Appendix B

Distribution-to-Distribution NDT

In this appendix, we start from the approximation of the log-likelihood of pdf addressed in the Point-to-Distribution NDT [7]. Then, we paraphrase the derivations of the objective function, the gradient vector, and the Hessian matrix of the Distribution-to-Distribution NDT [6].

Approximation of Log-likelihood of Gaussian Probability Density Function [7]

Let $p(x) = \mathcal{N}(x \mid \boldsymbol{\mu}, \sigma^2)$ be the probability density function of a Gaussian distribution. Minimizing $p(x)$ is often replaced by minimizing $\log(p(x))$ due to the property of logarithms. However, $\log(p(x))$ could be unbounded below when $p(x)$ is close to zero. To prevent this issue from happening, Magnusson [7] suggests to add a base pdf value $0 < c_2 < 1$ and a scale factor $c_1 > 0$ to approximate $p(x)$, that is,

$$
\begin{aligned}
p(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) && \text{(B.1)}\\
\approx \overline{p}(x) &= c_1 \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) + c_2. && \text{(B.2)}
\end{aligned}
$$

Afterwards, the second approximation is performed to $\log(\overline{p}(x))$

$$
\begin{aligned}
\log(\overline{p}(x)) &= \log(c_1 \exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}) + c_2) && \text{(B.3)}\\
\approx \quad \widetilde{p}(x) &= d_1 \exp(-\frac{d_2}{2}\frac{(x-\mu)^2}{\sigma^2}) + d_3 && \text{(B.4)}
\end{aligned}
$$

where $d_1, d_2 > 0, d_3 < 0$ are computed by assuming the approximated value and the original value are identical at μ, $\mu \pm \sigma$, and $\pm\infty$, that is,

$$
\log(\overline{p}(\mu)) = \widetilde{p}(\mu),\ \log(\overline{p}(\mu \pm \sigma)) = \widetilde{p}(\mu \pm \sigma),\ \log(\overline{p}(\pm\infty)) = \widetilde{p}(\pm\infty). \tag{B.5}
$$

Therefore, we have

$$d_1 = \log(c_1 + c_2) - d_3 \tag{B.6}$$

$$d_2 = -2\log\left(\frac{\log(c_1 \exp(-\frac{1}{2}) + c_2) - d_3}{d_1}\right) \tag{B.7}$$

$$d_3 = \log(c_2). \tag{B.8}$$

In the implementation in [7], these parameters are chosen as empirical values. Following the same approximation, we can derive the following multivariate version:

$$\arg\max\ \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma) = \arg\max\ \log(\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu}, \Sigma)) \tag{B.9}$$

$$\approx \arg\max\ d_1 \exp(-\frac{d_2}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x} - \boldsymbol{\mu})). \tag{B.10}$$

Note that d_3 is omitted since it is a constant and does not influence the optimization.

Objective Function of Distribution-to-Distribution NDT [6]

Recall eq. (3.9), the L_2 distance between two NDT models is

$$\begin{aligned} D_{L_2}(\mathcal{N}_p, \mathcal{N}_q) &= \int (\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_p}, \Sigma_p) - \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_q}, \Sigma_q))^2 \, d\boldsymbol{x} \\ &= \int (\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_p}, \Sigma_p))^2 \, d\boldsymbol{x} + \int (\mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_q}, \Sigma_q))^2 \, d\boldsymbol{x} \\ &\quad - 2 \int \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_p}, \Sigma_p)\, \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_q}, \Sigma_q) \, d\boldsymbol{x} \end{aligned} \tag{B.11}$$

The first and second terms in eq. (B.11) are constants, and the third term can be simplified as,

$$-2 \int \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_p}, \Sigma_p)\, \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_q}, \Sigma_q) \, d\boldsymbol{x} \tag{B.12}$$

$$= -2 \int \mathcal{N}(\boldsymbol{0} \mid \boldsymbol{\mu_p} - \boldsymbol{\mu_q}, \Sigma_p + \Sigma_q)\, \mathcal{N}(\boldsymbol{x} \mid \boldsymbol{\mu_c}, \Sigma_c) \, d\boldsymbol{x} \tag{B.13}$$

$$= -2\, \mathcal{N}(\boldsymbol{0} \mid \boldsymbol{\mu_p} - \boldsymbol{\mu_q}, \Sigma_p + \Sigma_q) \tag{B.14}$$

where $\boldsymbol{\mu_c} = (\Sigma_p^{-1} + \Sigma_q^{-1})^{-1}(\Sigma_p^{-1}\boldsymbol{\mu_p} + \Sigma_q^{-1}\boldsymbol{\mu_q}),\ \Sigma_c = (\Sigma_p^{-1} + \Sigma_q^{-1})^{-1}$. Equation (B.13) follows eq. (371) in Section 8.1.8 of the Matrix Cookbook [21] and can be proved by expansion.

After combining all the correspondences, the optimized parameter $\mathbf{\Theta}$, and the approximation in eq. (B.10), the final optimization problem then becomes eq. (3.12).

$$\arg\min_{\mathbf{\Theta}} \prod_{k=1}^{n} D_{L_2}(T(\mathcal{N}_{pk}, \mathbf{\Theta}), \mathcal{N}_{qk}) \tag{B.15}$$

$$= \arg\min_{R,\boldsymbol{t}} \prod_{k=1}^{n} D_{L_2}(\mathcal{N}(R\boldsymbol{\mu_{pk}} + \boldsymbol{t}, R\Sigma_{pk}R^T), \mathcal{N}(\boldsymbol{\mu_{qk}}, \Sigma_{qk})) \tag{B.16}$$

$$= \arg\min_{R,\boldsymbol{t}} -\prod_{k=1}^{n} \mathcal{N}(\mathbf{0} \mid R\boldsymbol{\mu_{pk}} + \boldsymbol{t} - \boldsymbol{\mu_{qk}}, R\Sigma_{pk}R^T + \Sigma_{qk}) \tag{B.17}$$

$$= \arg\min_{R,\boldsymbol{t}} \sum_{k=1}^{n} -\log(\mathcal{N}(\mathbf{0} \mid R\boldsymbol{\mu_{pk}} + \boldsymbol{t} - \boldsymbol{\mu_{qk}}, R\Sigma_{pk}R^T + \Sigma_{qk})) \tag{B.18}$$

$$\approx \arg\min_{R,\boldsymbol{t}} \sum_{k=1}^{n} -d_1 \exp\left(-\frac{d_2}{2}(R\boldsymbol{\mu_{pk}} + \boldsymbol{t} - \boldsymbol{\mu_{qk}})^T (R\Sigma_{pk}R^T + \Sigma_{qk})^{-1} (R\boldsymbol{\mu_{pk}} + \boldsymbol{t} - \boldsymbol{\mu_{qk}})\right) \tag{B.19}$$

Gradient Vector and Hessian Matrix of Distribution-to-Distribution NDT [6]

Following the notations in [6], we define

$$\boldsymbol{\mu_{pq}} = R\boldsymbol{\mu_p} + \boldsymbol{t} - \boldsymbol{\mu_q} \qquad B = (R\Sigma_p R^T + \Sigma_q)^{-1} \tag{B.20}$$

$$\boldsymbol{j_a} = \frac{\partial \boldsymbol{\mu_{pq}}}{\partial \Theta_a} = \frac{\partial (R\boldsymbol{\mu_p} + \boldsymbol{t})}{\partial \Theta_a} \qquad Z_a = \frac{\partial B^{-1}}{\partial \Theta_a} = \frac{\partial (R\Sigma_p R^T)}{\partial \Theta_a} \tag{B.21}$$

$$\boldsymbol{H_{ab}} = \frac{\partial^2 \boldsymbol{\mu_{pq}}}{\partial \Theta_a \partial \Theta_b} = \frac{\partial^2 (R\boldsymbol{\mu_p} + \boldsymbol{t})}{\partial \Theta_a \partial \Theta_b} \qquad Z_{ab} = \frac{\partial^2 B^{-1}}{\partial \Theta_a \partial \Theta_b} = \frac{\partial^2 (R\Sigma_i R^T)}{\partial \Theta_a \partial \Theta_b} \tag{B.22}$$

The objective function of a single correspondence is

$$f_{d2d\text{-}ndt}(\mathbf{\Theta}, \mathcal{N}_p, \mathcal{N}_q) = -d_1 \exp(-\frac{d_2}{2}\boldsymbol{\mu_{pq}}^T B \boldsymbol{\mu_{pq}}) \tag{B.23}$$

The components of the gradient vector are

$$\nabla f(\mathbf{\Theta})_a = \frac{d_1 d_2}{2} \exp(-\frac{d_2}{2}\boldsymbol{\mu_{pq}}^T B \boldsymbol{\mu_{pq}}) \cdot q_a \tag{B.24}$$

$$q_a = 2\boldsymbol{\mu_{pq}}^T B \boldsymbol{j_a} - \boldsymbol{\mu_{pq}}^T B Z_a B \boldsymbol{\mu_{pq}} \tag{B.25}$$

Note that we apply the property $\frac{\partial B}{\partial \Theta_a} = -B\frac{\partial B^{-1}}{\partial \Theta_a}B = -BZ_aB$.

After that, the components of the Hessian matrix are

$$\begin{aligned}\nabla^2 f(\boldsymbol{\Theta})_{ab} = d_1 d_2 \exp(-\frac{d_2}{2}\boldsymbol{\mu_{pq}}^T B\boldsymbol{\mu_{pq}})(\boldsymbol{j_b}^T B\boldsymbol{j_a} - \boldsymbol{\mu_{pq}}^T BZ_bB\boldsymbol{j_a} + \boldsymbol{\mu_{pq}}^T B\boldsymbol{H_{ab}} \\ -\boldsymbol{\mu_{pq}}^T BZ_aB\boldsymbol{j_b} + \boldsymbol{\mu_{pq}}^T BZ_aBZ_bB\boldsymbol{\mu_{pq}} \\ -\frac{1}{2}\boldsymbol{\mu_{pq}}^T BZ_{ab}B\boldsymbol{\mu_{pq}} - \frac{d_2}{4}q_aq_b).\end{aligned} \tag{B.26}$$

where q_b follows eq. (B.25) but use the derivatives to Θ_b.

Analytical Expressions of $\boldsymbol{j_a}, \boldsymbol{H_{ab}}, Z_a, Z_{ab}$

To compute these expressions, we first compute $\frac{\partial R}{\partial \Theta_a}$ and $\frac{\partial^2 R}{\partial \Theta_a \partial \Theta_b}$. We have

$$R = R_x(\theta_x)R_y(\theta_y)R_z(\theta_z) \tag{B.27}$$

$$= \begin{bmatrix} c_yc_z & -s_zc_y & s_y \\ s_xs_yc_z + s_zc_x & -s_xs_ys_z + c_xc_z & -s_xc_y \\ s_xs_z - s_yc_xc_z & s_xc_z + s_ys_zc_x & c_xc_y \end{bmatrix} \tag{B.28}$$

where $s_x = \sin\theta_x,\ s_y = \sin\theta_y,\ s_z = \sin\theta_z,\ c_x = \cos\theta_x,\ c_y = \cos\theta_y,\ c_z = \cos\theta_z$.

The derivatives of R with respect to translation parameters t_x, t_y, t_z are zero matrices since R is only related to $\theta_x, \theta_y, \theta_z$, i.e.,

$$\frac{\partial R}{\partial t_x} = \frac{\partial R}{\partial t_y} = \frac{\partial R}{\partial t_z} = 0_{3\times 3} \tag{B.29}$$

where $0_{3\times 3}$ represents the 3x3 zero matrix.

For the same reason, we have

$$\frac{\partial^2 R}{\partial \Theta_a \partial \Theta_b} = 0_{3\times 3} \tag{B.30}$$

where a or $b \in \{t_x, t_y, t_z\}$.

The next page shows the derivatives of R with respect to $\theta_x, \theta_y, \theta_z$.

$$\frac{\partial R}{\partial \theta_x} = \begin{bmatrix} 0 & 0 & 0 \\ -s_x s_z + s_y c_x c_z & -s_x c_z - s_y s_z c_x & -c_x c_y \\ s_x s_y c_z + s_z c_x & -s_x s_y s_z + c_x c_z & -s_x c_y \end{bmatrix} \tag{B.31}$$

$$\frac{\partial R}{\partial \theta_y} = \begin{bmatrix} -s_y c_z & s_y s_z & c_y \\ s_x c_y c_z & -s_x s_z c_y & s_x s_y \\ -c_x c_y c_z & s_z c_x c_y & -s_y c_x \end{bmatrix} \tag{B.32}$$

$$\frac{\partial R}{\partial \theta_z} = \begin{bmatrix} -s_z c_y & -c_y c_z & 0 \\ -s_x s_y s_z + c_x c_z & -s_x s_y c_z - s_z c_x & 0 \\ s_x c_z + s_y s_z c_x & -s_x s_z + s_y c_x c_z & 0 \end{bmatrix} \tag{B.33}$$

$$\frac{\partial^2 R}{\partial \theta_x^2} = \begin{bmatrix} 0 & 0 & 0 \\ -s_x s_y c_z - s_z c_x & s_x s_y s_z - c_x c_z & s_x c_y \\ -s_x s_z + s_y c_x c_z & -s_x c_z - s_y s_z c_x & -c_x c_y \end{bmatrix} \tag{B.34}$$

$$\frac{\partial^2 R}{\partial \theta_y^2} = \begin{bmatrix} -c_y c_z & s_z c_y & -s_y \\ -s_x s_y c_z & s_x s_y s_z & s_x c_y \\ s_y c_x c_z & -s_y s_z c_x & -c_x c_y \end{bmatrix} \tag{B.35}$$

$$\frac{\partial^2 R}{\partial \theta_z^2} = \begin{bmatrix} -c_y c_z & s_z c_y & 0 \\ -s_x s_y c_z - s_z c_x & s_x s_y s_z - c_x c_z & 0 \\ -s_x s_z + s_y c_x c_z & -s_x c_z - s_y s_z c_x & 0 \end{bmatrix} \tag{B.36}$$

$$\frac{\partial^2 R}{\partial \theta_x \partial \theta_y} = \begin{bmatrix} 0 & 0 & 0 \\ c_x c_y c_z & -s_z c_x c_y & s_y c_x \\ s_x c_y c_z & -s_x s_z c_y & s_x s_y \end{bmatrix} \tag{B.37}$$

$$\frac{\partial^2 R}{\partial \theta_x \partial \theta_z} = \begin{bmatrix} 0 & 0 & 0 \\ -s_x c_z - s_y s_z c_x & s_x s_z - s_y c_x c_z & 0 \\ -s_x s_y s_z + c_x c_z & -s_x s_y c_z - s_z c_x & 0 \end{bmatrix} \tag{B.38}$$

$$\frac{\partial^2 R}{\partial \theta_y \partial \theta_z} = \begin{bmatrix} s_y s_z & s_y c_z & 0 \\ -s_x s_z c_y & -s_x c_y c_z & 0 \\ s_z c_x c_y & c_x c_y c_z & 0 \end{bmatrix} \tag{B.39}$$

Therefore, we can derive $\boldsymbol{j_a}, \boldsymbol{H_{ab}}, Z_a, Z_{ab}$ from eq. (B.31) to eq. (B.39).

$$\boldsymbol{j_{t_x}} = [1,\ 0,\ 0]^T \qquad \boldsymbol{j_{t_y}} = [0,\ 1,\ 0]^T \qquad \boldsymbol{j_{t_z}} = [0,\ 0,\ 1]^T \tag{B.40}$$

$$\boldsymbol{j_{\theta_x}} = \frac{\partial R}{\partial \theta_x}\boldsymbol{\mu_p} \qquad \boldsymbol{j_{\theta_y}} = \frac{\partial R}{\partial \theta_y}\boldsymbol{\mu_p} \qquad \boldsymbol{j_{\theta_z}} = \frac{\partial R}{\partial \theta_z}\boldsymbol{\mu_p} \tag{B.41}$$

$$\boldsymbol{H_{\theta_x\theta_x}} = \frac{\partial^2 R}{\partial \theta_x^2}\boldsymbol{\mu_p} \qquad \boldsymbol{H_{\theta_y\theta_y}} = \frac{\partial^2 R}{\partial \theta_y^2}\boldsymbol{\mu_p} \qquad \boldsymbol{H_{\theta_z\theta_z}} = \frac{\partial^2 R}{\partial \theta_z^2}\boldsymbol{\mu_p} \tag{B.42}$$

$$\boldsymbol{H_{\theta_x\theta_y}} = \frac{\partial^2 R}{\partial \theta_x \theta_y}\boldsymbol{\mu_p} \qquad \boldsymbol{H_{\theta_x\theta_z}} = \frac{\partial^2 R}{\partial \theta_x \theta_z}\boldsymbol{\mu_p} \qquad \boldsymbol{H_{\theta_y\theta_z}} = \frac{\partial^2 R}{\partial \theta_y \theta_z}\boldsymbol{\mu_p} \tag{B.43}$$

Note that $\boldsymbol{H_{ab}} = \mathbf{0}$ when a or $b \in \{t_x, t_y, t_z\}$.

$$Z_{t_x} = Z_{t_y} = Z_{t_z} = 0_{3\times 3} \tag{B.44}$$

$$Z_{\theta_x} = \frac{\partial R}{\partial \theta_x}\Sigma_p R^T + (\frac{\partial R}{\partial \theta_x}\Sigma_p R^T)^T \tag{B.45}$$

$$Z_{\theta_y} = \frac{\partial R}{\partial \theta_y}\Sigma_p R^T + (\frac{\partial R}{\partial \theta_y}\Sigma_p R^T)^T \tag{B.46}$$

$$Z_{\theta_z} = \frac{\partial R}{\partial \theta_z}\Sigma_p R^T + (\frac{\partial R}{\partial \theta_z}\Sigma_p R^T)^T \tag{B.47}$$

$$Z_{\theta_x\theta_x} = (\frac{\partial^2 R}{\partial \theta_x^2}\Sigma_p R^T + \frac{\partial R}{\partial \theta_x}\Sigma_p(\frac{\partial R}{\partial \theta_x})^T) + (\frac{\partial^2 R}{\partial \theta_x^2}\Sigma_p R^T + \frac{\partial R}{\partial \theta_x}\Sigma_p(\frac{\partial R}{\partial \theta_x})^T)^T \tag{B.48}$$

$$Z_{\theta_y\theta_y} = (\frac{\partial^2 R}{\partial \theta_y^2}\Sigma_p R^T + \frac{\partial R}{\partial \theta_y}\Sigma_p(\frac{\partial R}{\partial \theta_y})^T) + (\frac{\partial^2 R}{\partial \theta_y^2}\Sigma_p R^T + \frac{\partial R}{\partial \theta_y}\Sigma_p(\frac{\partial R}{\partial \theta_y})^T)^T \tag{B.49}$$

$$Z_{\theta_z\theta_z} = (\frac{\partial^2 R}{\partial \theta_z^2}\Sigma_p R^T + \frac{\partial R}{\partial \theta_z}\Sigma_p(\frac{\partial R}{\partial \theta_z})^T) + (\frac{\partial^2 R}{\partial \theta_z^2}\Sigma_p R^T + \frac{\partial R}{\partial \theta_z}\Sigma_p(\frac{\partial R}{\partial \theta_z})^T)^T \tag{B.50}$$

$$Z_{\theta_x\theta_y} = (\frac{\partial^2 R}{\partial \theta_x \partial \theta_y}\Sigma_p R^T + \frac{\partial R}{\partial \theta_x}\Sigma_p(\frac{\partial R}{\partial \theta_y})^T) + (\frac{\partial^2 R}{\partial \theta_x \partial \theta_y}\Sigma_p R^T + \frac{\partial R}{\partial \theta_x}\Sigma_p(\frac{\partial R}{\partial \theta_y})^T)^T \tag{B.51}$$

$$Z_{\theta_x\theta_z} = (\frac{\partial^2 R}{\partial \theta_x \partial \theta_z}\Sigma_p R^T + \frac{\partial R}{\partial \theta_x}\Sigma_p(\frac{\partial R}{\partial \theta_z})^T) + (\frac{\partial^2 R}{\partial \theta_x \partial \theta_z}\Sigma_p R^T + \frac{\partial R}{\partial \theta_x}\Sigma_p(\frac{\partial R}{\partial \theta_z})^T)^T \tag{B.52}$$

$$Z_{\theta_y\theta_z} = (\frac{\partial^2 R}{\partial \theta_y \partial \theta_z}\Sigma_p R^T + \frac{\partial R}{\partial \theta_y}\Sigma_p(\frac{\partial R}{\partial \theta_z})^T) + (\frac{\partial^2 R}{\partial \theta_y \partial \theta_z}\Sigma_p R^T + \frac{\partial R}{\partial \theta_y}\Sigma_p(\frac{\partial R}{\partial \theta_z})^T)^T \tag{B.53}$$

Note that $Z_{ab} = 0_{3\times 3}$ when a or $b \in \{t_x, t_y, t_z\}$.

Finally, the values of eq. (B.24) and eq. (B.26) can be computed.