the best alignment of the point-clouds. Referring back to Section 6.3.2, we may actually be interested in $\hat{\mathbf{T}}_{iv_k}$, which can be recovered using

$$\hat{\mathbf{T}}_{iv_k} = \hat{\mathbf{T}}_{v_k i}^{-1} = \begin{bmatrix} \hat{\mathbf{C}}_{iv_k} & \hat{\mathbf{r}}_i^{v_k i} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{8.17}$$

Both forms of the transformation matrix are useful, depending on how the solution will be used.

8.1.3 Rotation Matrix Solution

The rotation-matrix case was originally studied outside of robotics by Green (1952) and Wahba (1965) and later within robotics by Horn (1987a) and Arun et al. (1987) and later by Umeyama (1991) considering the $\det \mathbf{C} = 1$ constraint. We follow the approach of de Ruiter and Forbes (2013), which captures all of the cases in which $\mathbf{C}$ can be determined uniquely. We also identify how many global and local solutions can exist for $\mathbf{C}$ when there is not a single global solution.

As in the previous section, we will use some simplified notation to avoid repeating sub- and super-scripts:

$$\mathbf{y}_j = \mathbf{r}_{v_k}^{p_j v_k}, \quad \mathbf{p}_j = \mathbf{r}_i^{p_j i}, \quad \mathbf{r} = \mathbf{r}_i^{v_k i}, \quad \mathbf{C} = \mathbf{C}_{v_k i}. \tag{8.18}$$

Also, we define

$$\mathbf{y} = \frac{1}{w}\sum_{j=1}^{M} w_j \mathbf{y}_j, \quad \mathbf{p} = \frac{1}{w}\sum_{j=1}^{M} w_j \mathbf{p}_j, \quad w = \sum_{j=1}^{M} w_j, \tag{8.19}$$

where the w_j are scalar weights for each point. Note that, as compared to the last section, some of the symbols are now 3×1 rather than 4×1.

We define an error term for each point:

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{C}(\mathbf{p}_j - \mathbf{r}). \tag{8.20}$$

Our estimation problem is then to globally minimize the cost function,

$$J(\mathbf{C}, \mathbf{r}) = \frac{1}{2}\sum_{j=1}^{M} w_j \mathbf{e}_j^T \mathbf{e}_j = \frac{1}{2}\sum_{j=1}^{M} w_j \left(\mathbf{y}_j - \mathbf{C}(\mathbf{p}_j - \mathbf{r})\right)^T \left(\mathbf{y}_j - \mathbf{C}(\mathbf{p}_j - \mathbf{r})\right), \tag{8.21}$$

subject to $\mathbf{C} \in SO(3)$ (i.e., $\mathbf{C}\mathbf{C}^T = \mathbf{1}$ and $\det \mathbf{C} = 1$).

Before carrying out the optimization, we will make a change of variables for the translation parameter. Define

$$\mathbf{d} = \mathbf{r} + \mathbf{C}^T \mathbf{y} - \mathbf{p}, \tag{8.22}$$

which is easy to isolate for $\mathbf{r}$ if all the other quantities are known. In this case, we can rewrite our cost function as

$$J(\mathbf{C}, \mathbf{d}) = \underbrace{\frac{1}{2} \sum_{j=1}^{M} w_j \left((\mathbf{y}_j - \mathbf{y}) - \mathbf{C}(\mathbf{p}_j - \mathbf{p}) \right)^T \left((\mathbf{y}_j - \mathbf{y}) - \mathbf{C}(\mathbf{p}_j - \mathbf{p}) \right)}_{\text{depends only on } \mathbf{C}} + \underbrace{\frac{1}{2} \mathbf{d}^T \mathbf{d}}_{\text{depends only on } \mathbf{d}}, \quad (8.23)$$

which is the sum of two semi-positive-definite terms, the first depending only on $\mathbf{C}$ and the second only on $\mathbf{d}$. We can minimize the second trivially by taking $\mathbf{d} = \mathbf{0}$, which in turn implies that

$$\mathbf{r} = \mathbf{p} - \mathbf{C}^T \mathbf{y}. \quad (8.24)$$

As in the quaternion case, this is simply the difference of the centroids of the two point-clouds, expressed in the stationary frame.

What is left is to minimize the first term with respect to $\mathbf{C}$. We note that if we multiply out each smaller term within the first large term, only one part actually depends on $\mathbf{C}$:

$$\begin{aligned} &\left((\mathbf{y}_j - \mathbf{y}) - \mathbf{C}(\mathbf{p}_j - \mathbf{p}) \right)^T \left((\mathbf{y}_j - \mathbf{y}) - \mathbf{C}(\mathbf{p}_j - \mathbf{p}) \right) \\ &= \underbrace{(\mathbf{y}_j - \mathbf{y})^T (\mathbf{y}_j - \mathbf{y})}_{\text{independent of } \mathbf{C}} - 2 \underbrace{\left((\mathbf{y}_j - \mathbf{y})^T \mathbf{C} (\mathbf{p}_j - \mathbf{p}) \right)}_{\operatorname{tr}(\mathbf{C}(\mathbf{p}_j - \mathbf{p})(\mathbf{y}_j - \mathbf{y})^T)} + \underbrace{(\mathbf{p}_j - \mathbf{p})^T (\mathbf{p}_j - \mathbf{p})}_{\text{independent of } \mathbf{C}}. \end{aligned} \quad (8.25)$$

Summing this middle term over all the (weighted) points, we have

$$\begin{aligned} \frac{1}{w} \sum_{j=1}^{M} w_j \left((\mathbf{y}_j - \mathbf{y})^T \mathbf{C} (\mathbf{p}_j - \mathbf{p}) \right) &= \frac{1}{w} \sum_{j=1}^{M} w_j \operatorname{tr} \left(\mathbf{C} (\mathbf{p}_j - \mathbf{p})(\mathbf{y}_j - \mathbf{y})^T \right) \\ &= \operatorname{tr} \left(\mathbf{C} \frac{1}{w} \sum_{j=1}^{M} w_j (\mathbf{p}_j - \mathbf{p})(\mathbf{y}_j - \mathbf{y})^T \right) = \operatorname{tr} \left(\mathbf{C} \mathbf{W}^T \right), \end{aligned} \quad (8.26)$$

where

$$\mathbf{W} = \frac{1}{w} \sum_{j=1}^{M} w_j (\mathbf{y}_j - \mathbf{y})(\mathbf{p}_j - \mathbf{p})^T. \quad (8.27)$$

This $\mathbf{W}$ matrix plays a similar role to the one in the quaternion section, by capturing the spread of the points (similar to an inertia matrix in dynamics), but it is not exactly the same. Therefore, we can define a new cost function that we seek to minimize with respect to $\mathbf{C}$ as

$$J(\mathbf{C}, \mathbf{\Lambda}, \gamma) = -\operatorname{tr}(\mathbf{C}\mathbf{W}^T) + \underbrace{\operatorname{tr}\left(\mathbf{\Lambda} (\mathbf{C}\mathbf{C}^T - \mathbf{1}) \right) + \gamma (\det \mathbf{C} - 1)}_{\text{Lagrange multiplier terms}}, \quad (8.28)$$

where $\mathbf{\Lambda}$ and γ are Lagrange multipliers associated with the two terms on the right; these are used to ensure that the resulting $\mathbf{C} \in SO(3)$. Note that when $\mathbf{C}\mathbf{C}^T = \mathbf{1}$ and $\det \mathbf{C} = 1$, these terms have no effect on the resulting cost. It is also worth noting that $\mathbf{\Lambda}$ is symmetric since we only need to enforce six orthogonality constraints. This new cost function will be minimized by the same $\mathbf{C}$ as our original one.

Taking the derivative of $J(\mathbf{C}, \mathbf{\Lambda}, \gamma)$ with respect to $\mathbf{C}$, $\mathbf{\Lambda}$, and γ, we have[5]

$$\frac{\partial J}{\partial \mathbf{C}} = -\mathbf{W} + 2\mathbf{\Lambda}\mathbf{C} + \gamma \underbrace{\det \mathbf{C}}_{1} \underbrace{\mathbf{C}^{-T}}_{\mathbf{C}} = -\mathbf{W} + \mathbf{L}\mathbf{C}, \tag{8.29a}$$

$$\frac{\partial J}{\partial \mathbf{\Lambda}} = \mathbf{C}\mathbf{C}^T - \mathbf{1}, \tag{8.29b}$$

$$\frac{\partial J}{\partial \gamma} = \det \mathbf{C} - 1, \tag{8.29c}$$

where we have lumped together the Lagrange multipliers as $\mathbf{L} = 2\mathbf{\Lambda} + \gamma\mathbf{1}$. Setting the first equation to zero, we find that

$$\mathbf{L}\mathbf{C} = \mathbf{W}. \tag{8.30}$$

At this point, our explanation can proceed in a simplified or detailed manner, depending on the level of fidelity we want to capture.

Before moving forward, we show that it is possible to arrive at (8.30) using our Lie group tools without the use of Lagrange multipliers. We consider a perturbation of the rotation matrix of the form

$$\mathbf{C}' = \exp\left(\boldsymbol{\phi}^\wedge\right)\mathbf{C}, \tag{8.31}$$

and then take the derivative of the objective function with respect to $\boldsymbol{\phi}$ and set this to zero for a critical point. For the derivative with respect to the ith element of $\boldsymbol{\phi}$ we have

$$\begin{aligned}
\frac{\partial J}{\partial \phi_i} &= \lim_{h\to 0} \frac{J(\mathbf{C}') - J(\mathbf{C})}{h} \\
&= \lim_{h\to 0} \frac{-\mathrm{tr}(\mathbf{C}'\mathbf{W}^T) + \mathrm{tr}(\mathbf{C}\mathbf{W}^T)}{h} \\
&= \lim_{h\to 0} \frac{-\mathrm{tr}(\exp(h\mathbf{1}_i^\wedge)\mathbf{C}\mathbf{W}^T) + \mathrm{tr}(\mathbf{C}\mathbf{W}^T)}{h} \\
&\approx \lim_{h\to 0} \frac{-\mathrm{tr}((\mathbf{1} + h\mathbf{1}_i^\wedge)\mathbf{C}\mathbf{W}^T) + \mathrm{tr}(\mathbf{C}\mathbf{W}^T)}{h} \\
&= \lim_{h\to 0} \frac{-\mathrm{tr}(h\mathbf{1}_i^\wedge\mathbf{C}\mathbf{W}^T)}{h} \\
&= -\mathrm{tr}\left(\mathbf{1}_i^\wedge\mathbf{C}\mathbf{W}^T\right).
\end{aligned} \tag{8.32}$$

[5] We require these useful facts to take the derivatives:
$\frac{\partial}{\partial \mathbf{A}} \det \mathbf{A} = \det(\mathbf{A})\,\mathbf{A}^{-T}$,
$\frac{\partial}{\partial \mathbf{A}} \mathrm{tr}(\mathbf{A}\mathbf{B}^T) = \mathbf{B}$,
$\frac{\partial}{\partial \mathbf{A}} \mathrm{tr}(\mathbf{B}\mathbf{A}\mathbf{A}^T) = (\mathbf{B} + \mathbf{B}^T)\mathbf{A}$.

Setting this to zero, we require

$$(\forall i) \ \mathrm{tr}\big(\mathbf{1}_i^\wedge \underbrace{\mathbf{C}\mathbf{W}^T}_{\mathbf{L}}\big) = 0. \tag{8.33}$$

Owing to the skew-symmetric nature of the $\wedge$ operator, this implies that $\mathbf{L} = \mathbf{C}\mathbf{W}^T$ is a symmetric matrix for a critical point. Taking the transpose and right-multiplying by $\mathbf{C}$, we come back to (8.30). We now continue with the main derivation.

Simplified Explanation

If we somehow knew that $\det \mathbf{W} > 0$, then we could proceed as follows. First, we postmultiply (8.30) by itself transposed to find

$$\mathbf{L}\underbrace{\mathbf{C}\mathbf{C}^T}_{\mathbf{1}}\mathbf{L}^T = \mathbf{W}\mathbf{W}^T. \tag{8.34}$$

Since $\mathbf{L}$ is symmetric, we have that

$$\mathbf{L} = \left(\mathbf{W}\mathbf{W}^T\right)^{\frac{1}{2}}, \tag{8.35}$$

which we see involves a matrix square-root. Substituting this back into (8.30), the optimal rotation is

$$\mathbf{C} = \left(\mathbf{W}\mathbf{W}^T\right)^{-\frac{1}{2}}\mathbf{W}. \tag{8.36}$$

This has the same form as the projection onto $SO(3)$ discussed in Section 7.2.1.

Unfortunately, this approach does not tell the entire story since it relies on assuming something about $\mathbf{W}$, and therefore does not capture all of the subtleties of the problem. With lots of non-coplanar points, this method will typically work well. However, there are some difficult cases for which we need a more detailed analysis. A common situation in which this problem occurs is when carrying out alignments using just three pairs of noisy points in the RANSAC algorithm discussed earlier. The next section provides a more thorough analysis of the solution that handles the difficult cases.

Detailed Explanation

The detailed explanation begins by first carrying out a singular-value decomposition (SVD)[6] on the (square, real) matrix, $\mathbf{W}$, so that

$$\mathbf{W} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \tag{8.37}$$

[6] The *singular-value decomposition* of a real $M \times N$ matrix, $\mathbf{A}$, is a factorization of the form $\mathbf{A} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ where $\mathbf{U}$ is an $M \times M$ real, orthogonal matrix (i.e., $\mathbf{U}^T\mathbf{U} = \mathbf{1}$), $\mathbf{D}$ is an $M \times N$ matrix with real entries $d_i \geq 0$ on the main diagonal (all other entries zero), and $\mathbf{V}$ is an $N \times N$ real, orthogonal matrix (i.e., $\mathbf{V}^T\mathbf{V} = \mathbf{1}$). The d_i are called the *singular values* and are typically ordered from largest to smallest along the diagonal of $\mathbf{D}$. Note that the SVD is not unique.

where $\mathbf{U}$ and $\mathbf{V}$ are square, orthogonal matrices and $\mathbf{D} = \mathrm{diag}(d_1, d_2, d_3)$ is a diagonal matrix of singular values, $d_1 \geq d_2 \geq d_3 \geq 0$.

Returning to (8.30), we can substitute in the SVD of $\mathbf{W}$ so that

$$\mathbf{L}^2 = \mathbf{L}\mathbf{L}^T = \mathbf{L}\mathbf{C}\mathbf{C}^T\mathbf{L}^T = \mathbf{W}\mathbf{W}^T = \mathbf{U}\mathbf{D}\underbrace{\mathbf{V}^T\mathbf{V}}_{\mathbf{1}}\mathbf{D}^T\mathbf{U}^T = \mathbf{U}\mathbf{D}^2\mathbf{U}^T. \tag{8.38}$$

Taking the matrix square-root, we can write that

$$\mathbf{L} = \mathbf{U}\mathbf{M}\mathbf{U}^T, \tag{8.39}$$

where $\mathbf{M}$ is the symmetric, matrix square root of $\mathbf{D}^2$. In other words,

$$\mathbf{M}^2 = \mathbf{D}^2. \tag{8.40}$$

It can be shown (de Ruiter and Forbes, 2013) that every real, symmetric $\mathbf{M}$ satisfying this condition can be written in the form

$$\mathbf{M} = \mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T, \tag{8.41}$$

where $\mathbf{S} = \mathrm{diag}(s_1, s_2, s_3)$ with $s_i = \pm 1$ and $\mathbf{Y}$ an orthogonal matrix (i.e., $\mathbf{Y}^T\mathbf{Y} = \mathbf{Y}\mathbf{Y}^T = \mathbf{1}$). An obvious example of this is $\mathbf{Y} = \mathbf{1}$ with $s_i = \pm 1$ and any values for d_i; a less obvious example that is a possibility when $d_1 = d_2$ is

$$\begin{aligned}\mathbf{M} &= \begin{bmatrix} d_1\cos\theta & d_1\sin\theta & 0 \\ d_1\sin\theta & -d_1\cos\theta & 0 \\ 0 & 0 & d_3 \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} & 0 \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{Y}} \underbrace{\begin{bmatrix} d_1 & 0 & 0 \\ 0 & -d_1 & 0 \\ 0 & 0 & d_3 \end{bmatrix}}_{\mathbf{DS}} \underbrace{\begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} & 0 \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}^T}_{\mathbf{Y}^T},\end{aligned} \tag{8.42}$$

for any value of the free parameter, θ. This illustrates an important point, that the structure of $\mathbf{Y}$ can become more complex in correspondence with repeated singular values (i.e., we cannot just pick any $\mathbf{Y}$). Related to this, we always have that

$$\mathbf{D} = \mathbf{Y}\mathbf{D}\mathbf{Y}^T, \tag{8.43}$$

due to the relationship between the block structure of $\mathbf{Y}$ and the multiplicity of the singular values in $\mathbf{D}$.

Now, we can manipulate the objective function that we want to minimize as follows:

$$\begin{aligned} J &= -\mathrm{tr}(\mathbf{C}\mathbf{W}^T) = -\mathrm{tr}(\mathbf{W}\mathbf{C}^T) = -\mathrm{tr}(\mathbf{L}) = -\mathrm{tr}(\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T\mathbf{U}^T) \\ &= -\mathrm{tr}(\underbrace{\mathbf{Y}^T\mathbf{U}^T\mathbf{U}\mathbf{Y}}_{\mathbf{1}}\mathbf{D}\mathbf{S}) = -\mathrm{tr}(\mathbf{D}\mathbf{S}) = -(d_1 s_1 + d_2 s_2 + d_3 s_3). \end{aligned} \tag{8.44}$$

There are now several cases to consider.

Case (i): $\det \mathbf{W} \neq 0$

Here we have that all of the singular values are positive. From (8.30) and (8.39) we have that

$$\begin{aligned} \det \mathbf{W} &= \det \mathbf{L} \underbrace{\det \mathbf{C}}_{1} = \det \mathbf{L} = \det(\mathbf{UYDSY}^T\mathbf{U}^T) \\ &= \underbrace{\det(\mathbf{Y}^T\mathbf{U}^T\mathbf{UY})}_{1} \det \mathbf{D} \det \mathbf{S} = \underbrace{\det \mathbf{D}}_{>0} \det \mathbf{S}. \end{aligned} \tag{8.45}$$

Since the singular values are positive, we have that $\det \mathbf{D} > 0$. Or in other words, the signs of the determinants of $\mathbf{S}$ and $\mathbf{W}$ must be the same, which implies that

$$\begin{aligned} \det \mathbf{S} &= \operatorname{sgn}(\det \mathbf{S}) = \operatorname{sgn}(\det \mathbf{W}) = \operatorname{sgn}\big(\det(\mathbf{UDV}^T)\big) \\ &= \operatorname{sgn}\big(\underbrace{\det \mathbf{U}}_{\pm 1} \underbrace{\det \mathbf{D}}_{>0} \underbrace{\det \mathbf{V}}_{\pm 1}\big) = \det \mathbf{U} \det \mathbf{V} = \pm 1. \end{aligned} \tag{8.46}$$

Note that we have $\det \mathbf{U} = \pm 1$ since $(\det \mathbf{U})^2 = \det(\mathbf{U}^T\mathbf{U}) = \det \mathbf{1} = 1$ and the same for $\mathbf{V}$. There are now four subcases to consider:

Subcase (i-a): $\det \mathbf{W} > 0$

Since $\det \mathbf{W} > 0$ by assumption, we must also have $\det \mathbf{S} = 1$ and therefore to uniquely minimize J in (8.44) we must pick $s_1 = s_2 = s_3 = 1$ since all of the d_i are positive and therefore we must have $\mathbf{Y}$ diagonal. Thus, from (8.30) we have

$$\begin{aligned} \mathbf{C} = \mathbf{L}^{-1}\mathbf{W} &= \left(\mathbf{UYDSY}^T\mathbf{U}^T\right)^{-1} \mathbf{UDV}^T \\ &= \mathbf{UY} \underbrace{\mathbf{S}^{-1}}_{\mathbf{S}} \mathbf{D}^{-1}\mathbf{Y}^T \underbrace{\mathbf{U}^T\mathbf{U}}_{\mathbf{1}} \mathbf{DV}^T = \mathbf{UYSD}^{-1} \underbrace{\mathbf{Y}^T\mathbf{D}}_{\mathbf{DY}^T} \mathbf{V}^T \\ &= \mathbf{UYSY}^T\mathbf{V}^T = \mathbf{USV}^T, \end{aligned} \tag{8.47}$$

with $\mathbf{S} = \operatorname{diag}(1,1,1) = \mathbf{1}$, which is equivalent to the solution provided in our 'simplified explanation' in the last section.

Subcase (i-b): $\det \mathbf{W} < 0$, $d_1 \geq d_2 > d_3 > 0$

Since $\det \mathbf{W} < 0$ by assumption, we have $\det \mathbf{S} = -1$, which means exactly one of the s_i must be negative. In this case, we can uniquely minimize J in (8.44) since the minimum singular value, d_3, is distinct, whereupon we must pick $s_1 = s_2 = 1$ and $s_3 = -1$ for the minimum. Since $s_1 = s_2 = 1$, we must have $\mathbf{Y}$ diagonal and can therefore from (8.30) we have that

$$\mathbf{C} = \mathbf{USV}^T, \tag{8.48}$$

with $\mathbf{S} = \operatorname{diag}(1, 1, -1)$.

Subcase (i-c): $\det \mathbf{W} < 0$, $d_1 > d_2 = d_3 > 0$

As in the last subcase, we have $\det \mathbf{S} = -1$, which means exactly one of the s_i must be negative. Looking to (8.44), since $d_2 = d_3$ we can pick either $s_2 = -1$ or $s_3 = -1$ and end up with the same value for J. With these values for the s_i we can pick any of the following for $\mathbf{Y}$:

$$\mathbf{Y} = \operatorname{diag}(\pm 1, \pm 1, \pm 1), \quad \mathbf{Y} = \begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & \pm \cos \frac{\theta}{2} & \mp \sin \frac{\theta}{2} \\ 0 & \pm \sin \frac{\theta}{2} & \pm \cos \frac{\theta}{2} \end{bmatrix}, \tag{8.49}$$

where θ is a free parameter. We can plug any of these $\mathbf{Y}$ in to find minimizing solutions for $\mathbf{C}$ using (8.30):

$$\mathbf{C} = \mathbf{U}\mathbf{Y}\mathbf{S}\mathbf{Y}^T\mathbf{V}^T, \tag{8.50}$$

with $\mathbf{S} = \operatorname{diag}(1, 1, -1)$ or $\mathbf{S} = \operatorname{diag}(1, -1, 1)$. Since θ can be anything, this means there are an infinite number of solutions that minimize the objective function.

Subcase (i-d): $\det \mathbf{W} < 0$, $d_1 = d_2 = d_3 > 0$

As in the last subcase, we have $\det \mathbf{S} = -1$, which means exactly one of the s_i must be negative. Looking to (8.44), since $d_1 = d_2 = d_3$ we can pick $s_1 = -1$ or $s_2 = -1$ or $s_3 = -1$ and end up with the same value for J, implying there an infinite number of minimizing solutions.

Case (ii): $\det \mathbf{W} = 0$

This time there are three subcases to consider depending on how many singular values are zero.

Subcase (ii-a): $\operatorname{rank} \mathbf{W} = 2$

In this case, we have $d_1 \geq d_2 > d_3 = 0$. Looking back to (8.44) we see that we can uniquely minimize J by picking $s_1 = s_2 = 1$ and since $d_3 = 0$, the value of s_3 does not affect J and thus it is a free parameter. Again looking to (8.30) we have

$$(\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T\mathbf{U}^T)\mathbf{C} = \mathbf{U}\mathbf{D}\mathbf{V}^T. \tag{8.51}$$

Multiplying by $\mathbf{U}^T$ from the left and $\mathbf{V}$ from the right, we have

$$\mathbf{D}\underbrace{\mathbf{U}^T\mathbf{C}\mathbf{V}}_{\mathbf{Q}} = \mathbf{D}, \tag{8.52}$$

since $\mathbf{DS} = \mathbf{D}$ due to $d_3 = 0$ and then $\mathbf{Y}\mathbf{D}\mathbf{Y}^T = \mathbf{D}$ from (8.43). The matrix, $\mathbf{Q}$, above will be orthogonal since $\mathbf{U}$, $\mathbf{C}$, and $\mathbf{V}$ are all orthogonal. Since $\mathbf{DQ} = \mathbf{D}$, $\mathbf{D} = \operatorname{diag}(d_1, d_2, 0)$, and $\mathbf{Q}\mathbf{Q}^T = \mathbf{1}$, we know that $\mathbf{Q} = \operatorname{diag}(1, 1, q_3)$ with $q_3 = \pm 1$. We also have that

$$q_3 = \det \mathbf{Q} = \det \mathbf{U} \underbrace{\det \mathbf{C}}_{1} \det \mathbf{V} = \det \mathbf{U} \det \mathbf{V} = \pm 1, \tag{8.53}$$

and therefore rearranging (and renaming $\mathbf{Q}$ as $\mathbf{S}$), we have

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{8.54}$$

with $\mathbf{S} = \text{diag}(1, 1, \det \mathbf{U} \det \mathbf{V})$.

Subcase (ii-b): $\text{rank}\, \mathbf{W} = 1$

In this case, we have $d_1 > d_2 = d_3 = 0$. We let $s_1 = 1$ to minimize J and now s_2 and s_3 do not affect J and are free parameters. Similarly to the last subcase, we end up with an equation of the form

$$\mathbf{D}\mathbf{Q} = \mathbf{D}, \tag{8.55}$$

which, along with $\mathbf{D} = \text{diag}(d_1, 0, 0)$ and $\mathbf{Q}\mathbf{Q}^T = \mathbf{1}$, implies that $\mathbf{Q}$ will have one of the following forms:

$$\underbrace{\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix}}_{\det \mathbf{Q} = 1} \quad \text{or} \quad \underbrace{\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & \sin\theta & -\cos\theta \end{bmatrix}}_{\det \mathbf{Q} = -1}, \tag{8.56}$$

with $\theta \in \mathbb{R}$ a free parameter. This means there are infinitely many minimizing solutions. Since

$$\det \mathbf{Q} = \det \mathbf{U} \underbrace{\det \mathbf{C}}_{1} \det \mathbf{V} = \det \mathbf{U} \det \mathbf{V} = \pm 1, \tag{8.57}$$

we have (renaming $\mathbf{Q}$ as $\mathbf{S}$) that

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{8.58}$$

with

$$\mathbf{S} = \begin{cases} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} & \text{if } \det \mathbf{U} \det \mathbf{V} = 1 \\ \\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & \sin\theta & -\cos\theta \end{bmatrix} & \text{if } \det \mathbf{U} \det \mathbf{V} = -1 \end{cases}. \tag{8.59}$$

Physically, this case corresponds to all of the points being collinear (in at least one of the frames) so that rotating about the axis formed by the points through any angle, θ, does not alter the objective function J.

Subcase (ii-c): $\text{rank}\, \mathbf{W} = 0$

This case corresponds to there being no points or all the points coincident and so any $\mathbf{C} \in SO(3)$ will produce the same value of the objective function, J.

Summary:

We have provided all of the solutions for $\mathbf{C}$ in our point-alignment problem; depending on the properties of $\mathbf{W}$, there can be one or infinitely many global solutions. Looking back through all the cases and subcases, we can see that if there is a unique global solution for $\mathbf{C}$, it is always of the form

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{8.60}$$

with $\mathbf{S} = \text{diag}(1, 1, \det \mathbf{U} \det \mathbf{V})$ and $\mathbf{W} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ is a singular-value decomposition of $\mathbf{W}$. The necessary and sufficient conditions for this unique global solution to exist are:

(i) $\det \mathbf{W} > 0$, or
(ii) $\det \mathbf{W} < 0$ and minimum singular value distinct: $d_1 \geq d_2 > d_3 > 0$, or
(iii) $\text{rank}\, \mathbf{W} = 2$.

If none of these conditions is true, there will be infinite solutions for $\mathbf{C}$. However, these cases are fairly pathological and do not occur frequently in practical situations.

Once we have solved for the optimal rotation matrix, we take $\hat{\mathbf{C}}_{v_k i} = \mathbf{C}$ as our estimated rotation. We build the estimated translation as

$$\hat{\mathbf{r}}_i^{v_k i} = \mathbf{p} - \hat{\mathbf{C}}_{v_k i}^T \mathbf{y} \tag{8.61}$$

and, if desired, combine the translation and rotation into an estimated transformation matrix,

$$\hat{\mathbf{T}}_{v_k i} = \begin{bmatrix} \hat{\mathbf{C}}_{v_k i} & -\hat{\mathbf{C}}_{v_k i}\hat{\mathbf{r}}_i^{v_k i} \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{8.62}$$

that provides the optimal alignment of the two point-clouds in a single quantity. Again, as mentioned in Section 6.3.2, we may actually be interested in $\hat{\mathbf{T}}_{i v_k}$, which can be recovered using

$$\hat{\mathbf{T}}_{i v_k} = \hat{\mathbf{T}}_{v_k i}^{-1} = \begin{bmatrix} \hat{\mathbf{C}}_{i v_k} & \hat{\mathbf{r}}_i^{v_k i} \\ \mathbf{0}^T & 1 \end{bmatrix}. \tag{8.63}$$

Both forms of the transformation matrix are useful, depending on how the solution will be used.

Example 8.1 We provide an example of *subcase (i-b)* to make things tangible. Consider the following two point-clouds that we wish to align, each consisting of six points:

$$\mathbf{p}_1 = 3 \times \mathbf{1}_1,\ \mathbf{p}_2 = 2 \times \mathbf{1}_2,\ \mathbf{p}_3 = \mathbf{1}_3,\ \mathbf{p}_4 = -3 \times \mathbf{1}_1,$$
$$\mathbf{p}_5 = -2 \times \mathbf{1}_2,\ \mathbf{p}_6 = -\mathbf{1}_3,$$
$$\mathbf{y}_1 = -3 \times \mathbf{1}_1,\ \mathbf{y}_2 = -2 \times \mathbf{1}_2,\ \mathbf{y}_3 = -\mathbf{1}_3,\ \mathbf{y}_4 = 3 \times \mathbf{1}_1,$$
$$\mathbf{y}_5 = 2 \times \mathbf{1}_2,\ \mathbf{y}_6 = \mathbf{1}_3,$$

where $\mathbf{1}_i$ is the ith column of the 3×3 identity matrix. The points in the first point-cloud are the centers of the faces of a rectangular prism and each point is associated with a point in the second point-cloud on the opposite face of another prism (that happens to be in the same location as the first)[7].

Using these points, we have the following:

$$\mathbf{p} = \mathbf{0}, \quad \mathbf{y} = \mathbf{0}, \quad \mathbf{W} = \frac{1}{6}\text{diag}(-18, -8, -2), \tag{8.64}$$

which means the centroids are already on top of one another so we only need to rotate to align the point-clouds.

Using the 'simplified approach', we have

$$\mathbf{C} = \left(\mathbf{W}\mathbf{W}^T\right)^{-\frac{1}{2}} \mathbf{W} = \text{diag}(-1, -1, -1). \tag{8.65}$$

Unfortunately, we can easily see that $\det \mathbf{C} = -1$ and so $\mathbf{C} \notin SO(3)$, which indicates this approach has failed.

For the more rigorous approach, a singular-value decomposition of $\mathbf{W}$ is

$$\begin{aligned} \mathbf{W} = \mathbf{U}\mathbf{D}\mathbf{V}^T, \quad \mathbf{U} = \text{diag}(1, 1, 1), \quad \mathbf{D} = \frac{1}{6}\text{diag}(18, 8, 2), \\ \mathbf{V} = \text{diag}(-1, -1, -1). \end{aligned} \tag{8.66}$$

We have $\det \mathbf{W} = -4/3 < 0$ and see that there is a unique minimum singular value, so we need to use the solution from *subcase (i-b)*. The minimal solution is therefore of the form $\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^T$ with $\mathbf{S} = \text{diag}(1, 1, -1)$. Plugging this in, we find

$$\mathbf{C} = \text{diag}(-1, -1, 1), \tag{8.67}$$

so that $\det \mathbf{C} = 1$. This is a rotation about the $\mathbf{1}_3$ axis through an angle π, which brings the error on four of the points to zero and leaves two of the points with non-zero error. This brings the objective function down to its minimum of $J = 4$.

Testing for Local Minima

In the previous section, we searched for global minima to the point-alignment problem and found there could be one or infinitely many. We did not, however, identify whether it was possible for local minima to exist, which we study now. Looking back to (8.30), this is the condition for a critical point in our optimization problem and therefore any solution that satisfies this criterion could be a minimum, a maximum, or a saddle point of the objective function, J.

If we have a solution, $\mathbf{C} \in SO(3)$, that satisfies (8.30), and we want

[7] As a physical interpretation, imagine joining each of the six point pairs by rubber bands. Finding the $\mathbf{C}$ that minimizes our cost metric is the same as finding the rotation that minimizes the amount of elastic energy stored in the rubber bands.

to characterize it, we can try perturbing the solution slightly and see whether the objective function goes up or down (or both). Consider a perturbation of the form

$$\mathbf{C}' = \exp\left(\boldsymbol{\phi}^\wedge\right)\mathbf{C}, \tag{8.68}$$

where $\boldsymbol{\phi} \in \mathbb{R}^3$ is a perturbation in an arbitrary direction, but constrained to keep $\mathbf{C}' \in SO(3)$. The change in the objective function δJ by applying the perturbation is

$$\delta J = J(\mathbf{C}') - J(\mathbf{C}) = -\mathrm{tr}(\mathbf{C}'\mathbf{W}^T) + \mathrm{tr}(\mathbf{C}\mathbf{W}^T) = -\mathrm{tr}\left((\mathbf{C}' - \mathbf{C})\mathbf{W}^T\right), \tag{8.69}$$

where we have neglected the Lagrange multiplier terms by assuming the perturbation keeps $\mathbf{C}' \in SO(3)$.

Now, approximating the perturbation out to second order, since this will tell us about the nature of the critical points, we have

$$\begin{aligned}\delta J &\approx -\mathrm{tr}\left(\left(\left(\mathbf{1} + \boldsymbol{\phi}^\wedge + \frac{1}{2}\boldsymbol{\phi}^\wedge\boldsymbol{\phi}^\wedge\right)\mathbf{C} - \mathbf{C}\right)\mathbf{W}^T\right) \\ &= -\mathrm{tr}\left(\boldsymbol{\phi}^\wedge\mathbf{C}\mathbf{W}^T\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\phi}^\wedge\boldsymbol{\phi}^\wedge\mathbf{C}\mathbf{W}^T\right).\end{aligned} \tag{8.70}$$

Then, plugging in the conditions for a critical point from (8.30), we have

$$\delta J = -\mathrm{tr}\left(\boldsymbol{\phi}^\wedge\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T\mathbf{U}^T\right) - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\phi}^\wedge\boldsymbol{\phi}^\wedge\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T\mathbf{U}^T\right). \tag{8.71}$$

It turns out that the first term is zero (because it is a critical point), which we can see from

$$\begin{aligned}\mathrm{tr}\left(\boldsymbol{\phi}^\wedge\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T\mathbf{U}^T\right) &= \mathrm{tr}\left(\mathbf{Y}^T\mathbf{U}^T\boldsymbol{\phi}^\wedge\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\right) \\ &= \mathrm{tr}\left(\left(\mathbf{Y}^T\mathbf{U}^T\boldsymbol{\phi}\right)^\wedge\mathbf{D}\mathbf{S}\right) = \mathrm{tr}\left(\boldsymbol{\varphi}^\wedge\mathbf{D}\mathbf{S}\right) = 0,\end{aligned} \tag{8.72}$$

where

$$\boldsymbol{\varphi} = \begin{bmatrix}\varphi_1 \\ \varphi_2 \\ \varphi_3\end{bmatrix} = \mathbf{Y}^T\mathbf{U}^T\boldsymbol{\phi}, \tag{8.73}$$

and owing to the properties of a skew-symmetric matrix (zeros on the diagonal). For the second term, we use the identity $\mathbf{u}^\wedge\mathbf{u}^\wedge = -\mathbf{u}^T\mathbf{u}\,\mathbf{1} + \mathbf{u}\mathbf{u}^T$ to write

$$\begin{aligned}\delta J &= -\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\phi}^\wedge\boldsymbol{\phi}^\wedge\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\mathbf{Y}^T\mathbf{U}^T\right) \\ &= -\frac{1}{2}\mathrm{tr}\left(\mathbf{Y}^T\mathbf{U}^T\left(-\boldsymbol{\phi}^T\boldsymbol{\phi}\,\mathbf{1} + \boldsymbol{\phi}\boldsymbol{\phi}^T\right)\mathbf{U}\mathbf{Y}\mathbf{D}\mathbf{S}\right) \\ &= -\frac{1}{2}\mathrm{tr}\left(\left(-\varphi^2\,\mathbf{1} + \boldsymbol{\varphi}\boldsymbol{\varphi}^T\right)\mathbf{D}\mathbf{S}\right),\end{aligned} \tag{8.74}$$

where $\varphi^2 = \boldsymbol{\varphi}^T\boldsymbol{\varphi} = \varphi_1^2 + \varphi_2^2 + \varphi_3^2$.

Manipulating a little further, we have

$$\begin{aligned}\delta J &= \frac{1}{2}\varphi^2 \mathrm{tr}(\mathbf{DS}) - \frac{1}{2}\boldsymbol{\varphi}^T\mathbf{DS}\boldsymbol{\varphi} \\ &= \frac{1}{2}\left(\varphi_1^2(d_2 s_2 + d_3 s_3) + \varphi_2^2(d_1 s_1 + d_3 s_3) + \varphi_3^2(d_1 s_1 + d_2 s_2)\right), \quad (8.75)\end{aligned}$$

the sign of which depends entirely on the nature of $\mathbf{DS}$.

We can verify the ability of this expression to test for a minimum using the unique global minima identified in the previous section. For *subcase (i-a)*, where $d_1 \geq d_2 \geq d_3$ and $s_1 = s_2 = s_3$, we have

$$\delta J = \frac{1}{2}\left(\varphi_1^2(d_2 + d_3) + \varphi_2^2(d_1 + d_3) + \varphi_3^2(d_1 + d_2)\right) > 0 \qquad (8.76)$$

for all $\boldsymbol{\varphi} \neq \mathbf{0}$, confirming a minimum. For *subcase (i-b)* where $d_1 \geq d_2 > d_3 > 0$ and $s_1 = s_2 = 1$, $s_3 = -1$, we have

$$\delta J = \frac{1}{2}\big(\varphi_1^2\underbrace{(d_2 - d_3)}_{>0} + \varphi_2^2\underbrace{(d_1 - d_3)}_{>0} + \varphi_3^2(d_1 + d_2)\big) > 0, \qquad (8.77)$$

for all $\boldsymbol{\varphi} \neq \mathbf{0}$, again confirming a minimum. Finally, for *subcase (ii-a)* where $d_1 \geq d_2 > d_3 = 0$ and $s_1 = s_2 = 1$, $s_3 = \pm 1$, we have

$$\delta J = \frac{1}{2}\left(\varphi_1^2 d_2 + \varphi_2^2 d_1 + \varphi_3^2(d_1 + d_2)\right) > 0, \qquad (8.78)$$

for all $\boldsymbol{\varphi} \neq \mathbf{0}$, once again confirming a minimum.

The more interesting question is whether there are any other local minima to worry about or not. This will become important when we use iterative methods to optimize rotation and pose variables. For example, let us consider *subcase (i-a)* a little further in the case that $d_1 > d_2 > d_3 > 0$. There are some other ways to satisfy (8.30) and generate a critical point. For example, we could pick $s_1 = s_2 = -1$ and $s_3 = 1$ so that $\det \mathbf{S} = 1$. In this case we have

$$\delta J = \frac{1}{2}\big(\varphi_1^2\underbrace{(d_3 - d_2)}_{<0} + \varphi_2^2\underbrace{(d_3 - d_1)}_{<0} + \varphi_3^2\underbrace{(-d_1 - d_2)}_{<0}\big) < 0, \qquad (8.79)$$

which corresponds to a maximum since any $\boldsymbol{\varphi} \neq \mathbf{0}$ will decrease the objective function. The other two cases, $\mathbf{S} = \mathrm{diag}(-1, 1, -1)$ and $\mathbf{S} = \mathrm{diag}(1, -1, -1)$, turn out to be saddle points since depending on the direction of the perturbation, the objective function can go up or down. Since there are no other critical points, we can conclude there are no local minima other than the global one.

Similarly for *subcase (i-b)*, we need $\det \mathbf{S} = -1$ and can show that $\mathbf{S} = \mathrm{diag}(-1, -1, -1)$ is a maximum and that $\mathbf{S} = \mathrm{diag}(-1, 1, 1)$ and $\mathbf{S} = \mathrm{diag}(1, -1, 1)$ are saddle points. Again, since there are no other critical points, we can conclude there are no local minima other than the global one.

Also, for *subcase (ii-a)* we in general have

$$\delta J = \frac{1}{2}\left(\varphi_1^2 d_2 s_2 + \varphi_2^2 d_1 s_1 + \varphi_3^2 (d_1 s_1 + d_2 s_2)\right), \tag{8.80}$$

and so the only way to create a local minimum is to pick $s_1 = s_2 = 1$, which is the global minimum we have discussed earlier. Thus, again there are no additional local minima.

Iterative Approach

We can also consider using an iterative approach to solve for the optimal rotation matrix, $\mathbf{C}$. We will use our $SO(3)$-sensitive scheme to do this. Importantly, the optimization we carry out is unconstrained, thereby avoiding the difficulties of the previous two approaches[8]. Technically, the result is not valid globally, only locally, as we require an initial guess that is refined from one iteration to the next; typically only a few iterations are needed. However, based on our discussion of local minima in the last section, we know that in all the important situations where there is a unique global minimum, there are no additional local minima to worry about.

Starting from the cost function where the translation has been eliminated,

$$J(\mathbf{C}) = \frac{1}{2}\sum_{j=1}^{M} w_j \left((\mathbf{y}_j - \mathbf{y}) - \mathbf{C}(\mathbf{p}_j - \mathbf{p})\right)^T \left((\mathbf{y}_j - \mathbf{y}) - \mathbf{C}(\mathbf{p}_j - \mathbf{p})\right), \tag{8.81}$$

we can insert the $SO(3)$-sensitive perturbation,

$$\mathbf{C} = \exp\left(\boldsymbol{\psi}^\wedge\right)\mathbf{C}_{\text{op}} \approx \left(\mathbf{1} + \boldsymbol{\psi}^\wedge\right)\mathbf{C}_{\text{op}}, \tag{8.82}$$

where $\mathbf{C}_{\text{op}}$ is the current guess and $\boldsymbol{\psi}$ is the perturbation; we will seek an optimal value to update the guess (and then iterate). Inserting the approximate perturbation scheme into the cost function turns it into a quadratic in $\boldsymbol{\psi}$ for which the minimizing value, $\boldsymbol{\psi}^\star$, is given by the solution to

$$\underbrace{\mathbf{C}_{\text{op}}\left(-\frac{1}{w}\sum_{j=1}^{M} w_j (\mathbf{p}_j - \mathbf{p})^\wedge (\mathbf{p}_j - \mathbf{p})^\wedge\right)\mathbf{C}_{\text{op}}^T}_{\text{constant}}\boldsymbol{\psi}^\star = -\frac{1}{w}\sum_{j=1}^{M} w_j (\mathbf{y}_j - \mathbf{y})^\wedge \mathbf{C}_{\text{op}}(\mathbf{p}_j - \mathbf{p}). \tag{8.83}$$

[8] The iterative approach does not require solving either an eigenproblem nor carrying out a singular-value decomposition.

At first glance, the right-hand side appears to require recalculation using the individual points at each iteration. Fortunately, we can manipulate it into a more useful form. The right-hand side is a 3×1 column, and its ith row is given by

$$\begin{aligned}
\mathbf{1}_i^T & \left(-\frac{1}{w} \sum_{j=1}^{M} w_j (\mathbf{y}_j - \mathbf{y})^\wedge \mathbf{C}_{\text{op}} (\mathbf{p}_j - \mathbf{p}) \right) \\
&= \frac{1}{w} \sum_{j=1}^{M} w_j (\mathbf{y}_j - \mathbf{y})^T \mathbf{1}_i^\wedge \mathbf{C}_{\text{op}} (\mathbf{p}_j - \mathbf{p}) \\
&= \frac{1}{w} \sum_{j=1}^{M} w_j \text{tr} \left(\mathbf{1}_i^\wedge \mathbf{C}_{\text{op}} (\mathbf{p}_j - \mathbf{p})(\mathbf{y}_j - \mathbf{y})^T \right) \\
&= \text{tr} \left(\mathbf{1}_i^\wedge \mathbf{C}_{\text{op}} \mathbf{W}^T \right), \quad (8.84)
\end{aligned}$$

where

$$\mathbf{W} = \frac{1}{w} \sum_{j=1}^{M} w_j (\mathbf{y}_j - \mathbf{y})(\mathbf{p}_j - \mathbf{p})^T, \tag{8.85}$$

which we already saw in the non-iterative solution. Letting

$$\mathbf{I} = -\frac{1}{w} \sum_{j=1}^{M} w_j (\mathbf{p}_j - \mathbf{p})^\wedge (\mathbf{p}_j - \mathbf{p})^\wedge, \tag{8.86a}$$

$$\mathbf{b} = \left[\text{tr} \left(\mathbf{1}_i^\wedge \mathbf{C}_{\text{op}} \mathbf{W}^T \right) \right]_i, \tag{8.86b}$$

the optimal update can be written in closed form as

$$\boldsymbol{\psi}^\star = \mathbf{C}_{\text{op}} \mathbf{I}^{-1} \mathbf{C}_{\text{op}}^T \, \mathbf{b}. \tag{8.87}$$

We apply this to the initial guess,

$$\mathbf{C}_{\text{op}} \leftarrow \exp \left(\boldsymbol{\psi}^{\star^\wedge} \right) \mathbf{C}_{\text{op}}, \tag{8.88}$$

and iterate to convergence, taking $\hat{\mathbf{C}}_{v_k i} = \mathbf{C}_{\text{op}}$ at the final iteration as our rotation estimate. After convergence, the translation is given as in the non-iterative scheme:

$$\hat{\mathbf{r}}_i^{v_k i} = \mathbf{p} - \hat{\mathbf{C}}_{v_k i}^T \mathbf{y}. \tag{8.89}$$

Notably, both $\mathbf{I}$ and $\mathbf{W}$ can be computed in advance, and therefore we do not require the original points during execution of the iterative scheme.

Three Non-collinear Points Required

Clearly, to solve uniquely for $\boldsymbol{\psi}^\star$ above, we need $\det \mathbf{I} \neq 0$. A sufficient condition is to have $\mathbf{I}$ positive-definite, which implies that for any $\mathbf{x} \neq \mathbf{0}$, we must have

$$\mathbf{x}^T \mathbf{I} \, \mathbf{x} > 0. \tag{8.90}$$

We then notice that

$$\begin{aligned}\mathbf{x}^T\mathbf{I}\,\mathbf{x} &= \mathbf{x}^T\left(-\frac{1}{w}\sum_{j=1}^{M} w_j(\mathbf{p}_j-\mathbf{p})^\wedge(\mathbf{p}_j-\mathbf{p})^\wedge\right)\mathbf{x} \\ &= \frac{1}{w}\sum_{j=1}^{M} w_j \underbrace{\left((\mathbf{p}_j-\mathbf{p})^\wedge\mathbf{x}\right)^T\left((\mathbf{p}_j-\mathbf{p})^\wedge\mathbf{x}\right)}_{\geq 0} \geq 0. \end{aligned} \tag{8.91}$$

Since each term in the sum is non-negative, the total must be non-negative. The only way to have the total be zero is if *every* term in the sum is also zero, or

$$(\forall j)\ (\mathbf{p}_j-\mathbf{p})^\wedge\mathbf{x} = \mathbf{0}. \tag{8.92}$$

In other words, we must have $\mathbf{x} = \mathbf{0}$ (not true by assumption), $\mathbf{p}_j = \mathbf{p}$, or $\mathbf{x}$ parallel to $\mathbf{p}_j - \mathbf{p}$. The last two conditions are never true as long as there are at least three points and they are not collinear.

Note that having three non-collinear points only provides a sufficient condition for a unique solution for $\boldsymbol{\psi}^\star$ at each iteration, and does not tell us about the number of possible global solutions to minimize our objective function in general. This was discussed at length in the previous sections, where we learned there could be one or infinitely many global solutions. Moreover, if there is a unique global minimum, there are no local minima to worry about.

8.1.4 Transformation Matrix Solution

Finally, for completeness, we also can provide an iterative approach to solving for the pose change using transformation matrices and their relationship to the exponential map[9]. As in the previous two sections, we will use some simplified notation to avoid repeating sub- and superscripts:

$$\begin{aligned}\boldsymbol{y}_j = \begin{bmatrix}\mathbf{y}_j \\ 1\end{bmatrix} = \begin{bmatrix}\mathbf{r}_{v_k}^{p_j v_k} \\ 1\end{bmatrix}, \quad \boldsymbol{p}_j = \begin{bmatrix}\mathbf{p}_j \\ 1\end{bmatrix} = \begin{bmatrix}\mathbf{r}_i^{p_j i} \\ 1\end{bmatrix}, \\ \mathbf{T} = \mathbf{T}_{v_k i} = \begin{bmatrix}\mathbf{C}_{v_k i} & -\mathbf{C}_{v_k i}\mathbf{r}_i^{v_k i} \\ \mathbf{0}^T & 1\end{bmatrix}. \end{aligned} \tag{8.93}$$

We have used a different font for the homogeneous representations of the points; we will be making connections back to the previous section on rotation matrices so we also keep the non-homogeneous point representations around for convenience.

We define our error term for each point as

$$\mathbf{e}_j = \boldsymbol{y}_j - \mathbf{T}\boldsymbol{p}_j, \tag{8.94}$$

[9] We will use the optimization approach outlined in Section 7.1.9.