S.1 Proofs

S.1.1 Lemma 3.1

Proof. From the matrix inversion lemma we have

$$\mathbf{A}^{-1} = \frac{1}{\alpha}\mathbf{I} - \frac{1}{\alpha^2}\mathbf{\Phi}_x^T\left(\frac{1}{\beta}\mathbf{I} + \frac{1}{\alpha}\mathbf{\Phi}_x\mathbf{\Phi}_x^T\right)^{-1}\mathbf{\Phi}_x = \frac{1}{\alpha}\mathbf{I} - \frac{1}{\alpha^2}\mathbf{\Phi}_x^T\left(\frac{1}{\beta}\mathbf{I} + \frac{1}{\alpha}\mathbf{K}_x^T\right)^{-1}\mathbf{\Phi}_x \tag{16}$$

$$= \frac{1}{\alpha}\mathbf{I} - \frac{1}{\alpha^2}\mathbf{\Phi}_x^T\left(\frac{1}{\beta}\mathbf{V}\mathbf{V}^T + \frac{1}{\alpha}\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\right)^{-1}\mathbf{\Phi}_x = \frac{1}{\alpha}\mathbf{I} - \frac{1}{\alpha^2}\mathbf{\Phi}_x^T\mathbf{V}\left(\frac{1}{\beta}\mathbf{I} + \frac{1}{\alpha}\mathbf{\Lambda}\right)^{-1}\mathbf{V}^T\mathbf{\Phi}_x$$

$$= \frac{1}{\alpha}\mathbf{I} - \mathbf{\Phi}_x^T\mathbf{V}\,\mathbf{diag}\left\{\alpha\lambda_i + \frac{\alpha^2}{\beta}\right\}^{-1}\mathbf{V}^T\mathbf{\Phi}_x \tag{17}$$

Now suppose that $\mathbf{A}^{-\frac{1}{2}} = \frac{1}{\sqrt{\alpha}}\mathbf{I} - \mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x$ for unknown diagonal matrix $\mathbf{D}$. Squaring, we obtain

$$\frac{1}{\alpha}\mathbf{I} - \frac{2}{\sqrt{\alpha}}\mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x + \mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{K}_x\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x \tag{18}$$

$$= \frac{1}{\alpha}\mathbf{I} - \frac{2}{\sqrt{\alpha}}\mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x + \mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x \tag{19}$$

$$= \frac{1}{\alpha}\mathbf{I} - \frac{2}{\sqrt{\alpha}}\mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x + \mathbf{\Phi}_x^T\mathbf{V}\mathbf{D}\mathbf{\Lambda}\mathbf{D}\mathbf{V}^T\mathbf{\Phi}_x \tag{20}$$

$$= \frac{1}{\alpha}\mathbf{I} - \mathbf{\Phi}_x^T\mathbf{V}\left(\frac{2}{\sqrt{\alpha}}\mathbf{D} - \mathbf{D}\mathbf{\Lambda}\mathbf{D}\right)\mathbf{V}^T\mathbf{\Phi}_x \tag{21}$$

$$= \frac{1}{\alpha}\mathbf{I} - \mathbf{\Phi}_x^T\mathbf{V}\,\mathbf{diag}\left\{\frac{2d_i}{\sqrt{\alpha}} - \lambda_i d_i^2\right\}\mathbf{V}^T\mathbf{\Phi}_x\ . \tag{22}$$

To solve the d_i we equate

$$\frac{1}{\alpha\lambda_i + \frac{\alpha^2}{\beta}} = \frac{2d_i}{\sqrt{\alpha}} - \lambda_i d_i^2\ , \tag{23}$$

which is quadratic in d_i. Solving yields

$$d_i = \frac{1}{\lambda_i}\left(\frac{1}{\sqrt{\alpha}} \pm \frac{1}{\sqrt{\alpha + \beta\lambda_i}}\right)\ . \tag{24}$$

Selecting the minus (vs the plus) will produce a pd matrix. We obtain

$$\mathbf{A}^{-\frac{1}{2}} = \frac{1}{\sqrt{\alpha}}\mathbf{I} + \mathbf{\Phi}_x^T\mathbf{V}\,\mathbf{diag}\left\{\frac{1}{\lambda_i}\left(\frac{1}{\sqrt{\alpha + \beta\lambda_i}} - \frac{1}{\sqrt{\alpha}}\right)\right\}\mathbf{V}^T\mathbf{\Phi}_x = \frac{1}{\sqrt{\alpha}}\mathbf{I} + \mathbf{\Phi}_x^T\mathbf{C}\mathbf{\Phi}_x\ . \tag{25}$$

It follows

$$\mathbf{\Phi}_x\mathbf{A}^{-\frac{1}{2}} = \frac{1}{\sqrt{\alpha}}\mathbf{\Phi}_x + \mathbf{K}_x\mathbf{V}\,\mathbf{diag}\left\{\frac{1}{\lambda_i}\left(\frac{1}{\sqrt{\alpha + \beta\lambda_i}} - \frac{1}{\sqrt{\alpha}}\right)\right\}\mathbf{V}^T\mathbf{\Phi}_x \tag{26}$$

$$= \frac{1}{\sqrt{\alpha}}\mathbf{\Phi}_x + \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T\mathbf{V}\,\mathbf{diag}\left\{\frac{1}{\lambda_i}\left(\frac{1}{\sqrt{\alpha + \beta\lambda_i}} - \frac{1}{\sqrt{\alpha}}\right)\right\}\mathbf{V}^T\mathbf{\Phi}_x \tag{27}$$

$$= \frac{1}{\sqrt{\alpha}}\mathbf{V}\mathbf{V}^T\mathbf{\Phi}_x + \mathbf{V}\,\mathbf{diag}\left\{\frac{1}{\sqrt{\alpha + \beta\lambda_i}} - \frac{1}{\sqrt{\alpha}}\right\}\mathbf{V}^T\mathbf{\Phi}_x \tag{28}$$

$$= \mathbf{V}\left(\frac{1}{\sqrt{\alpha}}\mathbf{I} + \mathbf{diag}\left\{\frac{1}{\sqrt{\alpha + \beta\lambda_i}} - \frac{1}{\sqrt{\alpha}}\right\}\right)\mathbf{V}^T\mathbf{\Phi}_x \tag{29}$$

$$= \mathbf{V}\,\mathbf{diag}\left\{\frac{1}{\sqrt{\alpha + \beta\lambda_i}}\right\}\mathbf{V}^T\mathbf{\Phi}_x = \mathbf{B}\mathbf{\Phi}_x \tag{30}$$

$\square$

S.1.2 Lemma 3.2

Proof. Observe:

$$\tilde{\mathbf{Q}}^T\tilde{\mathbf{Q}} = (\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\tilde{\mathbf{G}}\tilde{\mathbf{U}}^T)^T(\mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\tilde{\mathbf{G}}\tilde{\mathbf{U}}^T) \tag{31}$$

$$= \mathbf{I} - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\tilde{\mathbf{G}}\tilde{\mathbf{U}}^T - \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T - \tilde{\mathbf{U}}\tilde{\mathbf{G}}\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\tilde{\mathbf{G}}^T\tilde{\mathbf{U}}^T - \tilde{\mathbf{U}}\tilde{\mathbf{G}}^T\tilde{\mathbf{U}}^T + \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \tag{32}$$

$$= \mathbf{I} \tag{33}$$

□

S.1.3 Theorem 3.3

We begin with preliminary definitions. For $\mathbf{V} \in \text{St}(N, r)$, we define

$$\mathring{\mathcal{R}}(\mathbf{V}) \triangleq \{\mathbf{Y} : \mathbf{Y} \in \text{St}(N, r) \text{ and } \exists\, \mathbf{A} \in \mathcal{O}(r) \text{ such that } \mathbf{Y} = \mathbf{VA}\}$$
$$\mathring{\mathcal{N}}(\mathbf{V}) \triangleq \{\mathbf{Y} : \mathbf{Y} \in \text{St}(N, N-r) \text{ and } \mathbf{V}^T\mathbf{Y} = \mathbf{0}_{r\times(N-r)}\} \ .$$

It follows immediately that $\mathbf{Y} \in \mathring{\mathcal{R}}(\mathbf{V}) \Leftrightarrow \mathbf{V} \in \mathring{\mathcal{R}}(\mathbf{Y})$ and $\mathbf{Y} \in \mathring{\mathcal{N}}(\mathbf{V}) \Leftrightarrow \mathbf{V} \in \mathring{\mathcal{N}}(\mathbf{Y})$.

Given $\mathbf{Q} \in \mathcal{O}(N)$, we would like to know if $\mathbf{Q}$ can be expressed as $\mathbf{I}_N - \mathbf{V}(\mathbf{I}_r - \mathbf{G})\mathbf{V}^T$ for some orthogonal $\mathbf{G}$. If so, we can write $\mathbf{Q} = \mathbf{I}_N - \mathbf{V}(\mathbf{I}_r - \mathbf{G})\mathbf{V}^T = \mathbf{I}_N - (\mathbf{VA})(\mathbf{I}_r - \mathbf{A}^T\mathbf{GA})(\mathbf{VA})^T$, $\mathbf{A}$ orthogonal, implying that $\mathbf{Q}$ can be decomposed via any member of $\mathring{\mathcal{R}}(\mathbf{V})$. When this decomposition is possible, we say $\mathbf{Q}$ *is supported by* $\mathring{\mathcal{R}}(\mathbf{V})$.

Lemma S.1.1. *Given $\mathbf{Q} \in \mathcal{O}(N)$ and $\mathbf{V} \in St(N, r)$. $\mathbf{Q}$ is supported by $\mathring{\mathcal{R}}(\mathbf{V})$ if and only if for any $\mathbf{V}_\perp \in \mathring{\mathcal{N}}(\mathbf{V})$ we have $\mathbf{V}_\perp^T\mathbf{Q}\mathbf{V}_\perp = \mathbf{I}_{N-r}$.*

Proof. ($\Rightarrow$) With $\mathbf{Q}$ supported by $\mathring{\mathcal{R}}(\mathbf{V})$, there exists $\mathbf{G} \in \mathcal{O}(r)$ such that $\mathbf{Q} = \mathbf{I}_N - \mathbf{V}(\mathbf{I}_r - \mathbf{G})\mathbf{V}^T$. It follows that $\mathbf{V}_\perp^T\mathbf{Q}\mathbf{V}_\perp = \mathbf{V}_\perp^T\mathbf{V}_\perp - \mathbf{V}_\perp^T\mathbf{V}(\mathbf{I}_r - \mathbf{G})\mathbf{V}^T\mathbf{V}_\perp = \mathbf{I}_{N-r} - \mathbf{0} = \mathbf{I}_{N-r}$.
($\Leftarrow$) The matrix $\bar{\mathbf{V}} = (\mathbf{V} \mid \mathbf{V}_\perp)$ is an element of $\mathcal{O}(N)$. We may therefore write

$$\mathbf{Q} = \bar{\mathbf{V}}\bar{\mathbf{V}}^T\mathbf{Q}\bar{\mathbf{V}}\bar{\mathbf{V}}^T = \bar{\mathbf{V}}\begin{pmatrix}\mathbf{V}^T\\ \mathbf{V}_\perp^T\end{pmatrix}\mathbf{Q}\,(\mathbf{V} \mid \mathbf{V}_\perp)\,\bar{\mathbf{V}}^T = \bar{\mathbf{V}}\begin{pmatrix}\mathbf{V}^T\mathbf{QV} & \mathbf{V}^T\mathbf{Q}\mathbf{V}_\perp\\ \mathbf{V}_\perp^T\mathbf{QV} & \mathbf{V}_\perp^T\mathbf{Q}\mathbf{V}_\perp\end{pmatrix}\bar{\mathbf{V}}^T \tag{34}$$

$$= \bar{\mathbf{V}}\begin{pmatrix}\mathbf{V}^T\mathbf{QV} & \mathbf{V}^T\mathbf{Q}\mathbf{V}_\perp\\ \mathbf{V}_\perp^T\mathbf{QV} & \mathbf{I}\end{pmatrix}\bar{\mathbf{V}}^T \ . \tag{35}$$

Whenever an orthogonal matrix contains an ij-th element of ±1, the remaining i-th row and j-th column elements are 0. With $\bar{\mathbf{V}}^T\mathbf{Q}\bar{\mathbf{V}}$ orthogonal, the identity block in the bottom-right corner implies that $\mathbf{V}^T\mathbf{Q}\mathbf{V}_\perp = \mathbf{0}$ and $\mathbf{V}_\perp^T\mathbf{QV} = \mathbf{0}$. The result is a block diagonal orthogonal matrix; the first block is $\mathbf{V}^T\mathbf{QV}$ and the second is identity. It follows that $\mathbf{G} = \mathbf{V}^T\mathbf{QV} \in \mathcal{O}(r)$. We have

$$\mathbf{Q} = \bar{\mathbf{V}}\begin{pmatrix}\mathbf{V}^T\mathbf{QV} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\bar{\mathbf{V}}^T = (\mathbf{V} \mid \mathbf{V}_\perp)\begin{pmatrix}\mathbf{V}^T\mathbf{QV} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{pmatrix}\begin{pmatrix}\mathbf{V}^T\\ \mathbf{V}_\perp^T\end{pmatrix} \tag{36}$$

$$= \mathbf{V}_\perp\mathbf{V}_\perp^T + \mathbf{V}\mathbf{V}^T\mathbf{Q}\mathbf{V}\mathbf{V}^T \tag{37}$$

$$= (\mathbf{V}_\perp\mathbf{V}_\perp^T + \mathbf{V}\mathbf{V}^T) - (\mathbf{V}\mathbf{V}^T + \mathbf{VG}\mathbf{V}^T) \tag{38}$$

$$= \mathbf{I}_N - \mathbf{V}(\mathbf{I}_r - \mathbf{G})\mathbf{V}^T \tag{39}$$

The above simplification follows from $\mathbf{I}_N = \bar{\mathbf{V}}\bar{\mathbf{V}}^T = \mathbf{V}\mathbf{V}^T + \mathbf{V}_\perp\mathbf{V}_\perp^T$. □

Corollary S.1.2. *$\mathbf{I}_N$ is supported by $\mathring{\mathcal{R}}(\mathbf{V})$.*

We can now proceed to the optimization problem. The objective of interest, $f : \mathcal{O}(N)^m \to \mathbb{R}$, is

$$f(\mathbf{Q}_1, \dots, \mathbf{Q}_m) = \sum_{i<j} \|\mathbf{Z}_i\mathbf{Q}_i - \mathbf{Z}_j\mathbf{Q}_j\|_{\text{F}}^2 \ . \tag{40}$$

The next two lemmas are independent of f, but necessary for the final theorem.

Lemma S.1.3. *For $\mathbf{Y}_{1:m}$, we have $\frac{1}{m}\sum_{i=1}^m \mathbf{Y}_i = \arg\min_{\mathbf{C}} \sum_{i=1}^m \|\mathbf{Y}_i - \mathbf{C}\|_{\text{F}}^2$.*

Proof. Taking the derivative with respect to $\mathbf{C}$ and setting equal to zero, we have:

$$\mathbf{0} = 2\sum_{i=1}^{m}(\mathbf{C} - \mathbf{Y}_i) \qquad \Rightarrow \qquad \mathbf{C} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{Y}_i \tag{41}$$

□

Lemma S.1.4. *For* $\mathbf{C} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{Y}_i$ *we have* $\sum_{i<j}^{m}\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathrm{F}}^2 = \frac{1}{2}\sum_{i,j}^{m}\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathrm{F}}^2 = m\sum_{i=1}^{m}\|\mathbf{Y}_i - \mathbf{C}\|_{\mathrm{F}}^2$.

Proof. The first equality is due to

$$2\sum_{i<j}^{m}\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathrm{F}}^2 = \sum_{i<j}^{m}\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathrm{F}}^2 + \sum_{i>j}^{m}\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathrm{F}}^2 + \sum_{i=1}^{m}\|\mathbf{Y}_i - \mathbf{Y}_i\|_{\mathrm{F}}^2 \ . \tag{42}$$

The second equality is due to

$$\begin{aligned}
\sum_{i=1}^{m}\sum_{j=1}^{m}\|\mathbf{Y}_i - \mathbf{Y}_j\|_{\mathrm{F}}^2 &= \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i) - 2\,\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_j) + \mathrm{tr}(\mathbf{Y}_j^T\mathbf{Y}_j)\right) && (43)\\
&= \sum_{i=1}^{m}\left(m\,\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i) - 2m\,\mathrm{tr}(\mathbf{Y}_i^T\mathbf{C}) + \sum_{j=1}^{m}\mathrm{tr}(\mathbf{Y}_j^T\mathbf{Y}_j)\right) && (44)\\
&= \left(m\sum_{i=1}^{m}\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i)\right) - \left(2m^2\,\mathrm{tr}(\mathbf{C}^T\mathbf{C})\right) + \left(m\sum_{j=1}^{m}\mathrm{tr}(\mathbf{Y}_j^T\mathbf{Y}_j)\right) && (45)\\
&= \left(2m\sum_{i=1}^{m}\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i)\right) - \left(2m^2\,\mathrm{tr}(\mathbf{C}^T\mathbf{C})\right) && (46)
\end{aligned}$$

and

$$\begin{aligned}
m\sum_{i=1}^{m}\|\mathbf{Y}_i - \mathbf{C}\|_{\mathrm{F}}^2 &= m\sum_{i=1}^{m}\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i) - 2\,\mathrm{tr}(\mathbf{Y}_i^T\mathbf{C}) + \mathrm{tr}(\mathbf{C}^T\mathbf{C}) && (47)\\
&= \left(m\sum_{i=1}^{m}\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i)\right) - \left(2m^2\,\mathrm{tr}(\mathbf{C}^T\mathbf{C})\right) + \left(m^2\,\mathrm{tr}(\mathbf{C}^T\mathbf{C})\right) && (48)\\
&= \left(m\sum_{i=1}^{m}\mathrm{tr}(\mathbf{Y}_i^T\mathbf{Y}_i)\right) - \left(m^2\,\mathrm{tr}(\mathbf{C}^T\mathbf{C})\right) && (49)
\end{aligned}$$

□

Lemma S.1.5. *Let* $\mathbf{R}_{1:m} \in \mathcal{O}(N)$ *and let* $\mathcal{A} \subseteq \{1, 2, \ldots, m\}$*. The following algorithm, which updates* $\mathbf{R}_i$ *to* $\mathbf{R}_i'$*, results in* $f(\mathbf{R}_{1:m}) \geq f(\mathbf{R}_{1:m}')$*:*

1. $\mathbf{C} \leftarrow \frac{1}{m}\sum_{i=1}^{m}\mathbf{Z}_i\mathbf{R}_i$
2. *for each* $j \in \{1, 2, \ldots, m\}\backslash\mathcal{A}$*:* $\mathbf{R}_j' \leftarrow \mathbf{R}_j$
3. *for each* $j \in \mathcal{A}$*:* $\mathbf{R}_j' \leftarrow \arg\min_{\mathbf{Q}\in\mathcal{O}(N)}\|\mathbf{Z}_j\mathbf{Q} - \mathbf{C}\|_{\mathrm{F}}^2$

Proof. Let $\mathbf{C}' = \frac{1}{m}\sum_{i=1}^m \mathbf{Z}_i\mathbf{R}_i'$. We have

$$
\begin{aligned}
f(\mathbf{R}_{1:m}) &= m\sum_{i=1}^m \|\mathbf{Z}_i\mathbf{R}_i - \mathbf{C}\|_{\mathrm{F}}^2 && \text{from Lemma S.1.4} && (50)\\
&\geq m\sum_{i=1}^m \|\mathbf{Z}_i\mathbf{R}_i' - \mathbf{C}\|_{\mathrm{F}}^2 && \text{from the algorithm} && (51)\\
&\geq m\sum_{i=1}^m \|\mathbf{Z}_i\mathbf{R}_i' - \mathbf{C}'\|_{\mathrm{F}}^2 && \text{from Lemma S.1.3} && (52)\\
&= f(\mathbf{R}_{1:m}') && \text{from Lemma S.1.4 .} && (53)
\end{aligned}
$$

□

Lemma S.1.6. *f is bounded by below.*

Proof. f is the sum of nonnegative entries, so it is lower bounded by 0. □

Lemma S.1.7. *For $\mathbf{R} \in \mathcal{O}(N)$, $f(\mathbf{Q}_1, \mathbf{Q}_2, \ldots, \mathbf{Q}_m) = f(\mathbf{Q}_1\mathbf{R}, \mathbf{Q}_2\mathbf{R}, \ldots, \mathbf{Q}_m\mathbf{R})$.*

Proof. The Frobenius norm is unitariliy invariant: $\|\mathbf{Z}_i\mathbf{Q}_i\mathbf{R} - \mathbf{Z}_j\mathbf{Q}_j\mathbf{R}\|_{\mathrm{F}}^2 = \|\mathbf{Z}_i\mathbf{Q}_i - \mathbf{Z}_j\mathbf{Q}_j\|_{\mathrm{F}}^2$. □

In what follows, we use $f(\mathbf{Q}_{1:m}) = \frac{1}{2}\sum_{i,j}^m \|\mathbf{Z}_i\mathbf{Q}_i - \mathbf{Z}_j\mathbf{Q}_j\|_{\mathrm{F}}^2$ (Lemma S.1.4). With respect to $\mathbf{Q}_k$, the unconstrained derivative is

$$\frac{\partial f}{\partial \mathbf{Q}_k} = -\mathbf{Z}_k^T\sum_{i=1}^m \mathbf{Z}_i\mathbf{Q}_i \tag{54}$$

Thus, a critical point $(\mathbf{S}_1, \ldots, \mathbf{S}_m)$ must satisfy [5]:

$$\mathbf{S}_k^T\mathbf{Z}_k^T\sum_{i=1}^m \mathbf{Z}_i\mathbf{S}_i \in \mathbb{S}^N \tag{55}$$

for each k, where $\mathbb{S}^N$ denotes the set of $N \times N$ symmetric matrices.

Lemma S.1.8. *For $\mathbf{R} \in \mathcal{O}(N)$, if $(\mathbf{S}_1, \ldots, \mathbf{S}_m)$ is a critical point of f, then so is $(\mathbf{S}_1\mathbf{R}, \ldots, \mathbf{S}_m\mathbf{R})$.*

Proof. For each k we have $\mathbf{F}_k = \mathbf{S}_k^T\mathbf{Z}_k^T\sum_{i=1}^m \mathbf{Z}_i\mathbf{S}_i$ symmetric. If $\mathbf{F}_k$ is symmetric then so is $\mathbf{R}^T\mathbf{F}_k\mathbf{R} = (\mathbf{S}_k\mathbf{R})^T\mathbf{Z}_k^T\sum_{i=1}^m \mathbf{Z}_i(\mathbf{S}_i\mathbf{R})$. □

Let $r = mt$ and let $\mathbf{W} = [\mathbf{Z}_1^T\ \mathbf{Z}_2^T\ \cdots\ \mathbf{Z}_m^T]^T \in \mathbb{R}^{r\times N}$ have SVD $\mathbf{U\Sigma V}^T$, where $\mathbf{U} \in \mathcal{O}(r)$ and $\mathbf{V} \in \mathrm{St}(N, r)$. In the notation of kernel hyperalignment, we have $\mathbf{Z}_i = \mathbf{\Phi}_i\mathbf{A}_i^{-\frac{1}{2}}$ and $\mathbf{\Phi}_0 = [\mathbf{\Phi}_1^T\ \mathbf{\Phi}_2^T\ \cdots\ \mathbf{\Phi}_m^T]^T$.

Lemma S.1.9. $\mathbf{Z}_i = \mathbf{Z}_i\mathbf{V}\mathbf{V}^T$.

Proof. Let $\mathbf{E}_i = [\underbrace{\mathbf{0}_{t\times t}\ \cdots \mathbf{0}_{t\times t}}_{i-1}\ \mathbf{I}_t\ \underbrace{\mathbf{0}_{t\times t}\ \cdots \mathbf{0}_{t\times t}}_{m-i}]^T$. We have $\mathbf{Z}_i = \mathbf{E}_i^T\mathbf{W} = \mathbf{E}_i^T\mathbf{U\Sigma V}^T$ and $\mathbf{Z}_i\mathbf{V}\mathbf{V}^T = (\mathbf{E}_i^T\mathbf{W})\mathbf{V}\mathbf{V}^T = \mathbf{E}_i^T\mathbf{U\Sigma V}^T\mathbf{V}\mathbf{V}^T = \mathbf{E}_i^T\mathbf{U\Sigma V}^T$. □

Theorem S.1.10. *Any global minimizer of f can be mapped to another global minimizer whose entries are supported by $\mathring{\mathcal{R}}(\mathbf{V})$.*

Proof. Let $(\tilde{\mathbf{S}}_1, \ldots, \tilde{\mathbf{S}}_m)$ be a global minimizer of f. Because f is differentiable and bounded by below, $(\tilde{\mathbf{S}}_1, \ldots, \tilde{\mathbf{S}}_m)$ is a critical point. We form $\mathbf{S}_i = \tilde{\mathbf{S}}_i\mathbf{S}_1^T$. From Lemmas S.1.7 and S.1.8 $\mathbf{S}_{1:m}$ is also a global minimizer and a critical point. With $(\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_m) = (\mathbf{I}, \mathbf{S}_2, \ldots, \mathbf{S}_m)$, we have

$$\mathbf{S}_k^T\mathbf{Z}_k^T\sum_{i=1}^m \mathbf{Z}_i\mathbf{S}_i = \mathbf{S}_k^T\mathbf{Z}_k^T\mathbf{M} \in \mathbb{S}^N\ , \tag{56}$$

where $\mathbf{M} = \sum_{i=1}^{m} \mathbf{Z}_i \mathbf{S}_i$. Every symmetric matrix is diagonalizable [11] so there exists unitary $\mathbf{P}_k$ and diagonal $\mathbf{D}_k$ such that

$$\mathbf{S}_k^T \mathbf{Z}_k^T \mathbf{M} = \mathbf{M}^T \mathbf{Z}_k \mathbf{S}_k = \mathbf{P}_k \mathbf{D}_k \mathbf{P}_k^T \ . \tag{57}$$

Using Lemma S.1.9 we obtain

$$\begin{aligned}
\mathbf{M}^T \mathbf{Z}_k \mathbf{V}\mathbf{V}^T \mathbf{S}_k &= \mathbf{P}_k \mathbf{D}_k \mathbf{P}_k^T && (58)\\
\mathbf{M}^T \mathbf{Z}_k \mathbf{V}\mathbf{V}^T &= \mathbf{P}_k \mathbf{D}_k \mathbf{P}_k^T \mathbf{S}_k^T && (59)\\
\mathbf{S}_k \mathbf{M}^T \mathbf{Z}_k \mathbf{V}\mathbf{V}^T &= \mathbf{S}_k \mathbf{P}_k \mathbf{D}_k \mathbf{P}_k^T \mathbf{S}_k^T && (60)\\
\mathbf{S}_k \mathbf{M}^T \mathbf{Z}_k \mathbf{V}\mathbf{V}^T &= (\mathbf{S}_k \mathbf{P}_k) \mathbf{D}_k (\mathbf{S}_k \mathbf{P}_k)^T \ . && (61)
\end{aligned}$$

Let $\mathbf{V}_\perp \in \mathring{\mathcal{N}}(\mathbf{V})$. With $\mathbf{S}_k \mathbf{M}^T \mathbf{Z}_k \mathbf{V}\mathbf{V}^T$ symmetric and $(\mathbf{S}_k \mathbf{M}^T \mathbf{Z}_k \mathbf{V}\mathbf{V}^T)\mathbf{V}_\perp = \mathbf{0}$, it follows that $\mathbf{S}_k \mathbf{P}_k = [\mathbf{V}\mathbf{A}_k \mid \mathbf{V}_\perp]$ for some $\mathbf{A}_k \in \mathcal{O}(r)$ and $(N - r)$ of the eigenvalues are zero. We may therefore use

$$\mathbf{S}_k = [\mathbf{V}\mathbf{A}_k \mid \mathbf{V}_\perp]\mathbf{P}_k^T \quad , \quad \mathbf{P}_k = [\mathbf{S}_k^T \mathbf{V}\mathbf{A}_k \mid \mathbf{S}_k^T \mathbf{V}_\perp] \quad \text{and} \quad \mathbf{D}_k = \begin{pmatrix} \bar{\mathbf{D}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} , \tag{62}$$

where $\bar{\mathbf{D}}_k$ is $r \times r$ diagonal.

We now wish to show that $\mathbf{M}\mathbf{S}_k^T \mathbf{V}_\perp = \mathbf{0}$ for each k. We first note that

$$\begin{aligned}
\sum_{j=1}^{m} \mathbf{M}^T \mathbf{Z}_j \mathbf{S}_j = \sum_{j=1}^{m} \mathbf{M}^T \mathbf{Z}_j \mathbf{V}\mathbf{V}^T \mathbf{S}_j &= \sum_{j=1}^{m} \mathbf{P}_j \mathbf{D}_j \mathbf{P}_j^T && (63)\\
&= \mathbf{M}^T \sum_{j=1}^{m} \mathbf{Z}_j \mathbf{S}_j = \mathbf{M}^T \mathbf{M} = \sum_{i,j} \mathbf{S}_i^T \mathbf{Z}_i^T \mathbf{Z}_j \mathbf{S}_j \ . && (64)
\end{aligned}$$

We have

$$\begin{aligned}
\mathbf{V}_\perp^T \mathbf{S}_k \mathbf{M}^T \mathbf{M} \mathbf{S}_k^T \mathbf{V}_\perp &= \mathbf{V}_\perp^T \mathbf{S}_k \left(\sum_{j=1}^{m} \mathbf{P}_j \mathbf{D}_j \mathbf{P}_j^T \right) \mathbf{S}_k^T \mathbf{V}_\perp && (65)\\
&= \mathbf{V}_\perp^T \left(\mathbf{V}\mathbf{A}_k \mid \mathbf{V}_\perp \right) \mathbf{P}_k^T \left(\sum_{j=1}^{m} \mathbf{P}_j \mathbf{D}_j \mathbf{P}_j^T \right) \mathbf{P}_k \begin{pmatrix} \mathbf{A}_k^T \mathbf{V}^T \\ \mathbf{V}_\perp^T \end{pmatrix} \mathbf{V}_\perp && (66)\\
&= (\mathbf{0} \mid \mathbf{I}) \left(\sum_{j=1}^{m} (\mathbf{P}_k^T \mathbf{P}_j) \begin{pmatrix} \bar{\mathbf{D}}_j & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} (\mathbf{P}_j^T \mathbf{P}_k) \right) \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} && (67)\\
&= (\mathbf{0} \mid \mathbf{I}) \left(\sum_{j=1}^{m} \begin{pmatrix} * & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \right) \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} && (68)\\
&= (\mathbf{0} \mid \mathbf{I}) \begin{pmatrix} * & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} && (69)\\
&= \mathbf{0} \ . && (70)
\end{aligned}$$

It follows that

$$\|\mathbf{M}\mathbf{S}_k^T \mathbf{V}_\perp\|_{\mathrm{F}}^2 = \mathrm{tr}(\mathbf{V}_\perp^T \mathbf{S}_k \mathbf{M}^T \mathbf{M} \mathbf{S}_k^T \mathbf{V}_\perp) = \mathrm{tr}(\mathbf{0}) = 0 \tag{71}$$

and so necessarily we have $\mathbf{M}\mathbf{S}_k^T \mathbf{V}_\perp = \mathbf{0}$ (a norm separates points). With $\mathbf{S}_1 = \mathbf{I}$ we also have $\mathbf{M}\mathbf{V}_\perp = \mathbf{0}$.

The minimizer $\mathbf{S}_{1:m}$ is fixed and consequently so is $\mathbf{M}$. Let $\mathbf{C} = \frac{1}{m}\mathbf{M}$ be the centroid of the mappings. Corollary S.1.2 tells us that $\mathbf{S}_1 = \mathbf{I}$ is supported by $\mathring{\mathcal{R}}(\mathbf{V})$ so we consider $\mathbf{S}_{2:m}$. For each $k = 2, 3, \ldots, m$ we generate a new $\mathbf{S}_k$ by solving $\arg\min_{\mathbf{Q} \in \mathcal{O}(N)} \|\mathbf{Z}_k \mathbf{Q} - \mathbf{C}\|_{\mathrm{F}}^2$. Lemma S.1.5 guarantees that this new point will not increase the objective and so it will also be a global minimizer.

Starting with $k = 2$ we seek an orthogonal $\mathbf{Q}$ to minimize $\|\mathbf{Z}_2\mathbf{Q} - \mathbf{C}\|_{\mathrm{F}}^2 = \textsc{const} - 2\,\mathrm{tr}(\mathbf{Q}^T\mathbf{Z}_2^T\mathbf{C})$ or maximize $\mathrm{tr}(\mathbf{Q}^T\mathbf{Z}_2^T\mathbf{M})$. The problem at hand is the classical orthogonal Procrustes problem. A solution is found with full left and right singular matrices of $\mathbf{Y} = \mathbf{Z}_2^T\mathbf{M}$. Now, $\mathbf{V}_\perp^T\mathbf{Y} = \mathbf{0}$ because $\mathbf{Z}_2^T = \mathbf{V}\mathbf{V}^T\mathbf{Z}_2^T$ from Lemma S.1.9. Furthermore, $\mathbf{Y}\mathbf{V}_\perp = \mathbf{0}$ because $\mathbf{M}\mathbf{V}_\perp = \mathbf{0}$. Thus, the SVD of $\mathbf{Y}$ admits respective left and right singular matrices $[\mathbf{V}\mathbf{B} \mid \mathbf{V}_\perp]$ and $[\mathbf{V}\mathbf{B}' \mid \mathbf{V}_\perp]$ for some orthogonal $\mathbf{B}$ and $\mathbf{B}'$. It follows that $\mathbf{Q}_* = [\mathbf{V}\mathbf{B} \mid \mathbf{V}_\perp][\mathbf{V}\mathbf{B}' \mid \mathbf{V}_\perp]^T$ is a valid minimizer with which we update $\mathbf{S}_2$. This process is then repeated for $k = 3, \ldots, m$. Finally, we note that

$$\mathbf{V}_\perp^T[\mathbf{V}\mathbf{B} \mid \mathbf{V}_\perp][\mathbf{V}\mathbf{B}' \mid \mathbf{V}_\perp]^T\mathbf{V}_\perp = [\mathbf{0} \mid \mathbf{I}][\mathbf{0} \mid \mathbf{I}]^T = \mathbf{I}\,, \tag{72}$$

so from Lemma S.1.1, the updates produce a global minimizer supported by $\mathring{\mathcal{R}}(\mathbf{V})$. □

S.2 Derivation of (11)

$$\begin{aligned}
\mathbf{R}_i^T\mathbf{\Phi}_i^T\mathbf{\Psi} &= \frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\Big(\big[\mathbf{I} - \mathbf{\Phi}_0^T\mathbf{K}_0^{-1}\mathbf{\Phi}_0 + \mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{G}_i^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{\Phi}_0\big]\mathbf{\Phi}_i^T\mathbf{B}_i \\
&\qquad \mathbf{B}_j\mathbf{\Phi}_j\big[\mathbf{I} - \mathbf{\Phi}_0^T\mathbf{K}_0^{-1}\mathbf{\Phi}_0 + \mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\hat{\mathbf{G}}_j\mathbf{K}_0^{-\frac{1}{2}}\mathbf{\Phi}_0\big]\Big) &(73)\\
&= \textsc{const} + \frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\Big(\mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{G}_i^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{\Phi}_0\mathbf{\Phi}_i^T\mathbf{B}_i\mathbf{B}_j\mathbf{\Phi}_j \\
&\qquad \big[\mathbf{I} - \mathbf{\Phi}_0^T\mathbf{K}_0^{-1}\mathbf{\Phi}_0 + \mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\hat{\mathbf{G}}_j\mathbf{K}_0^{-\frac{1}{2}}\mathbf{\Phi}_0\big]\Big), &(74)
\end{aligned}$$

and so

$$\begin{aligned}
&\mathrm{tr}(\mathbf{R}_i^T\mathbf{\Phi}_i^T\mathbf{\Phi}_0) + \textsc{const} \\
&= \mathrm{tr}\left(\frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\mathbf{G}_i^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{K}_{0i}\mathbf{B}_i\mathbf{B}_j\mathbf{\Phi}_j\left(\mathbf{I} - \mathbf{\Phi}_0^T\mathbf{K}_0^{-1}\mathbf{\Phi}_0 + \mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\hat{\mathbf{G}}_j\mathbf{K}_0^{-\frac{1}{2}}\mathbf{\Phi}_0\right)\mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\right) \\
&= \mathrm{tr}\Big(\mathbf{G}_i^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{K}_{0i}\mathbf{B}_i\Big[\frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\mathbf{B}_j\big(\mathbf{\Phi}_j\mathbf{\Phi}_0^T - \mathbf{\Phi}_j\mathbf{\Phi}_0^T\mathbf{K}_0^{-1}\mathbf{\Phi}_0\mathbf{\Phi}_0^T \\
&\qquad + \mathbf{\Phi}_j\mathbf{\Phi}_0^T\mathbf{K}_0^{-\frac{1}{2}}\hat{\mathbf{G}}_j\mathbf{K}_0^{-\frac{1}{2}}\mathbf{\Phi}_0\mathbf{\Phi}_0^T\big)\Big]\mathbf{K}_0^{-\frac{1}{2}}\Big) \\
&= \mathrm{tr}\Big(\mathbf{G}_i^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{K}_{0i}\mathbf{B}_i\Big[\frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\mathbf{B}_j\big(\mathbf{K}_{j0} - \mathbf{K}_{j0}\mathbf{K}_0^{-1}\mathbf{K}_0 + \mathbf{K}_{j0}\mathbf{K}_0^{-\frac{1}{2}}\hat{\mathbf{G}}_j\mathbf{K}_0^{-\frac{1}{2}}\mathbf{K}_0\big)\Big]\mathbf{K}_0^{-\frac{1}{2}}\Big) \\
&= \mathrm{tr}\left(\mathbf{G}_i^T\mathbf{K}_0^{-\frac{1}{2}}\mathbf{K}_{0i}\mathbf{B}_i\left[\frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\mathbf{B}_j\mathbf{K}_{j0}\mathbf{K}_0^{-\frac{1}{2}}\hat{\mathbf{G}}_j\right]\right) &(75)\\
&= \mathrm{tr}\left(\mathbf{G}_i^T\tilde{\mathbf{B}}_i^T\left[\frac{1}{|\mathcal{A}|}\sum_{j\in\mathcal{A}}\tilde{\mathbf{B}}_j\hat{\mathbf{G}}_j\right]\right) &(76)
\end{aligned}$$

S.3 The Orthogonal Procrustes Problem for $n \geq 2t$

We can use the approach taken in formulating the kernel hyperalignment algorithm for solving $\arg\min_{\mathbf{R}\in\mathcal{O}(n)} \|\mathbf{X}\mathbf{R} - \mathbf{Y}\|_{\mathrm{F}}^2$. Here, $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{t\times n}$ and $n \geq 2t$. We form plane support via $\left[\mathbf{X}^T \mid \mathbf{Y}^T\right]^T \in \mathbb{R}^{2t\times n}$.

Let $\left[\mathbf{X}^T \mid \mathbf{Y}^T\right]^T$ have SVD $\bar{\mathbf{U}}\bar{\mathbf{\Sigma}}\bar{\mathbf{V}}^T$ with $\bar{\mathbf{\Sigma}}$ $r \times r$ diagonal, where r is the rank of the matrix. Impose $\mathbf{R} = \mathbf{I}_n - \bar{\mathbf{V}}(\mathbf{I}_r - \mathbf{G})\bar{\mathbf{V}}^T$ for some $\mathbf{G} \in \mathcal{O}(n)$. It follows that we wish to maximize $\mathrm{tr}(\bar{\mathbf{V}}^T\mathbf{G}^T\bar{\mathbf{V}}\mathbf{X}^T\mathbf{Y}) = \mathrm{tr}(\mathbf{G}^T\bar{\mathbf{V}}^T\mathbf{X}^T\mathbf{Y}\bar{\mathbf{V}})$. Thus, we take the full SVD of $\bar{\mathbf{V}}^T\mathbf{X}^T\mathbf{Y}\bar{\mathbf{V}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$ and set $\mathbf{G}_\star = \tilde{\mathbf{U}}\tilde{\mathbf{V}}^T$.

In total, this solution requires 2 SVDs: one of a $2t\times n$ matrix and one of an $r\times r$ matrix. Respectively, these SVDs cost $O(4t^2 n)$ and $O(r^3)$ operations. By construction $r \leq 2t$, so the total SVD cost in the worst case is $O(4t^2 n + 8t^3)$, which is linear in n.

As a final note, the storage requirement for the dense $\mathbf{R}$ is n^2 entries. However, given the imposed decomposition, the storage reduces to nr for $\bar{\mathbf{V}}$ and r^2 for $\mathbf{G}$, yielding a total of $nr+r^2$. Necessarily, there is always a storage benefit because $nr + r^2 \geq n^2 \Rightarrow n/r \leq (1+\sqrt{5})/2 \approx 1.618$ (Golden Ratio). With $r \leq 2t \leq n$, we have $n/r \geq 2 > (1+\sqrt{5})/2$.

S.4 Kernels used for experiments

Recall that n is the number of voxels.

- Linear

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle \tag{77}$$

- Quadratic

$$\left(\frac{n}{22.5} + \langle \mathbf{x}_1, \mathbf{x}_2 \rangle\right)^2 \tag{78}$$

- Gaussian

$$\exp\left\{\frac{\|\mathbf{x}_1 - \mathbf{x}_2\|_2^2}{22455}\right\} \tag{79}$$

- Sigmoid

$$\tanh\left(\frac{7}{n}\langle \mathbf{x}_1, \mathbf{x}_2 \rangle\right) \tag{80}$$