Supplementary Material: False Discovery and Its Control in Low-rank Estimation

A Appendix

A.1 Proof of Theorem 4 (main paper)

We first prove the basis-dependent bound. For each $\ell = 1, \ldots, B$ and for each $i = 1, \ldots, \dim(T^{\star\perp})$ we have that

$$\begin{aligned}\mathrm{trace}(\mathcal{P}_T \mathcal{P}_{\mathrm{span}(M_i)}) \;=\; & \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\mathrm{span}(M_i)}) + \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_{\mathrm{span}(M_i)}) \\ & + \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_{\mathrm{span}(M_i)}) + \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\mathrm{span}(M_i)}).\end{aligned} \tag{1}$$

The last two terms may be simplified as follows:

$$\begin{aligned}& \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_{\mathrm{span}(M_i)}) + \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\mathrm{span}(M_i)}) \\ =\; & \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} (\mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} - \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_T) \mathcal{P}_{\mathrm{span}(M_i)}) \\ & + \mathrm{trace}((\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_T - \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}) \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\mathrm{span}(M_i)}) \\ =\; & \mathrm{trace}([\mathcal{P}_T, \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}] \times [\mathcal{P}_{\mathrm{span}(M_i)}, \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}]).\end{aligned}$$

The first equality follows by noting that $\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} = \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} = 0$ for each $\ell = 1, \ldots, B$. The second equality follows from the definition of the commutator and the cyclicity of trace. We label the various terms of (1) combined with the above simplification in terms of commutators as follows for each $\ell = 1, \ldots, B$ and $i = 1, \ldots, \dim(T^{\star\perp})$:

$$\begin{aligned} f_{\ell,i} &= \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\mathrm{span}(M_i)}) \\ g_{\ell,i} &= \mathrm{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp} \mathcal{P}_{\mathrm{span}(M_i)}) \\ h_{\ell,i} &= \mathrm{trace}([\mathcal{P}_T, \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}] \times [\mathcal{P}_{\mathrm{span}(M_i)}, \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}]). \end{aligned} \tag{2}$$

Therefore, we have for each $\ell = 1, \ldots, B$ and $i = 1, \ldots, \dim(T^{\star\perp})$ that:

$$\mathrm{trace}(\mathcal{P}_T \mathcal{P}_{\mathrm{span}(M_i)}) = f_{\ell,i} + g_{\ell,i} + h_{\ell,i}.$$

Fix a pair of complementary bags indexed by $\{2j-1, 2j\}$ for some $j \in \{1, \ldots, \frac{B}{2}\}$. For this pair, we have that:

$$\begin{aligned}\mathrm{trace}(\mathcal{P}_T \mathcal{P}_{\mathrm{span}(M_i)}) &= \min\{f_{2j-1,i} + g_{2j-1,i} + h_{2j-1,i}, f_{2j,i} + g_{2j,i} + h_{2j,i}\} \\ &\leq \min\{f_{2j-1,i} + g_{2j-1,i}, f_{2j,i} + g_{2j,i}\} + \max\{h_{2j-1,i}, h_{2j,i}\} \\ &\leq \min\{f_{2j-1,i}, f_{2j,i}\} + g_{2j-1,i} + g_{2j,i} + \max\{h_{2j-1,i}, h_{2j,i}\}.\end{aligned} \tag{3}$$

The first equality holds because the two terms in the minimum are equal. The first inequality holds because $\min\{u_0 + v_0, u_1 + v_1\} \leq \min\{u_0, u_1\} + \max\{v_0, v_1\}$ if $u_0 + v_0 = u_1 + v_1$ (here $u_k = f_{2j-k,i} + g_{2j-k,i}$ and $v_k = h_{2j-k,i}$ for $k = 0, 1$). The second inequality holds because $\min\{u_0 + v_0, u_1 + v_1\} \leq \min\{u_0, u_1\} + v_0 + v_1$ for $v_0, v_1 \geq 0$ (here $u_k = f_{2j-k,i}$ and $v_k = g_{2j-k,i}$ for $k = 0, 1$). The bound (3) holds for all $j = 1, 2, \ldots, B/2$ and for each $i = 1, \ldots, \dim(T^{\star\perp})$. We can thus minimize the upper bounds as follows:

$$\begin{aligned}\operatorname{trace}(\mathcal{P}_T \mathcal{P}_{\operatorname{span}(M_i)}) &\leq \min_{j=1,2,\ldots,B/2} \min\{f_{2j-1,i}, f_{2j,i}\} + g_{2j-1,i} + g_{2j,i} + \max\{h_{2j-1,i}, h_{2j,i}\} \\ &\leq \frac{2}{B} \sum_{j=1}^{B/2} \min\{f_{2j-1,i}, f_{2j,i}\} + g_{2j-1,i} + g_{2j,i} + \max\{h_{2j-1,i}, h_{2j,i}\},\end{aligned}$$

where the second inequality follows from the fact that the minimum over a collection of numbers is bounded above by their average. Since $\operatorname{trace}(\mathcal{P}_T \mathcal{P}_{T^{\star\perp}}) = \sum_{i=1}^{\dim(T^{\star\perp})} \operatorname{trace}(\mathcal{P}_T \mathcal{P}_{\operatorname{span}(M_i)})$, we have the following bound after taking expectations:

$$\begin{aligned}\mathbb{E}\left[\operatorname{trace}(\mathcal{P}_T \mathcal{P}_{T^{\star\perp}})\right] \leq\ & \underbrace{\mathbb{E}\left[\sum_{i=1}^{\dim(T^{\star\perp})} \frac{2}{B} \sum_{j=1}^{B/2} \min\{f_{2j-1,i}, f_{2j,i}\}\right]}_{\text{Term 1}} + \underbrace{\mathbb{E}\left[\sum_{i=1}^{\dim(T^{\star\perp})} \frac{2}{B} \sum_{j=1}^{B/2} (g_{2j-1,i} + g_{2j,i})\right]}_{\text{Term 2}} \\ & + \underbrace{\mathbb{E}\left[\sum_{i=1}^{\dim(T^{\star\perp})} \frac{2}{B} \sum_{j=1}^{B/2} \max\{h_{2j-1,i}, h_{2j,i}\}\right]}_{\text{Term 3}}.\end{aligned}$$

We focus on bounding each term separately. First, considering Term 1, we have for each $\ell = 1, \ldots, B$ and each $i = 1, \ldots, \dim(T^{\star\perp})$ that:

$$\begin{aligned}f_{\ell,i} &= \operatorname{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\operatorname{span}(M_i)}) \\ &= \operatorname{trace}(\mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_{\operatorname{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)} \mathcal{P}_T) \\ &= \|\mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}(M_i)\|_F^2 \\ &\leq \|\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}(M_i)\|_F^2.\end{aligned}$$

Here the second equality follows from the idempotence of projection operators and the cyclicity of trace; the third equality by the definition of the Frobenius norm; and the inequality from the property that projection reduces the Frobenius norm of a matrix. With this relation, we bound Term 1 as follows:

$$\begin{aligned}\text{Term 1} &\leq \mathbb{E}\left[\sum_{i=1}^{\dim(T^{\star\perp})} \frac{2}{B} \sum_{j=1}^{B/2} \min\{\|\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}(M_i)\|_F^2, \|\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}(M_i)\|_F^2\}\right] \\ &\leq \mathbb{E}\left[\sum_{i=1}^{\dim(T^{\star\perp})} \frac{2}{B} \sum_{j=1}^{B/2} \|\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}(M_i)\|_F \|\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}(M_i)\|_F\right] \\ &= \sum_{i=1}^{\dim(T^{\star\perp})} [\mathbb{E}\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M_i)\|_F]^2.\end{aligned}$$

Here the second inequality follows from the property that minimum of two positive quantities is bounded above by the product of their square roots, and the equality follows from $\hat{T}(\mathcal{D}_{2j-1})$ and $\hat{T}(\mathcal{D}_{2j})$ being independent, and $\hat{T}(\mathcal{D}_\ell)$ being identically distributed for all $\ell = 1, 2, \ldots, \ell$. Turning next to Term 2, we have that:

$$\begin{aligned}
\text{Term 2} = 2\ \mathbb{E}\left[\frac{1}{B}\sum_{\ell=1}^{B}\sum_{i=1}^{\dim(T^{\star\perp})}\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\text{span}(M_i)})\right] \\
= 2\ \mathbb{E}\left[\frac{1}{B}\sum_{\ell=1}^{B}\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{T^{\star\perp}})\right] \\
\leq 2\ \mathbb{E}\left[\frac{1}{B}\sum_{\ell=1}^{B}\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp})\right] \\
= 2\ \mathbb{E}\left[\text{trace}(\mathcal{P}_T(\mathcal{I}-\mathcal{P}_{\texttt{avg}})\mathcal{P}_T)\right] \\
\leq 2(1-\alpha)\dim(T).
\end{aligned}$$

Here the second equality follows from $\sum_{i=1}^{\dim(T^{\star\perp})}\mathcal{P}_{\text{span}(M_i)} = \mathcal{P}_{T^{\star\perp}}$; the first inequality follows from the inequality $\text{trace}(AB) \leq \text{trace}(A)\|B\|_2$ for symmetric and positive-semidefinite A; the third equality from the definition of $\mathcal{P}_{\texttt{avg}}$, the idempotence of projection operators, and the cyclicity of trace; and the last inequality from the choice of T. Term 3 is simply taken as is. This concludes the basis-dependent bound.

Next we consider the basis-independent bound. We begin with a decomposition analogous to that of (1) along with the subsequent simplification in terms of commutators for each $\ell = 1, \ldots, B$:

$$\begin{aligned}
\text{trace}\left(\mathcal{P}_T\mathcal{P}_{T^{\star\perp}}\right) = \text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}\mathcal{P}_{T^{\star\perp}}\right) + \text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{T^{\star\perp}}\right) \\
+ \text{trace}\left(\left[\mathcal{P}_T, \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\right] \times \left[\mathcal{P}_{T^{\star\perp}}, \mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}\right]\right).
\end{aligned} \tag{4}$$

The remainder of the proof proceeds in an analogous fashion.

A.2 Proof of Proposition 5 (main paper)

We use the terminology of subsection A.1 above. We prove a bound on $\kappa_{\text{bag}}(\alpha)$ in both the basis-dependent and basis-independent settings based on the following observation for each $j = 1, \ldots, B/2$:

$$\max\left\{\sum_{i=1}^{\dim(T^{\star\perp})} h_{2j-1,i}, \sum_{i=1}^{\dim(T^{\star\perp})} h_{2j,i}\right\} \leq \sum_{i=1}^{\dim(T^{\star\perp})} \max\left\{h_{2j-1,i}, h_{2j,i}\right\}.$$

Taking expectations on both sides and averaging over the collection of complementary pairs of bags indexed by $j = 1, \ldots, B/2$, the left-hand-side corresponds to the basis-independent version of $\kappa_{\text{bag}}(\alpha)$ while the right-hand-side corresponds to the basis-dependent version of $\kappa_{\text{bag}}(\alpha)$. Consequently, it suffices to just bound the right-hand-side. Consider the following sets for each $j = 1, \ldots, B/2$ and each $i = 1, \ldots, \dim(T^{\star\perp})$:

$$\begin{aligned}
\mathcal{S}_j^1 &= \{i \mid h_{2j-1,i} = \max\{h_{2j-1}, h_{2j,i}\}\} \\
\mathcal{S}_j^0 &= \{i \mid h_{2j,i} = \max\{h_{2j-1}, h_{2j,i}\}\}.
\end{aligned}$$

If there are some i such that $h_{2j-1} = h_{2j,i}$, then the corresponding i should be assigned (arbitrarily) to one of $\mathcal{S}_j^0$ or $\mathcal{S}_j^1$, exclusively, so that the sets $\mathcal{S}_j^0, \mathcal{S}_j^1$ partition $\{1, \ldots, \dim(T^{\star\perp})\}$. With this notation, $\kappa_{\text{bag}}(\alpha)$ (basis-dependent or basis-independent) may be bounded as:

$$\kappa_{\text{bag}}(\alpha) \leq \mathbb{E}\left[\frac{2}{B}\sum_{j=1}^{B/2}\left\{\sum_{i\in\mathcal{S}_j^1} h_{2j-1,i} + \sum_{i\in\mathcal{S}_j^0} h_{2j,i}\right\}\right]. \tag{5}$$

We first bound the term $\sum_{i\in\mathcal{S}_j^0} h_{2j,i}$ as follows:

$$\begin{aligned}
\sum_{i\in\mathcal{S}_j^0} h_{2j,i} &= \text{trace}\left([\mathcal{P}_T, \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})^\perp}] \times \left[\sum_{i\in\mathcal{S}_j^0} \mathcal{P}_{\text{span}(M_i)}, \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\right]\right) \\
&\leq \|[\mathcal{P}_T, \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})^\perp}]\|_\star \left\|\left[\sum_{i\in\mathcal{S}_j^0} \mathcal{P}_{\text{span}(M_i)}, \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\right]\right\|_2 \\
&\leq \frac{1}{2}\|[\mathcal{P}_T, \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})^\perp}]\|_\star \\
&\leq \|\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})^\perp}\|_\star \\
&\leq \|\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})^\perp}\|_F\sqrt{\dim(T)}.
\end{aligned}$$

Here the first inequality holds because of the tracial Hölder inequality; the second inequality holds because the spectral norm of the commutator between two projection matrices is bounded above by $\frac{1}{2}$; the third inequality follows from the triangle inequality; and the final inequality follows from $\|A\|_\star \leq \|A\|_F\sqrt{\text{rank}(A)}$. We can similarly bound $\sum_{i\in\mathcal{S}_j^1} h_{2j-1,i}$. Applying this to (5), we obtain:

$$\begin{aligned}
\kappa_{\text{bag}}(\alpha) &\leq \mathbb{E}\left[\frac{2}{B}\sum_{j=1}^{B/2}(\|\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})^\perp}\|_F + \|\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})^\perp}\|_F)\sqrt{\dim(T)}\right] \\
&= 2\ \mathbb{E}\left[\left(\frac{1}{B}\sum_{\ell=1}^{B}\|\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\|_F\right)\sqrt{\dim(T)}\right] \\
&\leq 2\ \mathbb{E}\left[\left(\sqrt{\frac{1}{B}\sum_{\ell=1}^{B}\|\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\|_F^2}\right)\sqrt{\dim(T)}\right] \\
&\leq 2\ \mathbb{E}\left[\left(\sqrt{\frac{1}{B}\sum_{\ell=1}^{B}\text{trace}(\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}\mathcal{P}_T)}\right)\sqrt{\dim(T)}\right] \\
&= 2\ \mathbb{E}\left[\sqrt{\text{trace}(\mathcal{P}_T(\mathcal{I} - \mathcal{P}_{\texttt{avg}})\mathcal{P}_T)}\sqrt{\dim(T)}\right] \\
&\leq 2\sqrt{1-\alpha}\ \mathbb{E}[\dim(T)].
\end{aligned}$$

Here the second inequality follows from concavity of the square root function; and the final two steps follow from the definition of $\mathcal{P}_{\texttt{avg}}$ and the fact that T is a stable tangent space.

Next we conclude that $\mathbb{E}[\dim(T)] \leq \frac{q}{\alpha}$ via the following sequence of inequalities:

$$\mathbb{E}[\dim(T)] \leq \frac{1}{\alpha}\mathbb{E}[\sigma_{\min}(\mathcal{P}_T \mathcal{P}_{\mathtt{avg}} \mathcal{P}_T)] \leq \frac{1}{\alpha}\mathbb{E}[\mathrm{trace}(\mathcal{P}_T \mathcal{P}_{\mathtt{avg}} \mathcal{P}_T)] \leq \frac{1}{\alpha}\mathbb{E}[\mathrm{trace}(\mathcal{P}_{\mathtt{avg}})] = \frac{q}{\alpha}.$$

A.3 Proof of Bound in Remark 4 (main paper)

We employ the notation of f, g, h as presented in (2). Considering the decomposition (1) and noting that the projection operators $\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)}, \mathcal{P}_{\mathrm{span}(M_i)}, \mathcal{P}_T$ all commute with each other in variable selection, we have that $h_{2j-1,i} = h_{2j,i} = 0$ for each $j = 1, \dots, B/2$ and each $i = 1, \dots, \dim(T^{\star\perp})$. Hence, for each $j = 1, 2, \dots, B/2$ and each $i = 1, \dots, \dim(T^{\star\perp})$:

$$\begin{aligned}\mathrm{trace}\left(\mathcal{P}_T \mathcal{P}_{\mathrm{span}(M_i)}\right) &= f_{2j-1,i} + g_{2j-1,i} \\ \mathrm{trace}\left(\mathcal{P}_T \mathcal{P}_{\mathrm{span}(M_i)}\right) &= f_{2j,i} + g_{2j,i}.\end{aligned} \tag{6}$$

Furthermore,

$$\begin{aligned} g_{2j-1,i} &= \mathrm{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})^\perp} \mathcal{P}_T \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})^\perp} \mathcal{P}_{\mathrm{span}(M_i)}\right) \\ &= \mathrm{trace}\left(\mathcal{P}_{T \cap \mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})^\perp}\right). \end{aligned}$$

The second equality holds from commutativity of the projection operators in variable selection. Noticing that $\mathrm{trace}\left(\mathcal{P}_T \mathcal{P}_{\mathrm{span}(M_i)}\right) = \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right)$, we move $g_{2j-1,i}$ and $g_{2j,i}$ to the left side and conclude the relation :

$$\mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\right) = f_{2j-1} \quad ; \quad \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\right) = f_{2j}. \tag{7}$$

Notice $\mathcal{P}_{T\cap\mathrm{span}(M_i)}$ is a diagonal matrix with with all zeros except potentially one nonzero in the diagonal. Hence, $\mathrm{trace}(\mathcal{P}_{T\cap\mathrm{span}(M_i)}) = \{0, 1\}$ and is equal to 0 only if $\mathcal{P}_{T\cap\mathrm{span}(M_i)}$ is an identically zero matrix. Thus, an equivalent reformulation of (7) is:

$$\begin{aligned}\mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right) \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\right) &= f_{2j-1} \\ \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right) \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\right) &= f_{2j}.\end{aligned} \tag{8}$$

Taking the minimum over complementary bags yield:

$$\begin{aligned}\min\{f_{2j-1}, f_{2j}\} &= \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right) \min_k \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-k})}\right) \\ &\geq \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right) \prod_k \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-k})}\right) \\ &\geq \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right) \left\{\sum_k \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\hat{T}(\mathcal{D}_{2j-k})}\right) - 1\right\}.\end{aligned}$$

Here the first inequality follows from the fact that $\min\{a, b\} \geq ab$ for $a, b \in [0, 1]$. The second inequality follows from $ab \geq a + b - 1$ for $a, b \in [0, 1]$. We then bound $\frac{2}{B}\sum_{j=1}^{B/2} \min\{f_{2j-1}, f_{2j}\}$

$$\frac{2}{B}\sum_{j=1}^{B/2} \min\{f_{2j-1}, f_{2j}\} \quad \geq \quad \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)}\right) [2\ \mathrm{trace}\left(\mathcal{P}_{T\cap\mathrm{span}(M_i)} \mathcal{P}_{\mathtt{avg}}\right) - 1].$$

Suppose $\mathcal{P}_{T\cap\text{span}(M_i)}$ is not zero-dimensional. Then, as $T \in \mathcal{T}_\alpha$, we find that:

$$\text{trace}\left(\mathcal{P}_{T\cap\text{span}(M_i)}\right) \leq \frac{1}{2\alpha-1}\frac{2}{B}\sum_{j=1}^{B/2}\min\{f_{2j-1}, f_{2j}\}. \tag{9}$$

If $\mathcal{P}_{T\cap\text{span}(M_i)}$ is zero-dimensional, the bound (9) continues to hold as f_{2j-1}, f_{2j} are non-negative quantities. Via the inequality $\text{trace}(AB) \leq \text{trace}(A)\|B\|_2$ for positive-semidefinite A, we have that $f_{2j-1,i} = \text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\mathcal{P}_T\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\mathcal{P}_{\text{span}(M_i)}) \leq \text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\mathcal{P}_{\text{span}(M_i)})$. We substitute this into (9) to find:

$$\begin{aligned}
\mathbb{E}[\text{trace}(\mathcal{P}_T\mathcal{P}_{T^{\star\perp}})] &= \sum_{i=1}^{\dim(T^{\star\perp})} \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T\cap\text{span}(M_i)}\right)\right] \\
&\leq \sum_{i=1}^{\dim(T^{\star\perp})} \frac{\mathbb{E}\left[\frac{2}{B}\sum_{j=1}^{B/2}\min\{\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\mathcal{P}_{\text{span}(M_i)}), \text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\mathcal{P}_{\text{span}(M_i)})\}\right]}{2\alpha-1} \\
&= \sum_{i=1}^{\dim(T^{\star\perp})} \frac{\mathbb{E}\left[\frac{2}{B}\sum_{j=1}^{B/2}\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\mathcal{P}_{\text{span}(M_i)})\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\mathcal{P}_{\text{span}(M_i)})\right]}{2\alpha-1} \\
&= \sum_{i=1}^{\dim(T^{\star\perp})} \frac{\mathbb{E}[\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M_i)})]^2}{2\alpha-1}.
\end{aligned}$$

Here the second equality follows from $\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j-1})}\mathcal{P}_{\text{span}(M_i)}) \in \{0,1\}$ and $\text{trace}(\mathcal{P}_{\hat{T}(\mathcal{D}_{2j})}\mathcal{P}_{\text{span}(M_i)}) \in \{0,1\}$; the final equality holds from $\hat{T}(\mathcal{D}_{2j-1})$ and $\hat{T}(\mathcal{D}_{2j})$ being independent and that $\hat{T}(\mathcal{D}_\ell)$ is identically distributed for all $\ell = 1, 2, \ldots, B$.

A.4 When are Assumptions 1 and 2 in (3.6) Satisfied?

Are there reasonable estimators and models in the low-rank setting that satisfy Assumptions 1 and 2 in (3.7) (main paper)? This section aims to address this question.

Assumption 1 is rather benign. Specifically, fix any $k \leq \min\{p_1, p_2\}$. Let $U \in \mathbb{R}^{p_1\times k}$ and $V \in \mathbb{R}^{p_2\times k}$ be drawn respectively from a Haar measure on the Stiefel Manifold. Then it is straightforward to check that the tangent space $\hat{T} = T(\text{span}(U), \text{span}(V))$ satisfies the following condition:

$$\frac{\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T^{\star\perp}}\mathcal{P}_{\hat{T}}\right)\right]}{\dim(T^{\star\perp})} = \frac{\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T^\star}\mathcal{P}_{\hat{T}}\right)\right]}{\dim(T^\star)}.$$

In other words, the case of equality in Assumption 1 is satisfied if the row and column space estimates are drawn uniformly at random as above, and Assumption 1 merely requires that the low-rank estimator under consideration is better than such a procedure which makes no use of any observations.

Assumption 2 is more stringent, although it is fulfilled in some natural classes of models / estimators. In particular, this assumption is satisfied when the estimator as well as the data generation process are both invariant under orthogonal conjugation. Consider for example:

- Linear regression with Gaussian functionals: consider the linear matrix regression setting where we obtain n linear measurements of $L^\star$ in the form $y_i = \langle \mathcal{A}_i, L^\star \rangle + \epsilon_i$, with each $\mathcal{A}_i \in \mathbb{R}^{p_1 \times p_2}$ consisting of i.i.d. standard Gaussian entries (and the $\mathcal{A}_i$'s being independent of each other) and $\epsilon \in \mathbb{R}^n$ being a standard Gaussian vector. Consider estimators of the form:

$$\hat{L} = \arg\min_{L} \quad \textstyle\sum_{i=1}^n (y_i - \langle \mathcal{A}_i, L \rangle)^2 + \lambda\, \mathcal{R}(L),$$

  which includes as special cases a convex approach with $\mathcal{R}(L) = \|L\|_\star$ as well as a non-convex approach (solved via alternating least squares) with $\mathcal{R}(L) = \|U\|_F^2 + \|V\|_F^2$ with $L = UV'$ corresponding to the estimator (4.2) (main paper).

- Matrix denoising: suppose we are given n observations of $L^\star$ of the form $Y_i = L^\star + \epsilon_i$. Here ϵ_i is a random matrix with i.i.d. standard Gaussian entries (and the ϵ_i's are independent of each other). Consider any spectral estimator (such as soft thresholding or hard thresholding of the singular values) applied to $\bar{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$ to estimate $L^\star$.

In Section A.10, we provide a PCA model and a corresponding estimator that would satisfy a version of Assumption 2 suitable for subspace estimation problems.

A.5 Proof of Proposition 6 (main paper)

Recall that $F = \sum_{i=1}^{\dim(T^{\star\perp})} \mathbb{E}[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M_i)\|_F]^2$ in the basis-dependent bound. Consider a collection of rank-1 basis elements $\{M_i\}_{i=1}^{\dim(T^{\star\perp})}$. By Assumption 2, $\mathbb{E}[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M_i)\|_F]^2 = \mathbb{E}[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F]^2$ for any fixed rank-1 matrix $M \in T^{\star\perp}$ with $\|M\|_F = 1$. Letting $\delta_1 = \mathbb{E}[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F]$, we thus have $F = \dim(T^{\star\perp})\delta_1^2$. Define the quantity $\delta_2 = \mathbb{E}[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F^2]$. Then, F can be bounded in terms of δ_2 and $|\delta_1 - \delta_2|$ as follows:

$$\begin{aligned} F &= \dim(T^{\star\perp})\delta_1^2 \\ &= \dim(T^{\star\perp})\{\delta_2^2 + (\delta_1 - \delta_2)^2 + 2\delta_2(\delta_1 - \delta_2)\} \\ &\leq \dim(T^{\star\perp})\{\delta_2^2 + (\delta_1 - \delta_2)^2 + 2\delta_2|\delta_1 - \delta_2|\}. \end{aligned} \tag{10}$$

We focus on bounding δ_2 and $|\delta_1 - \delta_2|$. To bound δ_2, note that:

$$\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{T^{\star\perp}}\right)\right] + \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{T^\star}\right)\right] = \mathbb{E}[\dim(\hat{T}(\mathcal{D}(n/2)))] = q.$$

Employing "better than random guessing" Assumption 1, we find that:

$$\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{T^{\star\perp}}\right)\right]\left(1 + \frac{\dim(T^\star)}{\dim(T^{\star\perp})}\right) \leq q.$$

Since $\dim(T^\star) + \dim(T^{\star\perp}) = p_1 p_2$, we find

$$\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{T^{\star\perp}}\right)\right] \leq \frac{q}{p_1 p_2}\dim(T^{\star\perp}). \tag{11}$$

We now express the left-hand side of (11) in terms of δ_2. By Assumption 2, we have that $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{T^{\star\perp}}\right)\right] = \dim(T^{\star\perp})\delta_2$. Combining this with (11), we obtain the bound:

$$\delta_2 \leq \frac{q}{p_1 p_2}. \tag{12}$$

Now we focus on bounding $|\delta_1 - \delta_2|$. We proceed by bounding $\delta_1 - \delta_2$ and $\delta_2 - \delta_1$ by the same quantity. In particular, we show that $\delta_1 - \delta_2 \leq \kappa_{\text{indiv}}$ where $\kappa_{\text{indiv}} = \mathbb{E}\left\|\left[\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}, \mathcal{P}_{\text{span}(M)}\right]\right\|_F$:

$$\begin{aligned}
\delta_1 - \delta_2 &= \mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F\right] - \mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F^2\right] \\
&\overset{(a)}{=}\mathbb{E}\left[\left\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\right\|_F\right] - \mathbb{E}\left[\left\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\right\|_F^2\right] \\
&\overset{(b)}{=}\mathbb{E}\left[\left\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\right\|_F\right] - \mathbb{E}\left[\left\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\right\|_F\right] \\
&\overset{(c)}{=}\mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}]\|_F\right] \\
&\quad - \mathbb{E}\left[\left\|\mathcal{P}_{\text{span}(M)}\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))} + \left[\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}, \mathcal{P}_{\text{span}(M)}\right]\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\right\|_F\right] \\
&\overset{(d)}{\leq}\mathbb{E}\left\|\left[\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}, \mathcal{P}_{\text{span}(M)}\right]\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\right\|_F \overset{(e)}{\leq}\mathbb{E}\left\|\left[\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}, \mathcal{P}_{\text{span}(M)}\right]\right\|_F = \kappa_{\text{indiv}}.
\end{aligned}$$

Here $\overset{(a)}{=}$ follows from the property that $\mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F\right] = \mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\|_F\right]$ and that $\mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}(M)\|_F^2\right] = \mathbb{E}\left[\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\|_F^2\right]$; $\overset{(b)}{=}$ follows from noting that $\mathcal{P}_{\text{span}(M)}$ has rank-1 by construction so that $\left\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\right\|_F = \text{trace}\left(\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\right) = \left\|\mathcal{P}_{\hat{T}(\mathcal{D}(n/2))}\mathcal{P}_{\text{span}(M)}\right\|_F^2$; $\overset{(c)}{=}$ follows from the definition of a commutator; $\overset{(d)}{\leq}$ follows from reverse triangle inequality; and $\overset{(e)}{\leq}$ follows from the following reasoning for matrices A, B where B is a projection matrix: $\|AB\|_F = \sqrt{\|AB\|_F^2} = \sqrt{\text{trace}(A'AB)} \leq \sqrt{\text{trace}(A'A)\|B\|_2} \leq \sqrt{\text{trace}(A'A)} = \|A\|_F$.

Similar logic shows that $\delta_2 - \delta_1 \leq \kappa_{\text{indiv}}$ which leads to the conclusion that $|\delta_1 - \delta_2| \leq \kappa_{\text{indiv}}$. Plugging in the bounds for δ_2 and $|\delta_1 - \delta_2|$ into (10), we find that:

$$F \leq \frac{q^2}{p_1 p_2} + \dim(T^{\star\perp})\kappa_{\text{indiv}}^2 + 2q\kappa_{\text{indiv}},$$

as desired.

A.6 Goodness of the Data-Driven Heuristic in Remark 5

Recall that we chose $M = uv'$, where u, v are selected to be the smallest singular vectors associated with $\mathcal{P}_{\mathsf{avg}}^{\mathcal{C}}$ and $\mathcal{P}_{\mathsf{avg}}^{\mathcal{R}}$, respectively. Notice that $\mathcal{P}_{T^{\star\perp}}(M)$ will be of rank less than or equal to 1 since $\mathcal{P}_{T^{\star\perp}}(M) = \mathcal{P}_{\mathcal{C}^{\star\perp}}M\mathcal{P}_{\mathcal{R}^{\star\perp}}$. Hence, the cosine of the largest principal angle between $T^{\star\perp}$ and M, given by $\|\mathcal{P}_{T^{\star\perp}}\mathcal{P}_{\text{span}(M)}\|_F$, will be achieved between the direction spanned by M and a rank-1 direction in $T^{\star\perp}$. As such, we next prove that if the estimator has good power, $\|\mathcal{P}_{T^{\star\perp}}\mathcal{P}_{\text{span}(M)}\|_F$ will be close to 1.

Lemma 1. *Let* $\tau := \mathbb{E}[\min\{\sigma_{\min}(\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}(n/2))}\mathcal{P}_{\mathcal{C}^\star}), \sigma_{\min}(\mathcal{P}_{\mathcal{R}^\star}\mathcal{P}_{\hat{\mathcal{R}}(\mathcal{D}(n/2))}\mathcal{P}_{\mathcal{R}^\star})\}]$ *and* $\delta := \mathbb{E}[\max\{\sigma_{\min}(\mathcal{P}^{\mathcal{C}}_{avg}), \sigma_{\min}(\mathcal{P}^{\mathcal{R}}_{avg})\}]$. *Then, the expected cosine of the principal angle between the data-driven* M *and* $T^{\star\perp}$ *is lower-bounded by:*

$$\mathbb{E}\left[\|\mathcal{P}_{T^{\star\perp}}\mathcal{P}_{\text{span}(M)}\|_F^2\right] \geq 2\tau - 1 - 2(\delta + \sqrt{\delta}).$$

Evidently, when the estimator has good power, i.e. τ is close to 1, and the expected smallest singular values δ is close to 0, the data-driven approach produces an M that is close to $T^{\star\perp}$. We next prove this lemma.

Proof. Notice that:

$$\begin{aligned}
\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T^{\star\perp}}\mathcal{P}_{\text{span}(M)}\right)\right] &= \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{C}^{\star\perp}}\mathcal{P}_{\text{span}(u)}\right)\text{trace}\left(\mathcal{P}_{\mathcal{R}^{\star\perp}}\mathcal{P}_{\text{span}(v)}\right)\right] \\
&\geq \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{C}^{\star\perp}}\mathcal{P}_{\text{span}(u)}\right)\right] + \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{R}^{\star\perp}}\mathcal{P}_{\text{span}(v)}\right)\right] - 1 \\
&= 1 - \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\text{span}(u)}\right)\right] - \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{R}^\star}\mathcal{P}_{\text{span}(v)}\right)\right], \qquad (13)
\end{aligned}$$

where the first equality is due to the property that $\mathcal{P}_{T^{\star\perp}} = \mathcal{C}^{\star\perp} \otimes \mathcal{R}^{\star\perp}$ and the inequality is due to the property $ab \geq a + b - 1$ for $a, b \in [0,1]$. This decomposition implies that upper bounds for $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\text{span}(u)}\right)\right]$ and $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{R}^\star}\mathcal{P}_{\text{span}(v)}\right)\right]$ yield a lower-bound for $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T^{\star\perp}}\mathcal{P}_{\text{span}(M)}\right)\right]$. Proceeding with upper-bounding $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\text{span}(u)}\right)\right]$, we consider the following decomposition:

$$\begin{aligned}
\text{trace}\left(\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\text{span}(u)}\right) &= \text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\text{span}(u)}\right) + \text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\text{span}(u)}\right) \\
&+ \text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\text{span}(u)}\right) + \text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\text{span}(u)}\right) \\
&\leq \left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\right\|_2 + \text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\text{span}(u)}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\right) + \left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\text{span}(u)}\right\|_\star,
\end{aligned}$$

where the inequality is due to $\text{trace}(AB) \leq \text{trace}(A)\|B\|_2$ for $A \succeq 0$, $\text{trace}(AB) \leq \|A\|_\star\|B\|_2$, the idempotence of projection operators and that $\|[\mathcal{P}_{T_1}, \mathcal{P}_{T_2}]\|_2 \leq \frac{1}{2}$ for any two subspaces T_1 and T_2.

Since the choice of ℓ was arbitrary, we minimize over the entire collection:

$$\begin{aligned}
\mathbb{E}\left[\operatorname{trace}\left(\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\operatorname{span}(u)}\right)\right] &\leq \mathbb{E}\Bigg[\min_{\ell=1,2,\dots,B}\left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\right\|_2 + \operatorname{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\operatorname{span}(u)}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\right) \\
&+ \left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\operatorname{span}(u)}\right\|_\star\Bigg] \\
&\overset{(a)}{\leq} \frac{1}{B}\sum_{\ell=1}^{B}\mathbb{E}\left[\left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\right\|_2\right] + \mathbb{E}\left[\operatorname{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\operatorname{span}(u)}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\right)\right] \\
&+ \mathbb{E}\left[\left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)}\mathcal{P}_{\operatorname{span}(u)}\right\|_\star\right] \\
&\overset{(b)}{\leq} \frac{1}{B}\sum_{\ell=1}^{B}\mathbb{E}\left[\left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\right\|_2\right] + \mathbb{E}\left[\operatorname{trace}\left(\mathcal{P}_{\operatorname{span}(u)}\mathcal{P}^{\mathcal{C}}_{\texttt{avg}}\mathcal{P}_{\operatorname{span}(u)}\right)\right] \\
&+ \mathbb{E}\left[\sqrt{\operatorname{trace}\left(\mathcal{P}_{\operatorname{span}(u)}\mathcal{P}^{\mathcal{C}}_{\texttt{avg}}\mathcal{P}_{\operatorname{span}(u)}\right)}\right] \\
&\overset{(c)}{\leq} \frac{1}{B}\sum_{\ell=1}^{B}\mathbb{E}\left[\left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\star}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\right\|_2\right] + \delta + \sqrt{\delta} \\
&\overset{(d)}{=} 1 - \tau + \delta + \sqrt{\delta}.
\end{aligned}$$

Here $\overset{(a)}{\leq}$ follows from the fact that minimum over a collection is bounded by their average; $\overset{(b)}{\leq}$ follows from $\|A\|_\star \leq \|A\|_F \operatorname{rank}(A)$ and the concavity of square root function; and $\overset{(c)}{\leq}$ follows from the fact that u is selected to be the smallest singular vector of $\mathcal{P}^{\mathcal{C}}_{\texttt{avg}}$, concavity of square function and Jensen's inequality, and the definition of δ, and $\overset{(d)}{=}$ follows from the fact that $\|\mathcal{P}_{T_1^\perp}\mathcal{P}_{T_2}\mathcal{P}_{T_1^\perp}\|_2 = 1 - \sigma_{\min}(\mathcal{P}_{T_2}\mathcal{P}_{T_1}\mathcal{P}_{T_2})$ and that $\hat{\mathcal{C}}(\mathcal{D}_\ell)$ is identically distributed for all ℓ. Repeating the same steps for the row-space and combining with (13) gives the desired result. □

A.7 Sensitivity of Subspace Stability Selection to α

The tuning parameter $\alpha \in [0, 1]$ plays an important role in how much discovery is made by subspace stability selection. In our experience, the output of subspace stability selection (which selects a stable tangent space) is rather robust to α in moderate to high SNR settings. As a result, in all our experiments, we select α to equal 0.70.

To more systematically explore the sensitivity of the subspace stability selection algorithm to the choice of α, we consider the following matrix completion setup where $L^\star \in \mathbb{R}^{p\times p}$ with $p = 100$, rank of $L^\star$ in the set $\{1, 3, 5\}$, and row/column spaces chosen uniformly at random from the Steifel manifold. We select a fraction 7/10 of the total entries uniformly at random as the observation set Ω so that $|\Omega| = 7p^2/10$. These observations are corrupted with Gaussian noise with variance selected so that the SNR is one of the values $\{0.5, 0.8, 2\}$, for a total number of nine problem instances (three different noise levels and three different ranks). We use these observations as input to the estimator (4.1) (main paper), with λ selected based on holdout validation on a $n_{\text{test}} = 7/20p^2$ validation set. We fix $B = 100$ and vary the choice of in Algorithm 1 (main paper) over the values in the set $\alpha_{\text{set}} = \{0.6, 0.625, 0.65, 0.675, 0.7, 0.725, 0.75, 0.775, 0.8\}$. For each α, we obtain an associated stable tangent space $T_{\mathtt{S3}(\alpha)}$. Figure 1 demonstrates the variation in the normalized false discovery $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T_{\mathtt{S3}(\alpha)}}\mathcal{P}_{T^{\star\perp}}\right)\right]/\dim(T^{\star\perp})$ and normalized power $\mathbb{E}\left[\text{trace}\left(\mathcal{P}_{T_{\mathtt{S3}(\alpha)}}\mathcal{P}_{T^{\star}}\right)\right]/\dim(T^{\star})$ as a function of α. We notice that for SNR $= 2$, both the false discovery and power are very stable with respect to α for all ranks. Even for a lower value of SNR $= 0.8$, the normalized false discovery and power remain stable to changes in α for small ranks, but are less stable for larger ranks. In summary, this experiment indicates that the subspace stability selection algorithm tends to be robust to perturbations of α for moderate-to-high SNR regimes and small ranks.

We note that the choice of α can also be guided by our theoretical results. In particular, in cases where the signal strength is strong so that the commutator terms are small (see the theoretical statements in Section 3.2 (main paper)), we recommend selecting a large α to maximize power while controlling for false discoveries.

A.8 Proof of Proposition 7 (main paper)

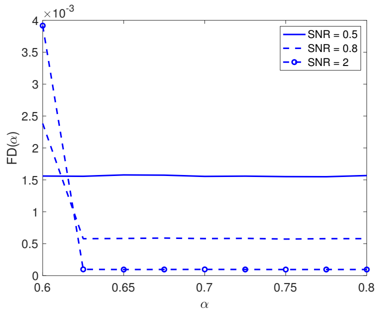
Let T be a tangent space produced by the modified algorithm with associated column and row spaces $(\mathcal{C}, \mathcal{R})$. We proceed by obtaining an upper bound on $\|\mathcal{P}_T(\mathcal{I} - \mathcal{P}_{\mathsf{avg}})\mathcal{P}_T\|_2$, which gives a lower bound on $\sigma_{\min}(\mathcal{P}_T\mathcal{P}_{\mathsf{avg}}\mathcal{P}_T)$:

$$
\begin{aligned}
\|\mathcal{P}_T(\mathcal{I} - \mathcal{P}_{\mathsf{avg}})\mathcal{P}_T\|_2 &= \max_{M\in T, \|M\|_F=1} \frac{1}{B}\text{trace}\Big(\sum_{\ell=1}^{B} M'\mathcal{P}_{\hat{T}(\mathcal{D}_\ell)^\perp}(M)\Big) \\
&\overset{(a)}{=} \max_{M\in T, \|M\|_F=1} \frac{1}{B}\sum_{\ell=1}^{B} \|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp} M \mathcal{P}_{\hat{\mathcal{R}}(\mathcal{D}_\ell)^\perp}\|_F^2 \\
&\overset{(b)}{\leq} \max_{M\in T, \|M\|_F=1} \frac{2}{B}\sum_{\ell=1}^{B} \|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}} M \mathcal{P}_{\hat{\mathcal{R}}(\mathcal{D}_\ell)^\perp}\|_F^2 \\
&+ \frac{2}{B}\sum_{\ell=1}^{B} \|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}^\perp} M \mathcal{P}_{\mathcal{R}}\mathcal{P}_{\hat{\mathcal{R}}(\mathcal{D}_\ell)^\perp}\|_F^2 \\
&\overset{(c)}{\leq} \max_{M\in T, \|M\|_F=1} \frac{2}{B}\sum_{\ell=1}^{B} \|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{C}} M\|_F^2 + \frac{2}{B}\sum_{\ell=1}^{B} \|\mathcal{P}_{\hat{\mathcal{R}}(\mathcal{D}_\ell)^\perp}\mathcal{P}_{\mathcal{R}} M'\|_F^2 \\
&= \max_{M\in T, \|M\|_F=1} 2\ \text{trace}(\mathcal{P}_{\mathcal{C}}(\mathcal{I} - \mathcal{P}_{\mathsf{avg}})\mathcal{P}_{\mathcal{C}} M M') + 2\ \text{trace}(\mathcal{P}_{\mathcal{R}}(\mathcal{I} - \mathcal{P}_{\mathsf{avg}})\mathcal{P}_{\mathcal{R}} M' M) \\
&\leq 2\ \|\mathcal{P}_{\mathcal{C}}(\mathcal{I} - \mathcal{P}_{\mathsf{avg}})\mathcal{P}_{\mathcal{C}}\|_2 + 2\ \|\mathcal{P}_{\mathcal{R}}(\mathcal{I} - \mathcal{P}_{\mathsf{avg}})\mathcal{P}_{\mathcal{R}}\|_2 \leq 4(1-\alpha).
\end{aligned}
$$

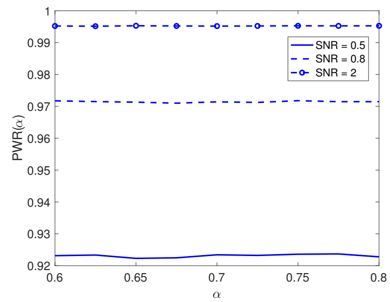
Here (a) follows from the cyclicity of the trace functional and the idempotence of projection maps; (b) from the fact that $M \in T$ implies that $M = \mathcal{P}_{\mathcal{C}} M + \mathcal{P}_{\mathcal{C}^\perp} M \mathcal{P}_{\mathcal{R}}$ and the elementary inequality $(a+b)^2 \leq 2a^2 + 2b^2$; and (c) from the property $\|A\mathcal{P}\|_F \leq \|A\|_F$ for any projection matrix $\mathcal{P}$.

A.9 Tangent Spaces for Column-Space Estimation

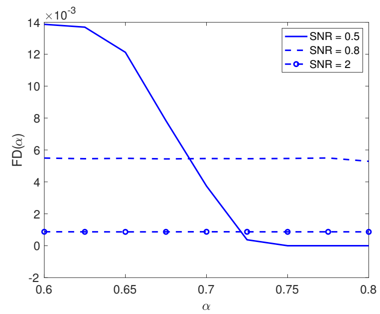
In certain domains such as hyperspectral imaging, one only requires estimates of the column-space of a low-rank matrix, and we seek an appropriate tangent space that represents the discoveries in this context.



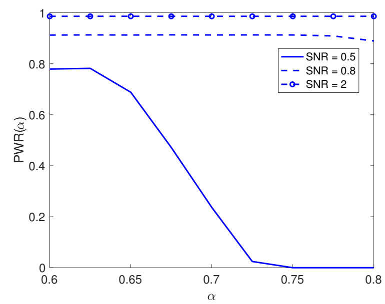
(a) FD normalized: rank = 1



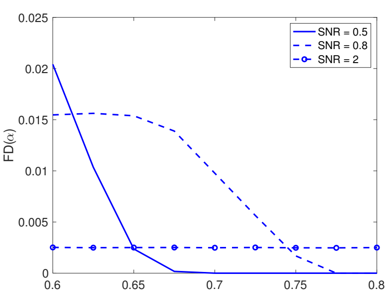
(b) normalized PWR : rank = 1



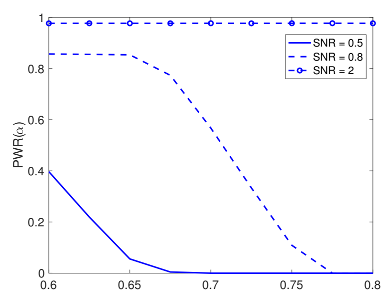
(c) normalized FD: rank = 3



(d) normalized PWR: rank = 3



(e) normalized FD: rank = 5



(f) normalized PWR: rank = 5

Figure 1: Variation in false discovery $\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{T_{S3}(\alpha)}\mathcal{P}_{T^{\star\perp}}\right)\right]/\mathrm{dim}(T^{\star\perp})$ *and power* $\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{T_{S3}(\alpha)}\mathcal{P}_{T^{\star}}\right)\right]/\mathrm{dim}(T^{\star})$ *as a function of* α *for different SNR and rank regimes.*

We begin by considering the tangent space with respect to the determinantal variety $\mathcal{V}(r) \subset \mathbb{R}^{p_1 \times p_2}$ at a rank-r matrix $L = UV' \in \mathbb{R}^{p_1 \times p_2}$ with column/row spaces $(\mathcal{C}, \mathcal{R})$. To compute this space, we consider differences of the form $(U + \Delta_1)(V + \Delta_2)' - UV' = \Delta_1 V' + U\Delta_2' + \Delta_1 \Delta_2' \approx \Delta_1 V' + U\Delta_2'$ for $\Delta_1 \in \mathbb{R}^{p_1 \times r}, \Delta_2 \in \mathbb{R}^{p_2 \times r}$ small. However, such elements involve attributes of the neighborhood of L that do not concern the estimation of an accurate column space, and therefore we must *quotient* out the irrelevant directions. Specifically, the directions consisting of column-space components in $\mathcal{C}^\perp$ are not relevant to the accurate estimation of $\mathcal{C}$. The matrices in $\mathcal{V}(r)$ that lie in a neighborhood around L with deviations in the column-space purely in directions in $\mathcal{C}^\perp$ are given by $L + \mathcal{P}_{\mathcal{C}^\perp}\Delta$ for $\Delta \in \mathbb{R}^{p_1 \times p_2}$. Therefore, we consider the following *equivalence class* associated to each rank-r matrix $L \in \mathcal{V}(r)$:

$$[L] = \{L + \mathcal{P}_{\mathcal{C}^\perp}\Delta\mathcal{P}_{\mathcal{R}} \mid \Delta \in \mathbb{R}^{p_1 \times p_2}\}. \tag{14}$$

The tangent space at L with respect to the *quotient manifold* $\mathcal{V}(r)\backslash[L]$ then signifies the discoveries of interest for column-space estimation. The tangent spaces at L with respect to the equivalence class $[L]$ and with respect to the quotient manifold $\mathcal{V}(r)\backslash[L]$ form complementary subspaces of the tangent space at L with respect to $\mathcal{V}(r)$, and these are known respectively as the *vertical space* and the *horizontal space*. One can check that the vertical space is given by $T_{\text{vertical}} = \{\mathcal{P}_{\mathcal{C}^\perp}\Delta\mathcal{P}_{\mathcal{R}} \mid \Delta \in \mathbb{R}^{p_1 \times p_2}\}$ while the horizontal space is given by $T_{\text{horizontal}} = \{\mathcal{P}_{\mathcal{C}}\Delta \mid \Delta \in \mathbb{R}^{p_1 \times p_2}\}$ so that $T(\hat{\mathcal{C}}, \hat{\mathcal{R}}) = T_{\text{vertical}} \oplus T_{\text{horizontal}}$. Our tangent space of interest is thus the subspace $T_{\text{horizontal}}$, which is solely a function of the column space $\mathcal{C}$.

Observing that $\mathcal{P}_{T_{\text{horizontal}}} = \mathcal{P}_{\mathcal{C}} \otimes \mathcal{I}$ and $\mathcal{P}_{T^\perp_{\text{horizontal}}} = \mathcal{P}_{\mathcal{C}^\perp} \otimes \mathcal{I}$, the expected false discovery, power, and false discovery rate in the context of column-space estimation associated to an estimator $\hat{C}$ are defined as:

$$\begin{aligned} \text{FD} = &\quad \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}}\mathcal{P}_{\mathcal{C}^{\star\perp}}\right)\right] \\ \text{PW} = &\quad \mathbb{E}\left[\text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}}\mathcal{P}_{\mathcal{C}^{\star}}\right)\right] \\ \text{FDR} = &\quad \mathbb{E}\left[\frac{\text{trace}\left(\mathcal{P}_{\hat{\mathcal{C}}}\mathcal{P}_{\mathcal{C}^{\star\perp}}\right)}{\dim(\hat{\mathcal{C}})}\right]. \end{aligned} \tag{15}$$

A.10 False Discovery Guarantees for Column-Space Estimation

In this section, we provide false discovery control guarantees of subspace stability selection for column-space estimation problems. Suppose there exists a population column-space $\mathcal{C}^\star \in \mathbb{R}^{p_1}$, and we are given i.i.d observations from a model parameterized by $\mathcal{C}^\star$. Let $\hat{\mathcal{C}}$ be a subspace estimator that operates on samples drawn from the model parameterized by $\mathcal{C}^\star$. Let $\mathcal{D}(n)$ denote a dataset consisting of n i.i.d observations from these models; we assume n is even and that we are given B subsamples $\{\mathcal{D}_\ell\}_{i=1}^B$ via complementary partitions of $\mathcal{D}(n)$.

We omit the proof of each of these statements as their proof is similar in spirit to those from the main paper.

Theorem 1 (False Discovery Control of Subspace Stability Selection)**.** *Consider the setup described above. Let $\hat{C}(\mathcal{D}_\ell)$ denote the subspace estimates obtained from each of the subsamples, and let $\mathcal{P}^{\mathcal{C}}_{avg}$ denote the associated average projection operator computed via (3.2) (main paper). Fix any $\alpha \in (0, 1)$ and let $\mathcal{C}$ denote any selection of an element of the associated set $\mathcal{T}_\alpha$ of stable tangent* *spaces. Then for any fixed collection of orthonormal basis elements* $\{M_i\}_{i=1}^{\dim(\mathcal{C}^{\star\perp})}$ *of* $\mathcal{C}^{\star\perp}$

$$\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{\mathcal{C}}\mathcal{P}_{\mathcal{C}^{\star\perp}}\right)\right] \leq F + \kappa_{bag}(\alpha) + 2(1-\alpha)\mathbb{E}[\dim(\mathcal{C})]. \tag{16}$$

For basis-dependent bound, $F \sum_{i=1}^{\dim(\mathcal{C}^{\star\perp})} \mathbb{E}\left[\left\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}(n/2))}(M_i)\right\|_F\right]^2$ *and* $\kappa_{bag}(\alpha) = \sum_{i=1}^{dim(\mathcal{C}^{\star\perp})} \frac{2}{B}\sum_{j=1}^{B/2}$ $\mathbb{E}[\max_{k\in\{0,1\}} \mathrm{trace}([\mathcal{P}_{\mathcal{C}}, \mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_{2j-k})^\perp}] \times [\mathcal{P}_{\mathrm{span}(M_i)}, \mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_{2j-k})}])]$*, whereas for a basis-independent bound,* $F \leq \mathbb{E}[\mathrm{trace}(\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}(n/2))}\mathcal{P}_{\mathcal{C}^{\star\perp}})^{1/2}]^2$ *and* $\kappa_{bag}(\alpha) = \frac{2}{B}\sum_{j=1}^{B/2}\mathbb{E}[\max_{k\in\{0,1\}} \mathrm{trace}([\mathcal{P}_{\mathcal{C}}, \mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_{2j-k})^\perp}] \times [\mathcal{P}_{\mathcal{C}^{\star\perp}}, \mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}_{2j-k})}])]$*. Here the expectation is with respect to randomness in the observations. The set* $\mathcal{D}(n/2)$ *denotes a collection of* $n/2$ *i.i.d. observations drawn from the model parametrized by* $\mathcal{C}^\star$*.*

The next proposition provides an upper bound for $\kappa_{\mathrm{bag}}(\alpha)$ and also provides a bag independent bound:

Proposition 1 (Bounding κ_{bag} and a Bag Independent Result)**.** *Consider the setup of Theorem 1. Then the following bound holds for both the basis-independent and basis-dependent* $\kappa_{bag}(\alpha)$*:* $\kappa_{bag}(\alpha) \leq 2\sqrt{1-\alpha}\mathbb{E}[dim(\mathcal{C})]$*. Furthermore, letting the average number of discoveries from* $n/2$ *observations be denoted by* $q := \mathbb{E}[\dim(\hat{\mathcal{C}}(\mathcal{D}(n/2)))]$*, we also have that* $\mathbb{E}[dim(\mathcal{C})] \leq \frac{q}{\alpha}$*. Thus, we obtain the following false discovery bound for any* $B \geq 2$*:*

$$\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{\mathcal{C}}\mathcal{P}_{\mathcal{C}^{\star\perp}}\right)\right] \quad\leq\quad F + \frac{2q}{\alpha}(1-\alpha+\sqrt{1-\alpha}). \tag{17}$$

Finally, we obtained a refined bound under "better than random guessing" and exchangeability assumptions:

$$\text{Assumption 3: } \frac{\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{\mathcal{C}^{\star\perp}}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}(n/2))}\right)\right]}{\dim(\mathcal{C}^{\star\perp})} \leq \frac{\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{\mathcal{C}^{\star}}\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}(n/2))}\right)\right]}{\dim(\mathcal{C}^{\star})} \tag{18}$$

Assumption 4: The distribution of $\|\mathcal{P}_{\hat{\mathcal{C}}(\mathcal{D}(n/2))}(M)\|_F$ is the same for all $M \in \mathcal{C}^{\star\perp}$, $\|M\|_F = 1$.

The idea behind these two assumptions are similar to Assumptions 1 and 2 in (3.7) (main paper). In particular, a similar argument as with Assumption 1 demonstrates that Assumption 3 is very benign. Assumption 4 is satisfied for data generation processes and estimators that are both invariant under orthogonal conjugation. In particular, consider the PCA model $y = \mathcal{B}^\star z + \epsilon$ for $\mathcal{B}^\star \in \mathbb{R}^{p_1\times k}$ and ϵ is a Gaussian vector with independent and identically distributed coordinates. Consider the PCA-estimator that finds top components of the empirical covariance of y from observations. Then the estimator satisfies Assumption 4 in (18).

Proposition 2 (Refined False Discovery Bound)**.** *Consider the setup in Theorem 1. Suppose additionally that Assumptions 3 and 4 in* (18) *are satisfied. Let the average number of discoveries from* $n/2$ *observations be denoted by* $q := \mathbb{E}[\dim(\hat{\mathcal{C}}(\mathcal{D}(n/2)))]$*. Then, for any fixed* $M \in \mathcal{C}^{\star\perp}$ *with* $\|M\|_2 = 1$*, the expected false discovery of a stable column-space* $\mathcal{C}$ *is bounded by:*

$$\mathbb{E}\left[\mathrm{trace}\left(\mathcal{P}_{\mathcal{C}}\mathcal{P}_{\mathcal{C}^{\star\perp}}\right)\right] \leq \frac{q^2}{p_1} + f\left(\kappa_{indiv}\right) + \frac{2q}{\alpha}(1-\alpha+\sqrt{1-\alpha}), \tag{19}$$

where $\kappa_{indiv} := \mathbb{E}\left[\|[\mathcal{P}_{\mathrm{span}(M)}, \mathcal{P}_{\mathcal{C}^{\star\perp}}]\|_F\right]$ *and* $f(\kappa_{indiv}) = p_1\kappa_{indiv}^2 + 2q\kappa_{indiv}$*.*

References

Meinshausen, N. and P. Bühlmann (2010) Stability selection. *Journal of Royal Statistical Methodology (Series B)*, *72*(4), 417–473.

Shah, R. and J. Samworth (2013) Variable selection and error control: another look at stability selection. *Journal of Royal Statistical Society (Series B)*, *75*(1), 55–80.