119) \Rightarrow \mathbf{y} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{C}_2 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} + \mathbf{n} \tag{121}$$

$$\Rightarrow \underset{(M\times 1)}{\mathbf{y}} = \underset{(M\times N_1)}{\mathbf{C}_1}\underset{(N_1\times 1)}{\mathbf{x}_1} + \underset{(M\times N_2)}{\mathbf{C}_2}\underset{(N_2\times 1)}{\mathbf{x}_2} + \underset{(M\times 1)}{\mathbf{n}} \tag{122}$$

To remove the dependence of $\mathbf{y}$ on $\mathbf{x}_1$ we can multiply (122) with $\mathbf{U}^T$ where $\mathbf{U}$ is of dimension $M \times (M - N_2)$ ideally with $\mathbf{U}^T\mathbf{C}_1 = \mathbf{0}$

$$(122) \Rightarrow \mathbf{U}^T\mathbf{y} = \mathbf{U}^T\mathbf{C}_1\mathbf{x}_1 + \mathbf{U}^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{U}^T\mathbf{n} \tag{123}$$

$$\Rightarrow \mathbf{U}^T\mathbf{y} = \mathbf{U}^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{U}^T\mathbf{n} \tag{124}$$

In what follows, we define different choices for $\mathbf{U}$:

A. Choice 1:

we obtain one choice for $\mathbf{U}$ is $\mathbf{Q}_2$

$$\underset{M\times N_1}{\mathbf{C}_1} = \underbrace{\begin{bmatrix} \underset{M\times N_1}{\mathbf{Q}_1} & \underset{M\times(M-N_1)}{\mathbf{Q}_2} \end{bmatrix}}_{\underset{M\times M}{\mathbf{Q}}} \underbrace{\begin{bmatrix} \underset{N_1\times N_1}{\mathbf{R}'} \\ \underset{(M-N_1)\times N_1}{\mathbf{0}} \end{bmatrix}}_{\underset{M\times N_1}{\mathbf{R}}} = \underset{(M\times N_1)}{\mathbf{Q}_1} \underset{(N_1\times N_1)}{\mathbf{R}'} \tag{125}$$

where

$$\mathbf{I} = \mathbf{Q}\mathbf{Q}^T = \mathbf{Q}_1\mathbf{Q}_1^T + \mathbf{Q}_2\mathbf{Q}_2^T \tag{126}$$

or

$$\mathbf{I} = \mathbf{Q}^T\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_1^T\mathbf{Q}_1 & \mathbf{Q}_1^T\mathbf{Q}_2 \\ \mathbf{Q}_2^T\mathbf{Q}_1 & \mathbf{Q}_2^T\mathbf{Q}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{127}$$

by using the fact that $\mathbf{Q}_1^T\mathbf{Q}_2 = \mathbf{0}$ and $\mathbf{Q}_2^T\mathbf{Q}_1 = \mathbf{0}$

Thus multiplying (122) by the left null space $\mathbf{Q}_2$ we can have:

$$\underset{(M-N_1)\times 1}{\mathbf{Q}_2^T\mathbf{y}} = \underset{[(M-N_1)\times M]}{\mathbf{Q}_2^T} \underset{(M\times N_2)}{\mathbf{C}_2} \underset{(N_2\times 1)}{\mathbf{x}_2} + \underset{(M-N_1)\times 1}{\mathbf{Q}_2^T\mathbf{n}} \tag{128}$$

where $\mathbf{Q}_2^T\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_2^T\mathbf{R}\mathbf{Q}_2)$

Note that if $\mathbf{R} = \sigma^2\mathbf{I} \Rightarrow \mathbf{Q}_2^T\mathbf{R}\mathbf{Q}_2 = \sigma^2\mathbf{I}_{(M-N_1)\times(M-N_1)}$

The resulting system in (128) is minimal and can be efficiently computed from the QR decomposition of $\mathbf{C}_1$ in (125).

Note that we must compute all of $\mathbf{Q}$ to get $\mathbf{Q}_2$. Thus, we cannot use ”thin” QR. We can, however (as we do in the MSCKF) employ given rotations and efficiently solve $\underset{(M-N_1)\times N_1}{\mathbf{0}} = \underset{[(M-N_1)\times M]}{\mathbf{Q}_2^T} \underset{(M\times N_1)}{\mathbf{C}_1}$

B. Choice 2:

Choose $\mathbf{U}$= $\mathbf{Q}'_2 = \mathbf{I}_M - \underset{(M\times N_1)}{\mathbf{Q}_1} \underset{(M\times N_1)}{\mathbf{Q}_1^T}$. Then $\mathbf{Q}'^T_2 = \mathbf{Q}'_2$ and $\mathbf{Q}'^T_2\mathbf{Q}'_1 = (\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{Q}_1^T = \mathbf{Q}_1 - \mathbf{Q}_1 = \mathbf{0}$

Multiplying $\mathbf{Q}'^T_2$ by (122), we have:

$$\underset{(M\times M)}{\mathbf{Q}'^T_2} \underset{(M\times 1)}{\mathbf{y}} = \underset{(M\times M)}{\mathbf{Q}'^T_2} \underset{(M\times N_2)}{\mathbf{C}_2} \underset{(N_2\times 1)}{\mathbf{x}_2} + \underset{(M\times M)}{\mathbf{Q}'^T_2} \underset{(M\times 1)}{\mathbf{n}} \tag{129}$$

with $\mathbf{Q}'^T_2\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}'^T_2\mathbf{R}\mathbf{Q}'_2)$

Note: If $\mathbf{R} = \sigma^2\mathbf{I}$ as is usually the case for measurements corresponding to visual feature observations, then:

$$\mathbf{Q}'^T_2\mathbf{R}\mathbf{Q}'_2 = \sigma^2(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T)(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T) \tag{130}$$

$$= \sigma^2(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T - \mathbf{Q}_1\mathbf{Q}_1^T + \mathbf{Q}_1\underbrace{\mathbf{Q}_1^T\mathbf{Q}_1}_{\mathbf{I}}\mathbf{Q}_1^T) \tag{131}$$

$$= \sigma^2(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T - \mathbf{Q}_1\mathbf{Q}_1^T + \mathbf{Q}_1\mathbf{Q}_1^T) \tag{132}$$

$$= \sigma^2(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T) \tag{133}$$

which is a new different covariance matrix.

The main limitation of this approach is that it results in a ”bloated” system with more equations than the independent ones (M in (129) instead of $M - N_1$ in (128). Some of these equations can be later removed by applying QR on $\mathbf{Q}_2^T\mathbf{C}_2$

Specifically, from (126) $\Rightarrow \mathbf{Q}'_2 = \mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T = \mathbf{Q}_2\mathbf{Q}_2^T$

and thus from (129)

$$\Rightarrow \mathbf{Q}_2\mathbf{Q}_2^T\mathbf{y} = \mathbf{Q}_2(\mathbf{Q}_2^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{Q}_2^T\mathbf{n}) \tag{134}$$

which is just (128) with extra equations due to the multiplication from the left with the tall matrix $\underset{(M\times N_1}{\mathbf{Q}_2}$ which ”mixed” the equations and create new ones or linear combination of the previous one.

From (129) we can go to (128) if we have $\mathbf{Q}_2$ and multiply $\mathbf{Q}_2^T$ both side of (134) as follow:

$$\underbrace{(\mathbf{Q}_2^T\mathbf{Q}_2)}_{\mathbf{I}}(\mathbf{Q}_2^T\mathbf{y}) = \underbrace{(\mathbf{Q}_2^T\mathbf{Q}_2)}_{\mathbf{I}}(\mathbf{Q}_2^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{Q}_2^T\mathbf{n}) \Leftrightarrow (128) \tag{135}$$

C. Choice 3:

$$\mathbf{y} = \mathbf{C}_1\mathbf{x}_1 + \mathbf{C}_2\mathbf{x}_2 \tag{136}$$

(110) can be written as

$$\underbrace{\mathbf{C}_2^T\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)}_{\mathbf{D}^T}\mathbf{x}_2 = \mathbf{C}_2^T\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\mathbf{y}. \tag{137}$$

Now consider again the equation:

$$\mathbf{y} = \mathbf{C}_1\mathbf{x}_1 + \mathbf{C}_2\mathbf{x}_2 \tag{138}$$

If we use $\mathbf{D}^T$ as $\mathbf{U}^T$ then:

$$\mathbf{D}^T\mathbf{C}_1 = \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1 - \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1 = \mathbf{0} \tag{139}$$

$$\Rightarrow \mathbf{D}^T\mathbf{y} = \mathbf{D}^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{D}^T\mathbf{n} \tag{140}$$

where $\mathbf{D}^T\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}^T\mathbf{R}\mathbf{D})$.

<u>*NOTE*1</u>: If $\mathbf{R} = \sigma^2\mathbf{I}$, then:

$$\mathbf{D}^T\mathbf{R}\mathbf{D} = \mathbf{C}_2^T\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\mathbf{R}\left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\mathbf{C}_2 \tag{141}$$

$$= \mathbf{C}_2^T\mathbf{R}^{-1}\left(\mathbf{I} - \mathbf{C}_1\left(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1\right)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right)\mathbf{R}\left(\mathbf{I} - \mathbf{R}^{-1}\mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\right)\mathbf{R}^{-1}\mathbf{C}_2 \tag{142}$$

$$= \mathbf{C}_2^T\mathbf{R}^{-1}\left[\mathbf{R} - \mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T - \mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T + \mathbf{C}_1(\mathbf{C}_1\mathbf{R}^{-1}\mathbf{C}_1^T)^{-1}\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}^T\right]\mathbf{R}^{-1}\mathbf{C}_2 \tag{143}$$

$$= \mathbf{C}_2^T\mathbf{R}^{-1}\left[\mathbf{R} - \mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}^T\right]\mathbf{R}^{-1}\mathbf{C}_2 \tag{144}$$

$$= \mathbf{C}_2^T(\frac{1}{\sigma^2}\mathbf{I})\left[\sigma^2\mathbf{I} - \sigma^2\mathbf{C}_1(\mathbf{C}_1^T\mathbf{C}_1)^{-1}\mathbf{C}_1^T\right](\frac{1}{\sigma^2}\mathbf{I})\mathbf{C}_2 = \frac{1}{\sigma^2}\mathbf{C}_2^T\left[\mathbf{I} - \mathbf{C}_1(\mathbf{C}_1^T\mathbf{C}_1)^{-1}\mathbf{C}^T\right]\mathbf{C}_2 \tag{145}$$

Employing (125), we have

$$\mathbf{C}_1^T\mathbf{C}_1 = \mathbf{R}'^T\mathbf{Q}_1^T\mathbf{Q}_1\mathbf{R}' = \mathbf{R}'^T\mathbf{R}' \tag{146}$$

$$\Rightarrow \mathbf{D}^T\mathbf{R}\mathbf{D} = \frac{1}{\sigma^2}\mathbf{C}_2^T\left[\mathbf{I} - \mathbf{Q}_1\mathbf{R}'(\mathbf{R}'^T\mathbf{R}')^{-1}\mathbf{R}'^T\mathbf{Q}_1^T\right]\mathbf{C}_2 \tag{147}$$

$$= \frac{1}{\sigma^2}\mathbf{C}_2^T\left[\mathbf{I} - \mathbf{Q}_1\mathbf{R}'\mathbf{R}'^{-1}\mathbf{R}'^{-T}\mathbf{R}'^T\mathbf{Q}_1^T\right]\mathbf{C}_2 \tag{148}$$

$$= \frac{1}{\sigma^2}\mathbf{C}_2^T\left(\mathbf{I} - \mathbf{Q}_1\mathbf{Q}_1^T\right)\mathbf{C}_2 \tag{149}$$

$$= \frac{1}{\sigma^2}\mathbf{C}_2^T\mathbf{Q}_2\mathbf{Q}_2^T\mathbf{C}_2 \tag{150}$$

$$\tag{151}$$

In general, the condition number of this matrix is worse than that of $\sigma^2\mathbf{I}_{\mathbf{M}-\mathbf{N}_1}$

Now, we examine the $\mathbf{D}$ matrix:

Note 2:

$$\mathbf{D}^T = \mathbf{C}_2^T\mathbf{R}^{-1}\left[\mathbf{I}-\mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\right] \tag{152}$$

$$= \frac{1}{\sigma^2}\mathbf{C}_2^T\left[\mathbf{I}-\mathbf{C}_1(\mathbf{C}_1^T\mathbf{C}_1)^{-1}\mathbf{C}_1^T\right] \tag{153}$$

$$= \frac{1}{\sigma^2}\mathbf{C}_2^T\left[\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T\right] \tag{154}$$

In "Choice 2" we used

$$\mathbf{Q}'^T = \mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T. \tag{155}$$

Note 3:
In "Choice 1" we used $\mathbf{U}^T = \mathbf{Q}_2^T = \mathbf{I}\mathbf{Q}_2^T = \mathbf{Q}_2^T(\mathbf{Q}_2\mathbf{Q}_2^T) = \mathbf{Q}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)$. Considering $\frac{1}{\sigma^2}$ in (154) which is just a scaling. The choices (154) and (155) are "similar" in terms of dimesion.

$\mathbf{C}_2^T : N_2 \times M$

$\mathbf{Q}_2^T : (M-N_1)\times M$

Revisiting the two different choices for reforming marginalization (125) we have:

$$\mathbf{y} = \mathbf{C}_1\mathbf{x}_1 + \mathbf{C}_2\mathbf{x}_2 + \mathbf{n} \tag{156}$$

$$\Rightarrow \mathbf{y} = \mathbf{Q}_1\mathbf{R}'\mathbf{x}_1 + \mathbf{C}_2\mathbf{x}_2 + \mathbf{n} \tag{157}$$

(1)

$$\mathbf{D}^T\mathbf{y} = \mathbf{D}^T\mathbf{Q}_1\mathbf{R}'\mathbf{x}_1 + \mathbf{D}^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{D}^T\mathbf{n} \tag{158}$$

$$\Rightarrow \frac{1}{\sigma^2}\mathbf{C}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{y} = \frac{1}{\sigma^2}\mathbf{C}_2^T\underbrace{(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{Q}_1}_{\mathbf{0}}\mathbf{R}'\mathbf{x}_1 + \frac{1}{\sigma^2}\mathbf{C}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{C}_2\mathbf{x}_2 + \frac{1}{\sigma^2}\mathbf{C}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{n} \tag{159}$$

$$\Leftrightarrow \mathbf{C}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{y} = \underset{N_2\times N_2}{\mathbf{C}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{C}_2}\underset{N_2\times 1}{\mathbf{x}_2} + \mathbf{C}_2^T(\mathbf{I}-\mathbf{Q}_1\mathbf{Q}_1^T)\mathbf{n} \tag{160}$$

which is square system versus:

$$\mathbf{Q}_2^T\mathbf{y} = \mathbf{Q}_2^T\mathbf{Q}_1\mathbf{R}'\mathbf{x}_1 + \mathbf{Q}_2^T\mathbf{C}_2\mathbf{x}_2 + \mathbf{Q}_2^T\mathbf{n} \tag{161}$$

$$\Leftrightarrow \underset{(M-N_1)\times 1}{\mathbf{Q}_2^T\mathbf{y}} = \underset{(M-N_1)\times M}{\mathbf{Q}_2^T}\underset{(M\times N_2)}{\mathbf{C}_2}\underset{(N_2)\times 1}{\mathbf{x}_2} + \mathbf{Q}_2^T\mathbf{n} \tag{162}$$

In general: M $\gg$ N_1+N_2 $\Rightarrow$ M-N_1 $\gg$ N_2 $\Rightarrow$ $\mathbf{Q}_2^T\mathbf{C}_2$ is a tall matrix and we will need to apply QR on that to perform measurement compression and result into a square system similar to that of (160) but with better numerical properties.

IV. Apply marginalization to the Jacobian of the localization and mapping problem

As mention in the section of Hessian structure, each of the Jacobian($\mathbf{H}_{ij}$ and $\overline{\mathbf{H}}_{i+1,i}$) corresponding to proprioceptive measurements has two consecutive nonzero block elements. Corresponding to the starting and end pose being linked.

This measurement (or continuation of multiple proprioceptive measurement between 2 pose at which exteroceptive measurement) were recorded.

Also, the exteroceptive measurement, Jacobian has 2 nonzero block elements:

NOTE: If a set of calibration parameters are considered (camera intrinsics) ;time synchronization,etc.) that are constant over time (i.e they are not part of the state vectors $\mathbf{r}_i$ and $\mathbf{f}_j$, and we call them $\boldsymbol{\theta}$, then we may have the block elements in both above Jacobians (with respect to $\boldsymbol{\theta}$ or a subset of them).

For simplicity, from now on we wil assume that we only have two nonzero block elements in Jacobian. In which case of we use the linearized model for each of them; we have (from 70):

Note that unlike in section Hessian, there we consider back to the case that have $\boldsymbol{\delta}\mathbf{y}$ and $\tilde{\mathbf{x}}$ for a better intuition:

$$\boldsymbol{\delta}\mathbf{y}_k - \mathbf{H}_k\tilde{\mathbf{x}} = \mathbf{n}_k \quad ; \quad k = 1, 2, ...K \tag{163}$$

or

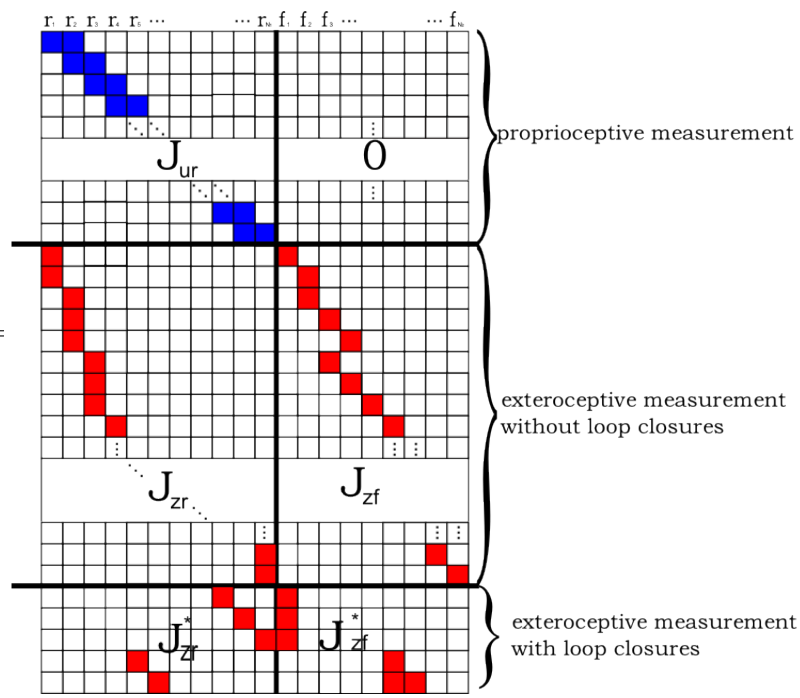
$$\mathbf{R}_k^{-\frac{1}{2}}\boldsymbol{\delta}\mathbf{y}_k = \mathbf{R}_k^{-\frac{1}{2}}\mathbf{H}_k\tilde{\mathbf{x}} + \underbrace{\mathbf{R}_k^{-\frac{1}{2}}\mathbf{n}_k}_{\mathbf{n}'_k}; \quad \mathbf{n}_k^{'} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{164}$$

stack all of the quantity from k=1,2,.....K in (164) we have:

$$\underbrace{\begin{bmatrix} \mathbf{R}_1^{-\frac{1}{2}}\mathbf{H}_1 \\ \mathbf{R}_2^{-\frac{1}{2}}\mathbf{H}_2 \\ \vdots \\ \mathbf{R}_K^{-\frac{1}{2}}\mathbf{H}_K \end{bmatrix}}_{\mathbf{J}:\ Jacobian} \tilde{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{R}_1^{-\frac{1}{2}}\boldsymbol{\delta}\mathbf{y}_1 \\ \mathbf{R}_2^{-\frac{1}{2}}\boldsymbol{\delta}\mathbf{y}_2 \\ \vdots \\ \mathbf{R}_K^{-\frac{1}{2}}\boldsymbol{\delta}\mathbf{y}_K \end{bmatrix}}_{residual} + \underbrace{\begin{bmatrix} \mathbf{n}_1^{'} \\ \mathbf{n}_2^{'} \\ \vdots \\ \mathbf{n}_K^{'} \end{bmatrix}}_{error} \tag{165}$$

From here we will concentrate on the structure of the Jacobian after we assume that $\tilde{\mathbf{x}}$ is partitioned as $\tilde{\mathbf{x}}= \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix}$

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}_{zr} & \mathbf{J}_{zf} \\ \mathbf{J}_{ur}^{*} & \mathbf{J}_{zf}^{*} \end{bmatrix} =$$



About size of Jacobian matrix:

$$\begin{array}{c} M_1 \{ \\ M_2 \{ \\ M_3 \{ \end{array} \underbrace{\begin{bmatrix} \mathbf{J_{ur}} \\ \mathbf{J_{zr}} \\ \mathbf{J^{*}{}_{zr}} \end{bmatrix}}_{N_1} \underbrace{\begin{bmatrix} \mathbf{0} \\ \mathbf{J}_{zf} \\ \mathbf{J}^{*}{}_{zf} \end{bmatrix}}_{N_2}$$

Our first objective is to "get rid of" $\mathbf{J}_{zf}$. Rewriting (165) we have:

$$\begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}_{zr} & \mathbf{J}_{zf} \\ \mathbf{J}_{zr}^{*} & \mathbf{J}_{zf}^{*} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\delta}\mathbf{y}_u \\ \boldsymbol{\delta}\mathbf{y}_z \\ \boldsymbol{\delta}\mathbf{y}_z^{*} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_u \\ \mathbf{n}_z \\ \mathbf{n}_z^{*} \end{bmatrix} \tag{166}$$

$$(166) \Rightarrow \mathbf{J}_{zr}\tilde{\mathbf{r}} + \underset{(M_2\times N_2)}{\mathbf{J}_{zf}}\tilde{\mathbf{f}} = \delta\mathbf{y}_z + \mathbf{n}_z,\ \mathbf{n}_z \sim \mathcal{N}(\mathbf{0},\mathbf{I}). \tag{167}$$

Do QR factorization on $\mathbf{J}_{zf}$:

$$\mathbf{J}_{zf} = \begin{bmatrix} \underset{(M_2\times N_2)}{\mathbf{Q}_1} & \underset{(M_2\times(M_2-N_2))}{\mathbf{Q}_2} \end{bmatrix} \begin{bmatrix} \mathbf{R}' \\ \mathbf{0} \end{bmatrix} = \mathbf{Q}_1 \underset{(N_2\times N_2)}{\mathbf{R}'} \tag{168}$$

Multiply $\mathbf{Q}_2^T$ to both sides of (167) we obtain:

$$\mathbf{Q}_2^T\mathbf{J}_{zr}\tilde{\mathbf{r}} + \underbrace{\mathbf{Q}_2^T\mathbf{J}_{zf}}_{\mathbf{0}}\tilde{\mathbf{f}} = \mathbf{Q}_2^T\delta\mathbf{y}_z + \underbrace{\mathbf{Q}_2^T\mathbf{n}_z}_{\mathbf{n}'_z} \quad ; \quad \mathbf{n}_z \sim \mathcal{N}(\mathbf{0},\mathbf{Q}_2^T\mathbf{Q}_2 = \mathbf{I}) \tag{169}$$

NOTE: This problem can be solved very efficiently if we collect all observations of each feature and apply given rotations as we did in MSCKF:

Specifically, assume that each feature $\mathbf{f}_j$ is "viewed" by poses $\mathbf{r}_{i1}, \mathbf{r}_{i2}, ..., \mathbf{r}_{i\Psi}$ then we collect and stack together to create the new Jacobian:

$$\begin{bmatrix} \mathbf{R}_{i_1 j}^{-\frac{1}{2}}\mathbf{H}_{i_1 j} \\ \mathbf{R}_{i_2 j}^{-\frac{1}{2}}\mathbf{H}_{i_2 j} \\ \vdots \\ \mathbf{R}_{i_\Psi j}^{-\frac{1}{2}}\mathbf{H}_{i_\Psi j} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \delta\mathbf{y}_{i_1 j} \\ \delta\mathbf{y}_{i_2 j} \\ \vdots \\ \delta\mathbf{y}_{i_\Psi j} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{i_1 j} \\ \mathbf{n}_{i_2 j} \\ \vdots \\ \mathbf{n}_{i_\Psi j} \end{bmatrix} \tag{170}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_{i_1 j}^{-\frac{1}{2}}\mathbf{H}_{i_1 j}^r & \mathbf{0} & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_{i_1 j}^{-\frac{1}{2}}\mathbf{H}_{i_1 j}^f & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{R}_{i_2 j}^{-\frac{1}{2}}\mathbf{H}_{i_2 j}^r & \dots & \mathbf{0} & \dots & \mathbf{0} & \mathbf{R}_{i_2 j}^{-\frac{1}{2}}\mathbf{H}_{i_2 j}^f & \mathbf{0} & \dots & \mathbf{0} \\ \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_{i_\Psi j}^{-\frac{1}{2}}\mathbf{H}_{i_\Psi j}^r & \dots & \mathbf{0} & \mathbf{R}_{i_\Psi j}^{-\frac{1}{2}}\mathbf{H}_{i_\Psi j}^f & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}}_1 \\ \vdots \\ \tilde{\mathbf{r}}_n \\ \tilde{\mathbf{f}}_1 \\ \vdots \\ \tilde{\mathbf{f}}_m \end{bmatrix} = \begin{bmatrix} \delta\mathbf{y}_{i_1 j} \\ \delta\mathbf{y}_{i_2 j} \\ \vdots \\ \delta\mathbf{y}_{i_\Psi j} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{i_1 j} \\ \mathbf{n}_{i_2 j} \\ \vdots \\ \mathbf{n}_{i_\Psi j} \end{bmatrix} \tag{171}$$

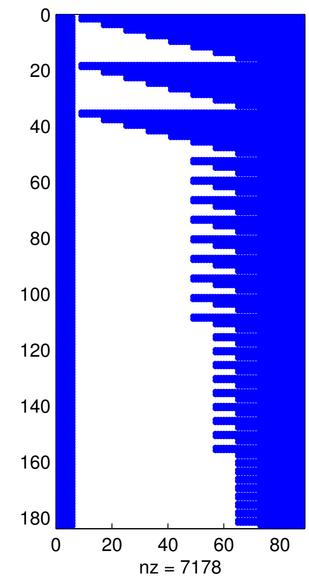
$$\Leftrightarrow \begin{bmatrix} \mathbf{R}_{i_1 j}^{-\frac{1}{2}}\mathbf{H}_{i_1 j}^r & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{i_2 j}^{-\frac{1}{2}}\mathbf{H}_{i_2 j}^r & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_{i_\Psi j}^{-\frac{1}{2}}\mathbf{H}_{i_\Psi j}^r \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}}_{i1} \\ \tilde{\mathbf{r}}_{i2} \\ \vdots \\ \tilde{\mathbf{r}}_{i\Psi} \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{R}_{i_1 j}^{-\frac{1}{2}}\mathbf{H}_{i_1 j}^f \\ \mathbf{R}_{i_2 j}^{-\frac{1}{2}}\mathbf{H}_{i_2 j}^f \\ \vdots \\ \mathbf{R}_{i_\Psi}^{-\frac{1}{2}}\mathbf{H}_{i_\Psi j}^f \end{bmatrix}}_{\mathbf{J}_{zrj}} \underset{3\times 1}{\tilde{\mathbf{f}}_j} = \begin{bmatrix} \delta\mathbf{y}_{i_1 j} \\ \delta\mathbf{y}_{i_2 j} \\ \vdots \\ \delta\mathbf{y}_{i_\Psi j} \end{bmatrix} + \begin{bmatrix} \mathbf{n}_{i_1 j} \\ \mathbf{n}_{i_2 j} \\ \vdots \\ \mathbf{n}_{i_\Psi j} \end{bmatrix} \tag{172}$$

Which we have obtained a very small matrix $\mathbf{J}_{zf_j}(4\times 3)$ compared to $\mathbf{J}_{zf}(M_2\times N_2)$. Then we will use the same method as QR factorization for $\mathbf{J}_{zf}$ to obtain the left null space of $\mathbf{J}_{zf_j}$; denote it as $\mathbf{U}_{zf_j}$ such that : $\mathbf{U}_{zf_j}^T\ \mathbf{J}_{zf_j}$=$\mathbf{0}$.

$$(172) \Rightarrow \underbrace{\mathbf{U}_{zf_j}^T\mathbf{J}_{zr_j}}_{\mathbf{J}_{zr_j}^{\mathbf{U}}} \begin{bmatrix} \tilde{\mathbf{r}}_{i_1} \\ \tilde{\mathbf{r}}_{i_2} \\ \vdots \\ \tilde{\mathbf{r}}_{i_\Psi} \end{bmatrix} + \underbrace{\mathbf{U}_{zf_j}^T\mathbf{J}_{zf_j}}_{\mathbf{0}}\tilde{\mathbf{f}}_j = \mathbf{u}_{zf_j}^T \begin{bmatrix} \delta\mathbf{y}_{i_1 j} \\ \delta\mathbf{y}_{i_2 j} \\ \vdots \\ \delta\mathbf{y}_{i_\Psi j} \end{bmatrix} + \mathbf{u}_{zf_j}^T \underbrace{\begin{bmatrix} \mathbf{n}_{i_1 j} \\ \mathbf{n}_{i_2 j} \\ \vdots \\ \mathbf{n}_{i_\Psi j} \end{bmatrix}}_{\mathbf{n}'\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \tag{173}$$

Notes:

1. If $\mathbf{r}_{i_1};\mathbf{r}_{i_2};...;\mathbf{r}_{i_\Psi}$ are consecutive poses; then $\mathbf{J}_{zr_j}^{\mathbf{U}}$ will have exactly the same structure as the matrix we get after we apply given rotation in the MSCKF case:

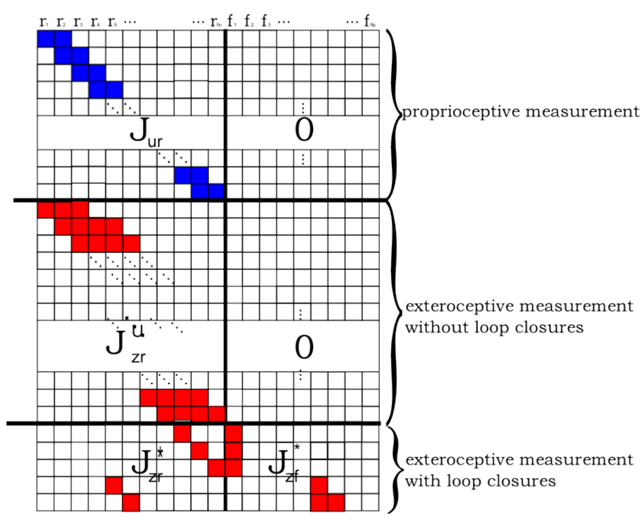


2. If $\mathbf{r}_{i_1};\mathbf{r}_{i_2};...;\mathbf{r}_{i_\Psi}$ are NOT consecutive poses. then $\mathbf{J}^{\mathbf{U}}_{zr_j}$ will have similar structure as in 1 but with spaces in between corresponding to poses that did not view the feature.

3. If in (173) we recover the full poses $\begin{bmatrix} \tilde{\mathbf{r}}_{i1} \\ \tilde{\mathbf{r}}_{i2} \\ \vdots \\ \tilde{\mathbf{r}}_{i\Psi} \end{bmatrix} \rightarrow \tilde{\mathbf{r}}$, then the corresponding $\mathbf{J}^{\mathbf{U}}_{zr}$ will be a sparse matrix with $2\mathbf{M}_{2j} \times 3$ rows, where $\mathbf{M}_{2j}$ is the member of poses that observed features j and $\mathbf{M}_{2j}$ nonzero block columns .

Employing the same process to marginalized all feature, we will obtain the new Jacobian: $\mathbf{J}'$

$$\mathbf{J}' = \begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}^{\mathbf{U}}_{zr} & \mathbf{0} \\ \mathbf{J}^{*}_{zr} & \mathbf{J}^{*}_{zf} \end{bmatrix} =$$



Note that $\mathbf{J}^{\mathbf{U}}_{zr}$ corresponds to the ”regular” odometry (e.g from IMU) where consecutive poses are ”linked”. $\mathbf{J}^{\mathbf{U}}_{zr}$ corresponds to visual odometry where multiple poses may be ”linked”.

Going back to the original Jacobian, the marginalization process we just describe can be expressed in a compact form as a left multiplication of:

$$\begin{bmatrix} \mathbf{I} & & \\ & \mathbf{U}_{zf}^T & \\ & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}_{zr} & \mathbf{J}_{zf} \\ \mathbf{J}_{zr}^* & \mathbf{J}_{zf}^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & & \\ & \mathbf{U}_{zf}^T & \\ & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\mathbf{y}_u \\ \delta\mathbf{y}_z \\ \delta\mathbf{y}_z^* \end{bmatrix} + \begin{bmatrix} \mathbf{I} & & \\ & \mathbf{U}_{zf}^T & \\ & & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{n}_u \\ \mathbf{n}_z \\ \mathbf{n}_z^* \end{bmatrix} \tag{174}$$

$$\Leftrightarrow \begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}_{zr}^{\mathbf{U}} & \mathbf{0} \\ \mathbf{J}_{zr}^* & \mathbf{J}_{zf}^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \delta\mathbf{y}_u \\ \delta\mathbf{y}_z^{\mathbf{U}} \\ \delta\mathbf{y}_z^* \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{n}_u \\ \mathbf{n}_z^{\mathbf{U}} \\ \mathbf{n}_z^* \end{bmatrix}}_{\mathbf{n}'\sim\mathcal{N}(\mathbf{0},\mathbf{I})} , where \quad \mathbf{U}_{zf}^T = \begin{bmatrix} \mathbf{U}_{zf_1}^T & & \\ & \ddots & \\ & & \mathbf{U}_{zf_{M_2}}^T \end{bmatrix} \tag{175}$$

Note that the upper two blocks of the Jacobian $\mathbf{J}' \begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}_{zr}^{\mathbf{U}} & \mathbf{0} \end{bmatrix}$ provides constraints only between the state $\tilde{\mathbf{r}}$ and also that it will be in general a tall matrix.

To condense the available information from all measurements (measurement compression), we employ QR factorization:

$$\begin{bmatrix} \mathbf{J}_{ur} \\ \mathbf{J}_{zr}^{\mathbf{U}} \end{bmatrix} = \underset{(M_1+M_2-N_2)\times N_2}{\mathbf{Q}_1} \underset{N_2\times N_2}{\mathbf{R}_0} \tag{176}$$

where R_0 is in general block upper diagonal band with the band equal to the number of feature tracking length.

Going back to (175):

$$\begin{bmatrix} (\mathbf{Q}_1^o)^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{J}_{ur} & \mathbf{0} \\ \mathbf{J}_{zr}^{\mathbf{u}} & \mathbf{0} \\ \mathbf{J}_{zu}^* & \mathbf{J}_{zf}^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} (\mathbf{Q}_1^o)^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \delta\mathbf{y}_u \\ \delta\mathbf{y}_z^{\mathbf{u}} \\ \delta\mathbf{y}_u^* \end{bmatrix} + \begin{bmatrix} (\mathbf{Q}_1^o)^T & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{n}_u \\ \mathbf{n}_z^{\mathbf{U}} \\ \mathbf{n}_n^* \end{bmatrix} \tag{177}$$

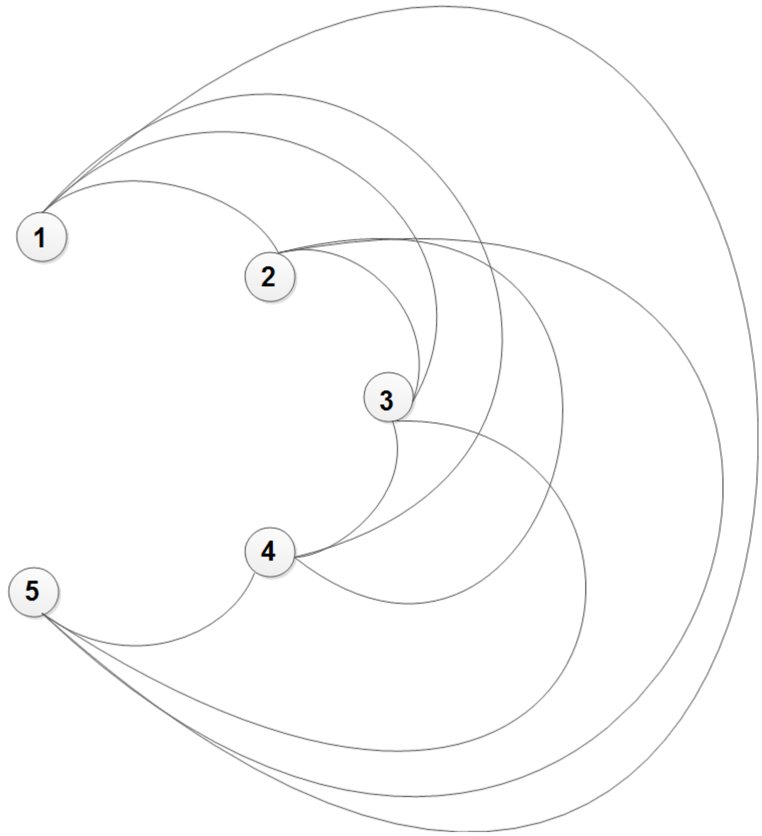
$$\begin{bmatrix} \mathbf{R}_0 & \mathbf{0} \\ \mathbf{J}_{zr}^* & \mathbf{J}_{zf}^* \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{r}} \\ \tilde{\mathbf{f}} \end{bmatrix} = \begin{bmatrix} \delta\mathbf{y}_0 \\ \delta\mathbf{y}_z \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{n}_0 \\ \mathbf{n}_z \end{bmatrix}}_{\mathbf{n}\sim\mathcal{N}(\mathbf{0},\mathbf{I})} \tag{178}$$

Notes:

1. The upper portion of (178) corresponds to all the form of odometry available to the robot.

$$\mathbf{R}_0\tilde{\mathbf{r}} = \delta\mathbf{y}_0 + \mathbf{n}_0 \tag{179}$$

If we were to depict this with graph we would have:



* This is not exactly right; The constraint coupling many poses cannot be represented as a high degree node so that we need to decompose it in multiple isolated constraints.

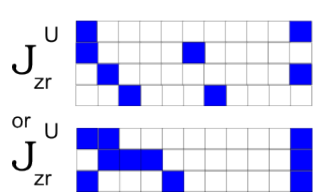
2. If we were to (we want to follow the similar process as Gaussian Elimination block as before) marginalize the feature (the lower part of 178):

$$\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & (\mathbf{U}^*_{zf})^T\end{bmatrix}\begin{bmatrix}\mathbf{R}_0 & \mathbf{0}\\ \mathbf{J}^*_{zr} & \mathbf{J}^*_{zf}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{r}}\\ \tilde{\mathbf{f}}\end{bmatrix} = \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & (\mathbf{U}^*_{zf})^T\end{bmatrix}\begin{bmatrix}\delta\mathbf{y}_0\\ \delta\mathbf{y}_z\end{bmatrix} + \begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & (\mathbf{U}^*_{zf})^T\end{bmatrix}\begin{bmatrix}\mathbf{n}_0\\ \mathbf{n}_z\end{bmatrix} \tag{180}$$

$$\Rightarrow \begin{bmatrix}\mathbf{R}_0 & \mathbf{0}\\ \mathbf{J}^{*\mathbf{U}}_{zr} & \mathbf{0}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{r}}\\ \tilde{\mathbf{f}}\end{bmatrix} = \begin{bmatrix}\delta\mathbf{y}_0\\ (\mathbf{U}^*_{zf})^T\delta\mathbf{y}_z\end{bmatrix} + \begin{bmatrix}\mathbf{n}_0\\ (\mathbf{U}^*_{zf})^T\mathbf{n}_z\end{bmatrix} \tag{181}$$

$$\Rightarrow \begin{bmatrix}\mathbf{R}_0\\ \mathbf{J}^{*\mathbf{U}}_{zr}\tilde{\mathbf{r}}\end{bmatrix}\tilde{\mathbf{r}} = \begin{bmatrix}\delta\mathbf{y}_0\\ (\mathbf{U}^*_{zf})^T\delta\mathbf{y}_z\end{bmatrix} + \begin{bmatrix}\mathbf{n}_0\\ (\mathbf{U}^*_{zf})^T\mathbf{n}_z\end{bmatrix} \tag{182}$$

Then the matrix $\mathbf{J}^*_{zr}$ will have "long distance" constraints. That is it will couple robot states that are far apart; e.g:



If we were were to apply measurement compression QR again on :

$$\begin{bmatrix}\mathbf{R}_0\\ \mathbf{J}_{zr}^{\mathbf{U}}\end{bmatrix} = \overline{\mathbf{Q}}\ \overline{\mathbf{R}} \tag{183}$$

the resulting $\overline{\mathbf{R}}$ will be "full" upper triangular matrix; that is:

$$(183) \Rightarrow \overline{\mathbf{Q}}^T \begin{bmatrix}\mathbf{R}_0\\ \mathbf{J}_{zr}\end{bmatrix} \tilde{\mathbf{r}} = \overline{\mathbf{Q}}^T \begin{bmatrix}\delta\mathbf{y}_0\\ (\mathbf{U}_{zf}^*)^T \delta\mathbf{y}_z\end{bmatrix} + \overline{\mathbf{Q}}^T \begin{bmatrix}\mathbf{n}_0\\ (\mathbf{U}_{zf}^*)^T \mathbf{n}_z\end{bmatrix} \tag{184}$$

$$\Leftrightarrow \quad \overline{\mathbf{R}}\tilde{\mathbf{r}} = \delta\overline{\mathbf{y}} + \overline{\mathbf{n}}\ ;\ \overline{\mathbf{n}} \sim \mathcal{N}(\mathbf{0},\mathbf{I}) \tag{185}$$

which we can solve for $\tilde{\mathbf{r}}$ and use back substitute in earlier equations to compute $\bar{\mathbf{f}}$.

APPENDIX

In this section, we will prove equation (117) in Hessian domain is equivalent to taking the integral over $\mathbf{x}_1$ of $\mathbf{p}(\mathbf{x}_1;\mathbf{x}_2)$ to get the marginal pdf $\mathbf{p}(\mathbf{x}_2)$. Where $\mathbf{x}_1;\mathbf{x}_2$ are all Gaussian vector and $\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix}$ is jointly Gaussian vector.

A. Relationship between Information and covariance form of Gaussian pdf

Denote x is Gaussian vector:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{P}) \Leftrightarrow \mathbf{p}(x) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{P}|^{\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{186}$$

$$(186) \Leftrightarrow \mathbf{p}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{P}|^{\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}\mathbf{x}^T\mathbf{P}^{-1}\mathbf{x} + \mathbf{x}^T\mathbf{P}^{-1}\boldsymbol{\mu} - \frac{1}{2}\boldsymbol{\mu}^T\mathbf{P}^{-1}\boldsymbol{\mu}\right) \tag{187}$$

Denote $\mathcal{I}$=$\mathbf{P}^{-1}$ as information matrix, $\boldsymbol{\nu} = \mathbf{P}^{-1}\boldsymbol{\mu}$ as information vector

$$\boldsymbol{\mu}^T\mathbf{P}^T\boldsymbol{\mu} = \boldsymbol{\mu}^T\mathbf{P}^{-1}\mathbf{P}\mathbf{P}^{-1}\boldsymbol{\mu} = (\mathbf{P}^{-1}\boldsymbol{\mu})^T\mathbf{P}(\mathbf{P}^{-1}\boldsymbol{\mu}) = \boldsymbol{\nu}^T\mathcal{I}^{-1}\boldsymbol{\nu} \tag{188}$$

$$(187) \Leftrightarrow \mathbf{p}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathcal{I}|^{-\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}\mathbf{x}^T\mathcal{I}\mathbf{x} + \mathbf{x}^T\boldsymbol{\nu} - \frac{1}{2}\boldsymbol{\nu}^T\mathcal{I}^{-1}\boldsymbol{\nu}\right) \tag{189}$$

OR we can denote:

$$\mathbf{x} \sim \mathcal{N}^{-1}(\boldsymbol{\nu},\mathcal{I}) \tag{190}$$

For the following next sections; $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu},\mathbf{P})$ and $\mathbf{x} \sim \mathcal{N}^{-1}(\boldsymbol{\nu},\mathcal{I})$ refer to the same

$$\mathbf{p}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{P}|^{\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathcal{I}|^{-\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}\mathbf{x}^T\mathcal{I}\mathbf{x} + \mathbf{x}^T\boldsymbol{\nu} - \frac{1}{2}\boldsymbol{\nu}^T\mathcal{I}^{-1}\boldsymbol{\nu}\right) \tag{191}$$

where $\mathcal{I} = \mathbf{P}^{-1}$ and $\boldsymbol{\nu} = \mathbf{P}^{-1}\boldsymbol{\mu}$

B. Marginal Gaussian pdf

Lemma 1: Suppose $\mathbf{x} = \begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} \sim \mathcal{N}\left[\underbrace{\begin{pmatrix}\boldsymbol{\mu}_1\\ \boldsymbol{\mu}_2\end{pmatrix}}_{\boldsymbol{\mu}},\ \underbrace{\begin{pmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{pmatrix}}_{\mathbf{P}}\right]$ Prove that $\int_{\mathbf{x}_1} \mathbf{p}(\mathbf{x}_1;\mathbf{x}_2)\mathrm{d}\mathbf{x}_1 = \mathbf{p}(\mathbf{x}_2)$, $\mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{P}_{22})$

Proof:

$$\mathbf{p}(\mathbf{x}_1,\mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{n_1+n_2}{2}}|\mathbf{P}|^{\frac{1}{2}}}\mathbf{exp}\underbrace{\left(-\frac{1}{2}\begin{bmatrix}\mathbf{x}_1-\boldsymbol{\mu}_1\\ \mathbf{x}_2-\boldsymbol{\mu}_2\end{bmatrix}^T\begin{bmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{x}_1-\boldsymbol{\mu}_1\\ \mathbf{x}_2-\boldsymbol{\mu}_2\end{bmatrix}\right)}_{\boldsymbol{\alpha}(\mathbf{x})} \tag{192}$$

$$\boldsymbol{\alpha}(\mathbf{x}_1;\mathbf{x}_2) = \begin{bmatrix}\mathbf{x}_1-\boldsymbol{\mu}_1\\ \mathbf{x}_2-\boldsymbol{\mu}_2\end{bmatrix}^T\begin{bmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{x}_1-\boldsymbol{\mu}_1\\ \mathbf{x}_2-\boldsymbol{\mu}_2\end{bmatrix} \tag{193}$$

$$= \begin{bmatrix}(\mathbf{x}_1-\boldsymbol{\mu}_1)^T & (\mathbf{x}_2-\boldsymbol{\mu}_2)^T\end{bmatrix}\begin{bmatrix}(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1} & -(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\\ -\mathbf{P}_{22}^{-1}\mathbf{P}_{21}(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1} & (\mathbf{P}_{22}-\mathbf{P}_{21}\mathbf{P}_{11}^{-1}\mathbf{P}_{12})^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1-\boldsymbol{\mu}_1\\ \mathbf{x}_2-\boldsymbol{\mu}_2\end{bmatrix} \tag{194}$$

$$= (\mathbf{x}_1-\boldsymbol{\mu}_1)^{-1}(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1)- \tag{195}$$

$$2(\mathbf{x_1}-\boldsymbol{\mu}_1)^T(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)+(\mathbf{x}_2-\boldsymbol{\mu}_2)^T(\mathbf{P}_{22}-\mathbf{P}_{21}\mathbf{P}_{11}^{-1}\mathbf{P}_{12})^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)$$

Using matrix inversion lemma:

$$(\mathbf{P}_{22}-\mathbf{P}_{21}\mathbf{P}_{11}^{-1}\mathbf{P}_{12})^{-1} = \mathbf{P}_{22}^{-1}+\mathbf{P}_{22}^{-1}\mathbf{P}_{21}(\mathbf{P}_{11}-\mathbf{P}_{22}\mathbf{P}_{21}^{-1}\mathbf{P}_{21})^{-1}\mathbf{P}_{12}\mathbf{P}_{22}^{-1} \tag{196}$$

$$\tag{197}$$

$$\Rightarrow \boldsymbol{\alpha}(\mathbf{x}_1;\mathbf{x}_2) = (\mathbf{x}_1-\boldsymbol{\mu}_1)^T(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1)-2(\mathbf{x}_1-\boldsymbol{\mu}_1)^T(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}\left[\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)\right]$$

$$+\left[(\mathbf{x}_2-\boldsymbol{\mu}_2)^T\mathbf{P}_{22}^{-1}\mathbf{P}_{21}\right](\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}\left[\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)\right]+(\mathbf{x}_2-\boldsymbol{\mu}_2)^T\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2) \tag{198}$$

$$= (\mathbf{x}_1-\boldsymbol{\mu}_1-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2))^T(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})^{-1}(\mathbf{x}_1-\boldsymbol{\mu}_1-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2))+(\mathbf{x}_2-\boldsymbol{\mu}_2)^T\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2) \tag{199}$$

$$|\mathbf{P}| = \underbrace{\left|\begin{bmatrix}\mathcal{I} & -\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\\ \mathbf{0} & \mathcal{I}\end{bmatrix}\right|}_{1}\left|\begin{bmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{bmatrix}\right| = \left|\begin{bmatrix}\mathcal{I} & -\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\\ \mathbf{0} & \mathcal{I}\end{bmatrix}\begin{bmatrix}\mathbf{P}_{11} & \mathbf{P}_{12}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{bmatrix}\right| = \begin{vmatrix}\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21} & \mathbf{0}\\ \mathbf{P}_{21} & \mathbf{P}_{22}\end{vmatrix} = |\mathbf{P}_{22}|\,|\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}| \tag{200}$$

$$\Rightarrow \mathbf{p}(\mathbf{x}_1;\mathbf{x}_2) = \frac{1}{(2\pi)^{\frac{n}{2}}|\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21}|^{\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}_1-\boldsymbol{\mu}_1-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2))^T(\mathbf{P}_{11}-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}\mathbf{P}_{21})(\mathbf{x}_1-\boldsymbol{\mu}_1-\mathbf{P}_{12}\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2))\right)$$

$$\times\frac{1}{(2\pi)^{\frac{n_2}{2}}|\mathbf{P}_{22}|^{\frac{1}{2}}}\mathbf{exp}\left(-\frac{1}{2}(\mathbf{x}_2-\boldsymbol{\mu}_2)^T\mathbf{P}_{22}^{-1}(\mathbf{x}_2-\boldsymbol{\mu}_2)\right) \tag{201}$$

$$= \mathbf{p}(\mathbf{x}_1|\mathbf{x}_2)\mathbf{p}(\mathbf{x}_2) \quad \text{where } \mathbf{x}_1|\mathbf{x}_2 \text{ and } \mathbf{x}_2 \text{ are both Guassian(based on formula (201))} \tag{202}$$

$$\Rightarrow \int_{\mathbf{x}_1}\mathbf{p}(\mathbf{x}_1;\mathbf{x}_2)\mathrm{d}\mathbf{x}_1 = \int_{\mathbf{x}_1}\mathbf{p}(\mathbf{x}_1|\mathbf{x}_2)\mathbf{p}(\mathbf{x}_2)\mathrm{d}\mathbf{x}_1 = \mathbf{p}(\mathbf{x}_2)\underbrace{\int_{\mathbf{x}_1}\mathbf{p}(\mathbf{x}_1|\mathbf{x}_2)\mathrm{d}\mathbf{x}_1}_{1} = \mathbf{p}(\mathbf{x}_2) : \text{Marginal Guassian pdf} \tag{203}$$

$$\Rightarrow \mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2;\mathbf{P}_{22}) \tag{204}$$

C. Relationship between Information and Covariance domain of Marginal Guassian pdf

In this section we find out the marginal pdf in view of information matrix/vector.

Suppose: $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$ or equivalently $\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N}^{-1}(\boldsymbol{\nu}, \mathcal{I})$

$\Rightarrow \mathbf{x}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \mathbf{P}_{22})$, Find $\boldsymbol{\nu}' = \mathbf{f}(\boldsymbol{\nu}_1, \boldsymbol{\nu}_2, \mathcal{I}_{11}, \mathcal{I}_{12}, \mathcal{I}_{21}, \mathcal{I}_{22})$ and $\mathcal{I}' = \mathbf{g}(\mathcal{I}_{11}, \mathcal{I}_{12}, \mathcal{I}_{21}, \mathcal{I}_{22})$ such that $\mathbf{x}_2 \sim \mathcal{N}^{-1}(\boldsymbol{\nu}', \mathcal{I}')$

$$\mathbf{P} = \mathcal{I}^{-1} = \left(\begin{bmatrix} (\mathcal{I}_{11} - \mathcal{I}_{12}\mathcal{I}_{22}^{-1}\mathcal{I}_{21})^{-1} & -\mathcal{I}_{11}^{-1}\mathcal{I}_{12}(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \\ -(\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}\mathcal{I}_{12})^{-1}\mathcal{I}_{21}\mathcal{I}_{11}^{-1} & (\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \end{bmatrix} \right) \tag{205}$$

$$\Rightarrow \mathbf{P}_{22} = (\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12})^{-1} \tag{206}$$

$$\Rightarrow \mathcal{I}' = \mathbf{P}_{22}^{-1} = \mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12} \tag{207}$$

$$\boldsymbol{\mu}_2 = \mathbf{P}_{22}\boldsymbol{\nu}' \Rightarrow \boldsymbol{\nu}' = \mathbf{P}_{22}^{-1}\boldsymbol{\mu}_2 = \mathbf{I}'\boldsymbol{\mu}_2 \tag{208}$$

$$\text{where} \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} = \mathbf{P}\boldsymbol{\nu} = \mathcal{I}^{-1} \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{bmatrix} \Rightarrow \boldsymbol{\mu}_2 = (\mathcal{I}')^{-1}\boldsymbol{\nu}_2 - \mathcal{I}_{22}^{-1}\mathcal{I}_{21}(\mathcal{I}')^{-1}\boldsymbol{\nu}_1 \tag{209}$$

$$\Rightarrow \boldsymbol{\nu}' = \mathcal{I}' \left[(\mathcal{I}')^{-1}\boldsymbol{\nu}_2 - (\mathcal{I}')^{-1}\mathcal{I}_{21}\mathcal{I}_{11}^{-1}\boldsymbol{\nu}_1 \right] = \boldsymbol{\nu}_2 - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\boldsymbol{\nu}_1 \tag{210}$$

After that, we found

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \sim \mathcal{N}^{-1} \left[\begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{pmatrix}, \begin{pmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{pmatrix} \right] \Rightarrow \mathbf{x}_2|\mathbf{y} \sim \mathcal{N}^{-1} \left(\boldsymbol{\nu}_2 - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\boldsymbol{\nu}_1; (\mathcal{I}_{22} - \mathcal{I}_{21}\mathcal{I}_{11}^{-1}\mathcal{I}_{12}) \right) \tag{211}$$

D. *Schur complement om Hessian domain versus marginal pdf in the Bundle adjustment problem*

Problem Statement:

For this section, we will show that

1. Taking marginal pdf $\mathbf{p}(\mathbf{x}_2|\mathbf{y})$
2. Pulling contribution factor to cost function $\mathbb{C}(\mathbf{x}_2)$
3. Taking derivative $\frac{\partial \mathbf{C}(\mathbf{x}_2)}{\partial \mathbf{x}_2}$=0

$\Leftrightarrow$ Doing Schur complement in equation (212)

As we have seen previously :

$$\max \mathbf{p}(\mathbf{x}) \Leftrightarrow \min \| \tag{212}$$

$$\mathbf{x}|\mathbf{y} \text{ is Gaussian vector} \tag{213}$$

$$\Rightarrow \mathbf{p}(\mathbf{x}|\mathbf{y}) \propto -(\frac{1}{2}\mathbf{x}^T\mathcal{I}\mathbf{x} - \mathbf{x}^T\boldsymbol{\nu}) \tag{214}$$

$$\Rightarrow \max \mathbf{p}(\mathbf{x}|\mathbf{y}) \Leftrightarrow \min(\frac{1}{2}\mathbf{x}^T\mathcal{I}\mathbf{x} - \mathbf{x}^T\boldsymbol{\nu}) \tag{215}$$

Previously:

$$\max \mathbf{p}(\mathbf{x}|\mathbf{y}) \Leftrightarrow \min \frac{1}{2}\|\mathbf{y} - \mathbf{C}\mathbf{x}\|_R^2 \tag{216}$$

$$\Leftrightarrow \min(\frac{1}{2}\mathbf{x}^T\mathbf{C}^T\mathbf{R}^{-1}\mathbf{C}\mathbf{x} - \mathbf{x}^T\mathbf{C}^T\mathbf{R}^{-1}\mathbf{y}) \tag{217}$$

From (215)

$$\Rightarrow \mathcal{I} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{C} \tag{218}$$

$$\boldsymbol{\nu} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{y} \tag{219}$$

$\Rightarrow$ Hessian information matrix

$$\mathcal{I} = \mathbf{C}^T\mathbf{R}^{-1}\mathbf{C} = \begin{bmatrix} \mathbf{C}_1^T\mathbf{R}^-1\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{R}^-1\mathbf{C}_2 \\ \mathbf{C}_2^T\mathbf{R}^-1\mathbf{C}_1 & \mathbf{C}_1^T\mathbf{R}^-1\mathbf{C}_2 \end{bmatrix} = \begin{bmatrix} \mathcal{I}_{11} & \mathcal{I}_{12} \\ \mathcal{I}_{21} & \mathcal{I}_{22} \end{bmatrix} \tag{220}$$

Information vector

$$\boldsymbol{\nu} = \begin{bmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{C}_1^T\mathbf{R}^-1\mathbf{y} \\ \mathbf{C}_2^T\mathbf{R}^-1\mathbf{y} \end{bmatrix} \tag{221}$$

Follow part C to obtain marginal pdf :

$$\Rightarrow \mathbf{x}_2|\mathbf{y} \sim \mathcal{N}^{-1}(\underbrace{\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{y} - \mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{y}}_{\mathbf{C}_2^T\mathbf{R}(\mathcal{I}-\mathbf{C}_2(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1})\mathbf{y}}, \underbrace{\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_2 - (\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1)(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2)}_{\mathcal{I}'}) \tag{222}$$

$$\tag{223}$$

$\Rightarrow$ contribution factor of $\mathbf{p}(\mathbf{x}_2|\mathbf{z})$ to the cost function is:

$$\mathbb{C}(\mathbf{x}_2) = \mathbf{x}_2^T\mathcal{I}'\mathbf{x}_2 - 2\mathbf{x}_2 T\boldsymbol{\nu}' \tag{224}$$

$$\Leftrightarrow \mathcal{I}'\mathbf{x}_2 = \boldsymbol{\nu}' (\text{this is equation equivalent to } \frac{\partial \mathbf{C}(\mathbf{x}_2)}{\partial \mathbf{x}_2} = 0 \text{ in Hessian domain}) \tag{225}$$

$$\Leftrightarrow \left[\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_2 - (\mathbf{C}_2^T\mathbf{R}^{-1}\mathbf{C}_1)(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1^{-1})(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_2)\right]\mathbf{x}_2 = \mathbf{C}_2^T\mathbf{R}^{-1}(\mathcal{I} - \mathbf{C}_1(\mathbf{C}_1^T\mathbf{R}^{-1}\mathbf{C}_1)^{-1}\mathbf{C}_1^T\mathbf{R}^{-1})\mathbf{y} \tag{226}$$

which is exactly as equation (117)