Multiple View Geometry: Solution Sheet 10

Prof. Dr. Daniel Cremers,
Shenhan Qian, Simon Weber, Anna Ribic, and Tarun Yenamandra
Computer Vision Group, TU Munich
Wednesdays 16:15–18:15 at Hörsaal 2, "Interims I" (5620.01.102), and on RBG Live

Exercise: July 2nd, 2025

---

1. **Gauss-Newton Method** When optimizing a function $F(x) = \frac{1}{2}\|r(x)\|_2^2$ with residual $r(x) \in \mathbb{R}^n, and x \in \mathbb{R}^m$, the Gauss-Newton method approximates the residual using a Taylor expansion:
$$r(x_0 + \Delta x) \approx r(x_0) + J_r(x_0)\Delta x \tag{1}$$
The minimization problem thus is
$$\min_{\Delta x} \frac{1}{2}\|r_0 + J\Delta x\|_2^2 \tag{2}$$
with a slight abuse of notation $J := J_r(x_0)$ and $r_0 := r(x_0)$.

   (a) Compute the gradient of $\frac{1}{2}\|r_0 + J\Delta x\|_2^2$ w.r.t. Δx.
$$\begin{aligned}\frac{d}{d\epsilon}\frac{1}{2}\|r_0 + J(\Delta x + \epsilon v)\|_2^2\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}\frac{1}{2}\left\langle r_0 + J(\Delta x + \epsilon v), r_0 + J(\Delta x + \epsilon v)\right\rangle\Big|_{\epsilon=0}\\ &=< r_0, Jv > + < J\Delta x, Jv >\\ &= r_0^\top Jv + (\Delta x)^\top J^\top Jv\\ &=< \underbrace{J^\top r_0 + J^\top J\Delta x}_{\text{gradient}}, v >\end{aligned}$$

   (b) Solve the optimality condition for Δx.
$$\begin{aligned}J^\top r_0 + J^\top J\Delta x = 0\\ \Rightarrow \Delta x = -(J^\top J)^{-1}J^\top r_0\end{aligned}$$

   (c) What problems can occur when solving for Δx?
      - The matrix $J^\top J$ might be ill-conditioned.

2. **Levenberg-Marquardt Method**

   One way to motivate the Levenberg-Marquardt method is to tackle the previously discussed problem by adding the damping term as follows:
$$\left(J^\top J + \lambda D^T D\right)\Delta x = -J^\top r_0. \tag{3}$$
However, this can also be seen as a regularized version of the Gauss-Newton method.
$$\min_{\Delta x} \frac{1}{2}\|r_0 + J\Delta x\|_2^2 + \frac{\lambda}{2}\|D\Delta x\|_2^2. \tag{4}$$

(a) Compute the gradient of the new cost function w.r.t. Δx.

$$\begin{aligned}\frac{d}{d\epsilon}\frac{\lambda}{2}\|D\Delta x\|_2^2\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}\frac{\lambda}{2}\left\langle D(\Delta x+\epsilon v), D(\Delta x+\epsilon v)\right\rangle\Big|_{\epsilon=0} \\ &=< \underbrace{\lambda D^T D\Delta x}_{gradient}, v >\end{aligned}$$

And the gradient of the new cost function w.r.t. Δx is (due to linearity of the derivative):

$$J^\top r_0 + J^\top J\Delta x + \lambda D^T D\Delta x$$

(b) Solve the optimality condition for Δx.

$$\begin{aligned}J^\top r_0 + J^\top J\Delta x + \lambda D^T D\Delta x = 0 \\ \Rightarrow (J^\top J + \lambda D^T D)\Delta x = -J^\top r_0 \\ \Rightarrow \Delta x = -(J^\top J + \lambda D^T D)^{-1} J^\top r_0\end{aligned}$$

(c) What is the effect of λ on the solution? The effect of λ is to scale the damping term, i.e. how much we stay in the trust region.

- If λ is large, the damping term will dominate the solution, we stay in the trust region.
- If λ is small, the solution will be close to the Gauss-Newton solution.

3. **Levenberg-Marquardt for Bundle Adjustment**

Now, we apply the Levenberg-Marquardt method to the bundle adjustment problem. The variables are as follows:

- n_p: number poses
- n_l: number landmarks
- d_p: number of camera parameters
- $x_p \in \mathbb{R}^{n_p d_p}$: camera parameters
- $x_l \in \mathbb{R}^{n_l 3}$: landmark positions
- $x = \begin{bmatrix} x_p \\ x_l \end{bmatrix}$

We resuse the results from the previous problem with square diagonal matrix D

$$\min_{\Delta x} \frac{1}{2}\|r + J\Delta x\|_2^2 + \frac{\lambda}{2}\|D\Delta x\|_2^2. \tag{5}$$

which is the following optimality condition

$$\underbrace{(J^\top J + \lambda D^T D)}_{H}\Delta x = -J^\top r_0. \tag{6}$$

Now we split the Jacobian and damping into two parts $J = \begin{bmatrix} J_p & J_l \end{bmatrix}$ and $D = \begin{bmatrix} D_p & D_l \end{bmatrix}$ correponding to the camera parameters and the landmark positions.

(a) What is the dimension of H? What problems can occur when solving for Δx? What are the dimensions of J_p, J_ℓ, D_p, D_ℓ? Lets see what we can do... The matrix $J^\top J$ might be too large to invert. With I being the dimension of r, the dimensions of the respective matrices are:

- $H : \mathbb{R}^{(n_p d_p + n_l 3)\times(n_p d_p + n_l 3)}$
- J_p: $\mathbb{R}^{I \times n_p d_p}$
- J_l: $\mathbb{R}^{I \times n_l 3}$
- D_p: $\mathbb{R}^{n_p d_p + n_l 3 \times n_p d_p}$
- D_l: $\mathbb{R}^{n_p d_p + n_l 3 \times n_l 3}$

(b) Rewrite the optimality condition by rewriting the matrix H into the block matrix form, yielding the normal equation:

$$\begin{pmatrix} U & W \\ W^\top & V \end{pmatrix} \begin{pmatrix} \Delta x_p \\ \Delta x_l \end{pmatrix} = \begin{pmatrix} b_p \\ b_l \end{pmatrix}. \tag{7}$$

What are U, W, V, b_p, b_l and their dimensions?

$$U = J_p^\top J_p + \lambda D_p^\top D_p \in \mathbb{R}^{n_p d_p \times n_p d_p}, \quad W = J_p^\top J_l \in \mathbb{R}^{n_p d_p \times n_l 3}, \quad V = J_l^\top J_l + \lambda D_l^\top D_l \in \mathbb{R}^{n_l 3 \times n_l 3}$$

and

$$b_p = -J_p^\top r_0 \in \mathbb{R}^{n_p d_p}, \quad b_l = -J_l^\top r_0 \in \mathbb{R}^{n_l 3}$$

(c) The Schur complement is allowing us to first solve for Δx_p using the Schur complement S. Derive the Schur complement S and the vector $\tilde{b}$ for the reduced system: $S\Delta x_p = \tilde{b}$.

$$\begin{aligned} &\begin{bmatrix} U & W \\ W^\top & V \end{bmatrix} \begin{bmatrix} \Delta x_p \\ \Delta x_l \end{bmatrix} = \begin{bmatrix} b_p \\ b_l \end{bmatrix} \\ \Rightarrow &\begin{bmatrix} U & W \\ -WV^{-1}W^\top & -WV^{-1}V \end{bmatrix} \begin{bmatrix} \Delta x_p \\ \Delta x_l \end{bmatrix} = \begin{bmatrix} b_p \\ -WV^{-1}b_l \end{bmatrix} \\ \Rightarrow &\underbrace{(U - WV^{-1}W^\top)}_{:=S} \Delta x_p = \underbrace{b_p - WV^{-1}b_l}_{:=\tilde{b}} \end{aligned}$$

(d) What is the dimension of S? The dimension of S is $n_p d_p \times n_p d_p$.

4. **Power Bundle Adjustment** The goal of Power Bundle Adjustment is to solve the reduced system $S\Delta x_p = \tilde{b}$ efficiently.

(a) From the lecture, we know that computing the inverse of the Schur component can be approximated by a matrix power series. Specifically, we have:

$$\begin{aligned} S &= U(I - U^{-1}WV^{-1}W^\top) \\ \rightarrow S^{-1} &= (I - U^{-1}WV^{-1}W^\top)^{-1}U^{-1} \\ \rightarrow S^{-1} &\approx \sum_{i=0}^{m} (U^{-1}WV^{-1}W^\top)^i U^{-1}. \end{aligned} \tag{8}$$

To apply the matrix power series, we need to guarantee the spectral norm of the matrix is smaller than 1, i.e. show that all the eigenvalues μ of $U^{-1}WV^{-1}W^\top$ satisfy $0 \leq \mu < 1$. Hint: Consider the similar matrix $U^{-1/2}WV^{-1}WU^{-1/2}$ for $U^{-1}WV^{-1}W^\top$ and show $U^{-1/2}WV^{-1}WU^{-1/2}$ is positive semi-definite. Additionally, the similar matrix $U^{-1/2}SU^{-1/2}$ for $U^{-1}S$ and show it is positive definite.

Mathematical properties:

(*) Similar matrices: $A \sim P^{-1}AP$ for invertible P and A implies that A and $P^{-1}AP$ have the same eigenvalues. (Prove as homework)

(**) The Schur-complement S of a matrix H is positive definite if H is positive definite and S is positive semi-definite if H is positive semi-definite.

(°) Eigenvalues of pos. (sem.) matrices are bigger (or equal) to 0.

(°°) U diagonalizeable with pos EVs in D

$$U = PDP^{-1} = P\sqrt{D}\sqrt{D}P^{-1} = P\sqrt{D}P^{-1}P\sqrt{D}P^{-1} = U^{1/2}U^{1/2}$$

(°°°) U pos. def $\Rightarrow U^{-1}$ pos. def.

(°°°°) A, B pos. def $\Rightarrow ABA$ pos. def.

Note that U, V are symmetric pos. definite

- " $\mu \geq 0$": $A = U^{-1}WV^{-1}W^\top$ and $P = U^{-1/2}$.
  Therefore, we can construct the similar matrix to A by $P^{-1}AP = U^{-1/2}WV^{-1}W^\top U^{-1/2}$, which is symmetric. To this end, we examine the eigenvalues of $U^{-1/2}WV^{-1}WU^{-1/2}$ instead.
  - $U^{-1/2}WV^{-1}WU^{-1/2}$ positive semi-definite: Considering a non-zero vector $x \in \mathbb{R}^{n_p \times n_p}$, we can denote $y = WU^{-1/2}x$ and get:

    $$x^\top U^{-1/2}WV^{-1}WU^{-1/2}x = y^\top V^{-1}y.$$

    Because V is positive definite, it's inverse V^{-1} is positive definite.

    $$y^\top V^{-1}y > 0$$

    This means that the eigenvalues of $U^{-1/2}WV^{-1}WU^{-1/2}$ are larger or equal to 0.

  (eigenvalues of $U^{-1/2}WV^{-1}WU^{-1/2}$) $\geq 0 \Rightarrow$ (eigenvalues of A)$\geq 0 \Rightarrow \mu \geq 0$.
- " $\mu < 1$": Similarly, we can construct the similar matrix $U^{-1/2}SU^{-1/2}$ for the matrix $U^{-1}S$ to analyze the eigenvalues λ of $U^{-1}S$. To this end, we examine the eigenvalues of $U^{-1/2}SU^{-1/2}$ instead.
  - Because U and S (Schur complement of symm matrix) are positive definite, $U^{-1/2}SU^{-1/2}$ is positive definite. and the eigenvalues of $U^{-1/2}SU^{-1/2}$ are strictly bigger than 0.

  (eigenvalues of $U^{-1/2}SU^{-1/2}$) $> 0 \Rightarrow$ (eigenvalues of $U^{-1}S$)$> 0 \Rightarrow \lambda > 0$. Because

  $$A = I - U^{-1}S,$$

  it follows for the eigenvalues of A

  $$\mu = 1 - \lambda \Rightarrow \mu < 1.$$

  i.e. they are strictly smaller than 1.

5. **Dense RGB-D Tracking** In the previous bundle adjustment problem, we have seen how to optimize the camera parameters x_p and landmark positions x_l. Here, in the context of direct tracking, we optimize for the extrinsic camera parameters $x_p = [\xi_1, ..., \xi_{n_p}]$ using the photometric error as a residual and frame wise depth map h provided. With known camera poses, the 3D geometry can thus be densely be reconstructed. No need to optimize for landmark positions x_l. The residual is as follows:

$$E(x_p) = \sum_i \int_{\Omega_1} \| \underbrace{I_1(x) - I_i(\Pi g_{\xi_i}(hx))}_{r_x(\xi_i)} \|^2 dx \tag{9}$$

where I_1 and I_i are the intensity images, Π is the projection operator, g_{ξ_i} is the rigid transorm depending on the camera pose. The integral is over the image domain Ω_1 with x here being the homogeneous image coordinate and h its depth in the first frame.

(a) Using the results from previous problems, state the optimality condition for minimizing $\|r_x(\xi_i)\|_2^2$ using the Levenberg-Marquardt method.

$$(J^\top J + \lambda D^T D)\Delta\xi_i = -J^\top r_0$$

where J is the total derivative of the residual $r_x(\xi_i)$ w.r.t. the camera parameters ξ_i. It is a transposed gradient, if $I : \mathbb{R}^2 \to \mathbb{R}$, i.e. the image is in grayscale. D is the damping term.

(b) Compute the total derivative of the residual $r_x(\xi_i)$ w.r.t. the camera parameters ξ_i using the chain rule. You don't have to explicitly compute $\frac{d}{d\xi_i} g_{\xi_i}(hx)$.

$$\begin{aligned} J = \frac{d}{d\xi_i} r_x(\xi_i) &= \frac{d}{d\xi_i} I_1(x) - \frac{d}{d\xi_i} I_i(\Pi g_{\xi_i}(hx)) \\ &= -\nabla I_i \frac{d\Pi}{dg} \frac{d}{d\xi_i} g_{\xi_i}(hx) \end{aligned}$$