Supplementary Material for CVPR Submission 2020

Anonymous Authors

April 10, 2017

Abstract

This supplementary material contains a complete proof of Theorem 2.1 and Lemma 3.1 of the main paper. Section 1 discusses some technical notations and geometric observations, Section 2 and 3 restates and shows a complete proof for the theorem and lemma respectively, and Section 4 contains two auxiliary lemmas used in the main proof.

1 Notations and Recaps

Before commencing our main proof, we introduce some notations, recap the simplified objective function $\widehat{\varphi}$ and a few basic observations on its geometry. In this note, we use $\boldsymbol{C}_{\boldsymbol{a}}$ or $\boldsymbol{C}_{\widetilde{\boldsymbol{a}}}$ exchangeably to denote the $m \times m$ circulant matrix generated by a k-length short convolutional kernel $\boldsymbol{a}$ without ambiguity.

Reversed Circulant Matrix Let $\check{\boldsymbol{C}}_{\boldsymbol{a}_0}$ be the reversed circulant matrix for $\boldsymbol{a}_0$ such that

$$\check{\boldsymbol{C}}_{\boldsymbol{a}_0} = \left[\begin{array}{c|c|c|c|c} s_0\left[\widetilde{\boldsymbol{a}_0}\right] & s_{-1}\left[\widetilde{\boldsymbol{a}_0}\right] & s_{-2}\left[\widetilde{\boldsymbol{a}_0}\right] & \dots & s_{-(m-1)}\left[\widetilde{\boldsymbol{a}_0}\right] \end{array}\right] \in \mathbb{R}^{m\times m}, \tag{1.1}$$

Here, $s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]$ denotes a cyclic shift as defined in Equation (3) of the main paper. Since the i-th entry of $\boldsymbol{C}_{\boldsymbol{a}}^*\widetilde{\boldsymbol{a}_0}$ satisfies

$$\left[\boldsymbol{C}_{\boldsymbol{a}}^*\widetilde{\boldsymbol{a}_0}\right]_i = \left\langle s_i\left[\widetilde{\boldsymbol{a}}\right], \widetilde{\boldsymbol{a}_0}\right\rangle = \left\langle \widetilde{\boldsymbol{a}}, s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right]\right\rangle = \left\langle \boldsymbol{a}, \boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right]\right\rangle, \tag{1.2}$$

therefore following equation always holds

$$\boldsymbol{C}_{\boldsymbol{a}}^*\widetilde{\boldsymbol{a}_0} = \check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\widetilde{\boldsymbol{a}}. \tag{1.3}$$

Projection onto I Let $\boldsymbol{e}_0, \dots, \boldsymbol{e}_{m-1}$ denote the standard basis vectors. For index set $I = \left\{i_1, \dots, i_{|I|}\right\} \neq \emptyset$, let

$$\boldsymbol{V}_I \doteq \left[\boldsymbol{e}_{i_1} \mid \boldsymbol{e}_{i_2} \mid \cdots \mid \boldsymbol{e}_{i_{|I|}}\right] \in \mathbb{R}^{m\times|I|}, \tag{1.4}$$

and then

$$\boldsymbol{P}_I = \boldsymbol{V}_I\boldsymbol{V}_I^* \tag{1.5}$$

denote the projection onto the coordinates indexed by I while setting other entries zero.

Partial Signed Support Let $\boldsymbol{u}, \boldsymbol{v}$ be two vectors of the same dimension. If $\operatorname{supp}(\boldsymbol{u}) \subseteq \operatorname{supp}(\boldsymbol{v})$ and $\boldsymbol{u}(i)\boldsymbol{v}(i) \geq 0$ for all $i \in \operatorname{supp}(\boldsymbol{v})$, then $\boldsymbol{u}$ attains the partial signed support of $\boldsymbol{v}$, denoted as $\boldsymbol{u} \trianglelefteq \boldsymbol{v}$.

Piecewise Quadratic Function Let $\boldsymbol{x}^*(\boldsymbol{a})$ denote the minimizer for simplified objective function

$$\widehat{\varphi}(\boldsymbol{a}) = \min_{\boldsymbol{x}} \tfrac{1}{2} \left\|\widetilde{\boldsymbol{a}_0}\right\|_2^2 + \tfrac{1}{2} \left\|\boldsymbol{x}\right\|_2^2 - \langle \boldsymbol{a} \circledast \boldsymbol{x}, \widetilde{\boldsymbol{a}_0} \rangle + \lambda \left\|\boldsymbol{x}\right\|_1, \tag{1.6}$$

with sign $\boldsymbol{\sigma}$ and support I defined as

$$\boldsymbol{\sigma} \doteq \operatorname{sign}\left(\boldsymbol{x}^*(\boldsymbol{a})\right) \in \{-1, 0, 1\}^m, \quad I \doteq \operatorname{supp}(\boldsymbol{\sigma}) \subseteq \{0, 1, \dots, m-1\}^{1}. \tag{1.7}$$

By stationary condition for $\boldsymbol{x}^*(\boldsymbol{a})$, we obtain

$$\boldsymbol{x}^*(\boldsymbol{a}) = \mathrm{SOFT}_\lambda \left[\boldsymbol{C}_{\boldsymbol{a}}^* \widetilde{\boldsymbol{a}_0}\right] = \mathrm{SOFT}_\lambda \left[\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^* \boldsymbol{\iota} \boldsymbol{a}\right], \tag{1.8}$$

where $\mathrm{SOFT}_\lambda[u] = \operatorname{sign}(u) \max\{|u| - \lambda, 0\}$ is the entry-wise soft-thresholding operator.

For each sign pattern $\boldsymbol{\sigma} = \operatorname{supp}(\boldsymbol{x}^*) \in \{-1, 0, 1\}^m$, there exists corresponding region on the sphere such that

$$R_{\boldsymbol{\sigma}} = \left\{\boldsymbol{a} \mid \operatorname{sign}\left(\mathrm{SOFT}_\lambda \left[\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^* \boldsymbol{\iota} \boldsymbol{a}\right]\right) = \boldsymbol{\sigma}\right\}. \tag{1.9}$$

On the relative interior of each $R_{\boldsymbol{\sigma}}$, the function $\widehat{\varphi}$ has a simple expression:

$$\widehat{\varphi}(\boldsymbol{a}) = \widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) \doteq -\tfrac{1}{2} \boldsymbol{a}^* \boldsymbol{\iota}^* \check{\boldsymbol{C}}_{\boldsymbol{a}_0} \boldsymbol{P}_I \check{\boldsymbol{C}}_{\boldsymbol{a}_0}^* \boldsymbol{\iota} \boldsymbol{a} + \lambda \boldsymbol{\sigma}^* \boldsymbol{P}_I \check{\boldsymbol{C}}_{\boldsymbol{a}_0}^* \boldsymbol{\iota} \boldsymbol{a} + \tfrac{1}{2} - \tfrac{\lambda^2 |I|}{2}. \tag{1.10}$$

Therefore, the objective function is piecewise quadratic and can be rewritten as

$$\widehat{\varphi}_{\boldsymbol{\sigma}} = \tfrac{1}{2} \boldsymbol{a}^* \boldsymbol{M}_{\boldsymbol{\sigma}} \boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^* \boldsymbol{a} + Const_{\boldsymbol{\sigma}}, \tag{1.11}$$

with

$$\boldsymbol{M}_{\boldsymbol{\sigma}} = -\boldsymbol{\iota}^* \check{\boldsymbol{C}}_{\boldsymbol{a}_0} \boldsymbol{P}_I \check{\boldsymbol{C}}_{\boldsymbol{a}_0}^* \boldsymbol{\iota}, \quad \boldsymbol{b}_{\boldsymbol{\sigma}} = \lambda \boldsymbol{\iota}^* \check{\boldsymbol{C}}_{\boldsymbol{a}_0} \boldsymbol{P}_I \boldsymbol{\sigma}. \tag{1.12}$$

With above notations clarified, we are now ready to present a proof for the main theorem.

2 Proof of Theorem 2.1

Theorem 2.1 *Let*

$$\Xi = \left\{\operatorname{sign}\left(\mathrm{SOFT}_\lambda \left[\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^* \boldsymbol{\iota} \boldsymbol{a}\right]\right) \mid \boldsymbol{a} \in \mathbb{S}^{k-1}\right\}. \tag{2.1}$$

Define $\mathcal{I} = \{\operatorname{supp}(\boldsymbol{\sigma}) \mid \boldsymbol{\sigma} \in \Xi\}$. *For each nonempty* $I = \left\{i_1 < i_2 < \cdots < i_{|I|}\right\} \in \mathcal{I}$, *let*

$$\boldsymbol{W}_I = \left[\frac{\boldsymbol{\iota}^* s_{-i_1}\left[\widetilde{\boldsymbol{a}_0}\right]}{\left\|\boldsymbol{\iota}^* s_{-i_1}\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2} \middle| \frac{\boldsymbol{\iota}^* s_{-i_2}\left[\widetilde{\boldsymbol{a}_0}\right]}{\left\|\boldsymbol{\iota}^* s_{-i_2}\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2} \middle| \cdots \middle| \frac{\boldsymbol{\iota}^* s_{-i_{|I|}}\left[\widetilde{\boldsymbol{a}_0}\right]}{\left\|\boldsymbol{\iota}^* s_{-i_{|I|}}\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2} \right] \in \mathbb{R}^{k \times |I|} \tag{2.2}$$

Suppose that for every $I \in \mathcal{I}$ *of size greater than one,*

$$\left\|\boldsymbol{W}_I^* \boldsymbol{W}_I - \boldsymbol{I}\right\|_{\ell^2 \to \ell^2} < \frac{\lambda^2}{6} \tag{2.3}$$

and that $\lambda < 1$. *Then every local minimum* $\bar{\boldsymbol{a}}$ *of* $\widehat{\varphi}$ *over* $\mathbb{S}^{k-1}$ *satisfies either* $\bar{\boldsymbol{a}} \in R_{\mathbf{0}}$ *(in which case* $\bar{\boldsymbol{a}}$ *is also a global maximum), or*

$$\bar{\boldsymbol{a}} = \pm \frac{\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]}{\left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2} \tag{2.4}$$

for some shift τ.

[1] As in the main paper, we assume that m dimensional vectors are indexed by the integers $0, 1, \dots, m-1$.

Proof On the relative interior of each $R_{\boldsymbol{\sigma}}$, the simplified objective function

$$\widehat{\varphi}_{\boldsymbol{\sigma}} = \tfrac{1}{2}\boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\boldsymbol{a} + c_{\boldsymbol{\sigma}} \tag{2.5}$$

has Euclidean derivative and Hessian

$$\begin{aligned} \nabla\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) &= \boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}, & (2.6)\\ \nabla^2\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) &= \boldsymbol{M}_{\boldsymbol{\sigma}}. & (2.7) \end{aligned}$$

As we assume $\boldsymbol{a}$ to have unit Frobenius norm, or to live on a sphere, the more natural Riemannian gradient and Hessian are defined as

$$\begin{aligned} \operatorname{grad}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\boldsymbol{a}) &= \boldsymbol{P}_{\boldsymbol{a}^\perp}\nabla\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) & (2.8)\\ &= \boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}} - \boldsymbol{a}(\boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\boldsymbol{a}), & (2.9)\\ \operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\boldsymbol{a}) &= \boldsymbol{P}_{\boldsymbol{a}^\perp}\Big(\nabla^2\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) - \langle\nabla\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}), \boldsymbol{a}\rangle\,\boldsymbol{I}\Big)\boldsymbol{P}_{\boldsymbol{a}^\perp} & (2.10)\\ &= \boldsymbol{P}_{\boldsymbol{a}^\perp}\Big(\boldsymbol{M}_{\boldsymbol{\sigma}} - (\boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\boldsymbol{a})\boldsymbol{I}\Big)\boldsymbol{P}_{\boldsymbol{a}^\perp}. & (2.11) \end{aligned}$$

Here, $\boldsymbol{P}_{\boldsymbol{a}^\perp} = \boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^*$ denotes projection onto the tangent space over the sphere at $\boldsymbol{a}$. As in the Euclidean space, a stationary point on the sphere needs to satisfy $\operatorname{grad}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\boldsymbol{a}) = \boldsymbol{0}$. At a stationary point $\bar{\boldsymbol{a}}$, if $\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}})$ is positive semidefinite, the function is convex and $\bar{\boldsymbol{a}}$ is a local minimum; if $\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}})$ has a negative eigenvalue, then there exists a direction alone which the objective value decreases and hence $\bar{\boldsymbol{a}}$ is a saddle point [AMS07].

Let $I = \left\{i_1 < i_2 < \cdots < i_{|I|}\right\}$ and define

$$\begin{aligned} \eta_i &= \left\|\boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2 \quad \forall i \in I, & (2.12)\\ \boldsymbol{\eta} &= (\eta_{i_1}, \eta_{i_2}, \ldots, \eta_{i_{|I|}}) \in \mathbb{R}^{|I|}, & (2.13) \end{aligned}$$

and

$$\boldsymbol{U} = \left[\frac{\sigma_{i_1}\boldsymbol{\iota}^* s_{-i_1}\left[\widetilde{\boldsymbol{a}_0}\right]}{\eta_{i_1}}\middle|\frac{\sigma_{i_2}\boldsymbol{\iota}^* s_{-i_2}\left[\widetilde{\boldsymbol{a}_0}\right]}{\eta_{i_2}}\middle|\cdots\middle|\frac{\sigma_{i_{|I|}}\boldsymbol{\iota}^* s_{-i_{|I|}}\left[\widetilde{\boldsymbol{a}_0}\right]}{\eta_{i_{|I|}}}\right] \in \mathbb{R}^{k\times|I|}. \tag{2.14}$$

Here, columns of $\boldsymbol{U}$ have unit ℓ^2 norm. Then we have

$$\boldsymbol{M}_{\boldsymbol{\sigma}} = -\boldsymbol{U}\operatorname{diag}(\boldsymbol{\eta})^2\boldsymbol{U}^*, \quad \boldsymbol{b}_{\boldsymbol{\sigma}} = \lambda\boldsymbol{U}\boldsymbol{\eta}. \tag{2.15}$$

As I is defined via soft thresholding, we have $\left|[\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a}]_i\right| > \lambda$ holds for every $i \in I$. Hence,

$$\begin{aligned} \boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\boldsymbol{a} &= -\boldsymbol{a}^*\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_I\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} + \lambda\boldsymbol{\sigma}^*\boldsymbol{P}_I\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} & (2.16)\\ &= -\left\|\boldsymbol{P}_I\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a}\right\|_2^2 + \lambda\left\|\boldsymbol{P}_I\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a}\right\|_1 & (2.17)\\ &< 0 & (2.18) \end{aligned}$$

holds at any $\boldsymbol{a} \in \operatorname{cl}(R_{\boldsymbol{\sigma}}) \setminus R_{\boldsymbol{0}}$.

Stationary point and implications Consider any stationary point $\bar{\boldsymbol{a}} \in \operatorname{cl}(R_{\boldsymbol{\sigma}}) \setminus R_{\boldsymbol{0}}$ of $\widehat{\varphi}$. By continuity of the gradient of $\widehat{\varphi}$ (proved in Lemma 2.2), $\bar{\boldsymbol{a}}$ is also a stationary point of $\widehat{\varphi}_{\boldsymbol{\sigma}}$. By definition, $\operatorname{grad}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}}) = \boldsymbol{0}$, which implies that

$$(\bar{\boldsymbol{a}}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\bar{\boldsymbol{a}} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\bar{\boldsymbol{a}})\bar{\boldsymbol{a}} = \boldsymbol{M}_{\boldsymbol{\sigma}}\bar{\boldsymbol{a}} + \boldsymbol{b}_{\boldsymbol{\sigma}}. \tag{2.19}$$

Note that since $\bar{\boldsymbol{a}}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\bar{\boldsymbol{a}} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\bar{\boldsymbol{a}} \neq 0$ and $\boldsymbol{b}_{\boldsymbol{\sigma}} \in \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}})$, this implies that $\bar{\boldsymbol{a}} \in \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}})$.

Let $\gamma = -\left(\boldsymbol{a}^*\boldsymbol{M}_{\boldsymbol{\sigma}}\boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^*\boldsymbol{a}\right) > 0$, then the condition for a stationary point $\bar{\boldsymbol{a}}$ becomes

$$\gamma\bar{\boldsymbol{a}} = \boldsymbol{U}\operatorname{diag}(\boldsymbol{\eta})^2\boldsymbol{U}^*\bar{\boldsymbol{a}} - \lambda\boldsymbol{U}\boldsymbol{\eta}. \tag{2.20}$$

Let $\boldsymbol{\alpha} = \boldsymbol{U}^* \bar{\boldsymbol{a}}$, and note that for each j, $\alpha_j > 0$ and $\alpha_j \eta_j > \lambda$. In terms of $\boldsymbol{U}$, the stationarity condition becomes

$$\gamma \boldsymbol{\alpha} = \boldsymbol{U}^* \boldsymbol{U} \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \lambda \boldsymbol{U}^* \boldsymbol{U} \boldsymbol{\eta}. \tag{2.21}$$

Since the diagonal elements of $\boldsymbol{U}^* \boldsymbol{U}$ are all ones, and hence can be written as

$$\boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I} + \boldsymbol{\Delta}. \tag{2.22}$$

We have

$$\gamma \boldsymbol{\alpha} = \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \lambda \boldsymbol{\eta} + \boldsymbol{\Delta} \Big\{ \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \lambda \boldsymbol{\eta} \Big\}. \tag{2.23}$$

As $\boldsymbol{\alpha} \succ \mathbf{0}$ and $\operatorname{diag}(\boldsymbol{\eta}) \boldsymbol{\alpha} \succ \lambda \cdot \mathbf{1}$[2], together with an auxiliary Lemma 4.2 proved in Section 4, we have

$$\begin{aligned}
\left\| \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \lambda \boldsymbol{\eta} \right\|_2 &\leq \|\boldsymbol{\alpha}\|_2 && (2.24)\\
&= \|\boldsymbol{U}^* \bar{\boldsymbol{a}}\|_2 && (2.25)\\
&\leq \|\boldsymbol{U}\|_{\ell^2 \to \ell^2} && (2.26)\\
&\leq \sqrt{3/2}, && (2.27)
\end{aligned}$$

whence

$$\left\| \boldsymbol{\Delta} \Big\{ \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \lambda \boldsymbol{\eta} \Big\} \right\|_\infty \leq \sqrt{3/2} \times \|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^\infty}. \tag{2.28}$$

Suppose that $\boldsymbol{\Delta}$ is small enough that the right hand side of (2.28) is bounded by $\lambda^2/2$, i.e.,

$$\|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^\infty} \leq \frac{\lambda^2}{\sqrt{6}}. \tag{2.29}$$

Plugging back into the stationary condition $\operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \gamma \boldsymbol{\alpha} = \lambda \boldsymbol{\eta} - \boldsymbol{\Delta} \Big\{ \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{\alpha} - \lambda \boldsymbol{\eta} \Big\}$ gives

$$(\operatorname{diag}(\boldsymbol{\eta})^2 - \gamma) \boldsymbol{\alpha} \succ \lambda \boldsymbol{\eta} - \lambda^2/2 \succ \mathbf{0}. \tag{2.30}$$

Since $\alpha_i < 1$ and $\eta_i > \lambda$ for all i, which implies that

$$\gamma < \eta_{\min}^2 - \lambda \eta_{\min} + \lambda^2/2, \tag{2.31}$$

where $\eta_{\min}$ is the smallest of the η_i.

Negative curvature in Hessian Recall the Riemannian Hessian on the sphere is defined as

$$\begin{aligned}
\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\boldsymbol{a}) &= \boldsymbol{P}_{\boldsymbol{a}^\perp} \Big(\boldsymbol{M}_{\boldsymbol{\sigma}} - (\boldsymbol{a}^* \boldsymbol{M}_{\boldsymbol{\sigma}} \boldsymbol{a} + \boldsymbol{b}_{\boldsymbol{\sigma}}^* \boldsymbol{a}) \boldsymbol{I} \Big) \boldsymbol{P}_{\boldsymbol{a}^\perp} && (2.32)\\
&= \boldsymbol{P}_{\boldsymbol{a}^\perp} \Big(-\boldsymbol{U} \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{U}^* + \gamma \boldsymbol{I} \Big) \boldsymbol{P}_{\boldsymbol{a}^\perp}. && (2.33)
\end{aligned}$$

Below argument shows that this Riemmanian Hessian has negative eigenvalues. Let $\tilde{\boldsymbol{U}}$ be an orthonormal matrix generated via

$$\tilde{\boldsymbol{U}} \doteq \boldsymbol{U} (\boldsymbol{U}^* \boldsymbol{U})^{-1/2} \tag{2.34}$$

Whenever $\|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^2} < 1/2$ holds, Lemma 4.2 guarantees $\left\| \boldsymbol{U} - \tilde{\boldsymbol{U}} \right\|_{\ell^2 \to \ell^2} < 3 \|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^2}$. Under this condition, we can lower bound the smallest nonzero eigenvalue of $\boldsymbol{U} \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{U}^*$, as

$$\lambda_{\min}(\boldsymbol{U} \operatorname{diag}(\boldsymbol{\eta})^2 \boldsymbol{U}^*) = \sigma_{\min}(\boldsymbol{U} \operatorname{diag}(\boldsymbol{\eta}))^2 \tag{2.35}$$

[2]Here, $\succ$ denotes element-wise inequality between vectors.

$$\geq \left(\max\left\{\sigma_{\min}\left(\tilde{\boldsymbol{U}}\operatorname{diag}(\boldsymbol{\eta})\right) - \left\|\boldsymbol{U}-\tilde{\boldsymbol{U}}\right\|\left\|\operatorname{diag}(\boldsymbol{\eta})\right\|, 0\right\}\right)^2 \tag{2.36}$$

$$= \left(\max\left\{\eta_{\min} - \left\|\boldsymbol{U}-\tilde{\boldsymbol{U}}\right\|\eta_{\max}, 0\right\}\right)^2 \tag{2.37}$$

$$\geq \eta_{\min}^2 - 3\eta_{\max}\eta_{\min}\left\|\boldsymbol{\Delta}\right\|_{\ell^2\to\ell^2}. \tag{2.38}$$

Since $\lambda < \eta_{\min} \leq \eta_{\max} \leq 1$, additionally if $\|\boldsymbol{\Delta}\|_{\ell^2\to\ell^2} \leq \frac{\lambda}{6}$, we have

$$3\eta_{\max}\eta_{\min}\left\|\boldsymbol{\Delta}\right\|_{\ell^2\to\ell^2} \leq \lambda\eta_{\min} - \lambda^2/2, \tag{2.39}$$

Together with (2.29) and (2.31), we can obtain

$$\lambda_{\min}(\boldsymbol{U}\operatorname{diag}(\boldsymbol{\eta})^2\boldsymbol{U}^*) > \gamma, \tag{2.40}$$

or

$$\lambda_{\max}(\boldsymbol{M}_{\boldsymbol{\sigma}}) < -\gamma. \tag{2.41}$$

Thus, whenever the following conditions are satisfied

$$\lambda < 1, \quad \|\boldsymbol{\Delta}\|_{\ell^2\to\ell^\infty} \leq \frac{\lambda^2}{\sqrt{6}}, \quad \|\boldsymbol{\Delta}\|_{\ell^2\to\ell^2} \leq \frac{\lambda}{6}, \tag{2.42}$$

we have $\lambda_{\max}(\boldsymbol{M}_{\boldsymbol{\sigma}}) < -\gamma$ as desired.

Above calculations imply that for every $\boldsymbol{\xi} \in \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}}) \subseteq \mathbb{R}^k$,

$$\boldsymbol{\xi}^*\left(\boldsymbol{M}_{\boldsymbol{\sigma}} + \gamma\boldsymbol{I}\right)\boldsymbol{\xi} < 0. \tag{2.43}$$

Since

$$\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}}) = \boldsymbol{P}_{\bar{\boldsymbol{a}}^\perp}\Big(\boldsymbol{M}_{\boldsymbol{\sigma}} + \gamma\boldsymbol{I}\Big)\boldsymbol{P}_{\bar{\boldsymbol{a}}^\perp}, \tag{2.44}$$

for $\boldsymbol{\xi} \in \bar{\boldsymbol{a}}^\perp \cap \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}})$,

$$\boldsymbol{\xi}^*\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}})\,\boldsymbol{\xi} < 0. \tag{2.45}$$

Hence, on the relative interior $\operatorname{relint}(R_{\boldsymbol{\sigma}})$, $\widehat{\varphi} \equiv \widehat{\varphi}_{\boldsymbol{\sigma}}$ obtains, and so this implies that for $\|\boldsymbol{\sigma}\|_0 > 1$, there are no local minima in $\operatorname{relint}(R_{\boldsymbol{\sigma}})$.

Relative boundaries We first note that if $\|\boldsymbol{\sigma}\|_0 = 1$ and $I = \{i\}$, either $R_{\boldsymbol{\sigma}}$ is empty when $\|\boldsymbol{\iota}^* s_{-i}[\widetilde{\boldsymbol{a}_0}]\|_2 \leq \lambda$, or it contains an open ball around $\operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}}) \bigcap \mathbb{S}^{k-1} = \pm\frac{\boldsymbol{\iota}^* s_{-i}[\widetilde{\boldsymbol{a}_0}]}{\|\boldsymbol{\iota}^* s_{-i}[\widetilde{\boldsymbol{a}_0}]\|_2}$. Hence, if $\bar{\boldsymbol{a}} \in \operatorname{relbdy}(R_{\boldsymbol{\sigma}})$ is a stationary point and $\boldsymbol{\sigma} \neq \mathbf{0}$, we necessarily have $\|\boldsymbol{\sigma}\|_0 \geq 2$.

Since $\bar{\boldsymbol{a}}$ is on the boundary of $R_{\boldsymbol{\sigma}}$, it is also in $\operatorname{relbdy}(\operatorname{cl}(R_{\boldsymbol{\sigma}'}))$ for some $\boldsymbol{\sigma}' \neq \boldsymbol{\sigma}$. Let

$$\Xi = \left\{\boldsymbol{\sigma}' \mid \bar{\boldsymbol{a}} \in \operatorname{relbdy}(\operatorname{cl}(R_{\boldsymbol{\sigma}'}))\right\}. \tag{2.46}$$

Suppose that for every $\boldsymbol{\sigma}' \in \Xi$, $\boldsymbol{\sigma} \trianglelefteq \boldsymbol{\sigma}'$. Hence, $\operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}}) \subseteq \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}'})$ for every $\boldsymbol{\sigma}' \in \Xi$ and

$$\boldsymbol{\xi}^*\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}'}\right](\bar{\boldsymbol{a}})\,\boldsymbol{\xi} < 0, \quad \forall\, \boldsymbol{\xi} \in \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}}),\ \boldsymbol{\sigma}' \in \Xi. \tag{2.47}$$

By continuity of the gradients, $\bar{\boldsymbol{a}}$ is a stationary point for *every* $\widehat{\varphi}_{\boldsymbol{\sigma}'}$ such that $\boldsymbol{\sigma}' \in \Xi$. If we choose an arbitrary nonzero $\boldsymbol{\xi} \in \operatorname{range}(\boldsymbol{M}_{\boldsymbol{\sigma}})$, we have that for every $\boldsymbol{\sigma}' \in \Xi$,

$$\widehat{\varphi}_{\boldsymbol{\sigma}'}(\mathcal{P}_{\mathbb{S}^{k-1}}[\bar{\boldsymbol{a}} + t\boldsymbol{\xi}]) < \widehat{\varphi}(\bar{\boldsymbol{a}}) - \Omega(t^2). \tag{2.48}$$

There exists a neighborhood N of $\bar{\boldsymbol{a}}$ for which, at every $\boldsymbol{a} \in N \cap R_{\boldsymbol{\sigma}'}$, $\widehat{\varphi}(\boldsymbol{a}) = \widehat{\varphi}_{\boldsymbol{\sigma}'}(\boldsymbol{a}) \leq \widehat{\varphi}(\bar{\boldsymbol{a}})$ for some $\boldsymbol{\sigma}' \in \Xi$. Hence, $\bar{\boldsymbol{a}}$ is not a local minimum of $\widehat{\varphi}$.

Local minima If $\|\boldsymbol{\sigma}\|_0 = 1$ and $I = \{i\}$, then the simplified objective function is

$$\widehat{\varphi}_{\boldsymbol{\sigma}} = -\tfrac{1}{2} \left\langle \sigma_i \boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right], \boldsymbol{a} \right\rangle^2 + \lambda \left\langle \sigma_i \boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right], \boldsymbol{a} \right\rangle + c_{\boldsymbol{\sigma}}. \tag{2.49}$$

The minimizer appears at the boundary for $\left\langle \sigma_i \boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right], \boldsymbol{a} \right\rangle$, namely $\left\|\boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2$[3] obtained by

$$\bar{\boldsymbol{a}} = \sigma_i \frac{\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]}{\left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2}. \tag{2.50}$$

It can be easily verified that

$$\begin{aligned}
\operatorname{grad}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}}) &= \Big(-\left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2^2 + \lambda \left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2\Big) \times \left(\boldsymbol{I} - \bar{\boldsymbol{a}}\bar{\boldsymbol{a}}^*\right)\bar{\boldsymbol{a}} && (2.51)\\
&= \mathbf{0} && (2.52)\\
\operatorname{Hess}\left[\widehat{\varphi}_{\boldsymbol{\sigma}}\right](\bar{\boldsymbol{a}}) &= \boldsymbol{P}_{\boldsymbol{a}^\perp}\Big(-\left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2^2 \bar{\boldsymbol{a}}\bar{\boldsymbol{a}}^* + (1 - \lambda\left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2)\boldsymbol{I}\Big)\boldsymbol{P}_{\boldsymbol{a}^\perp} && (2.53)\\
&= (1 - \lambda\left\|\boldsymbol{\iota}^* s_\tau\left[\widetilde{\boldsymbol{a}_0}\right]\right\|_2)\boldsymbol{P}_{\boldsymbol{a}^\perp}\boldsymbol{P}_{\boldsymbol{a}^\perp} && (2.54)\\
&\succeq 0 && (2.55)
\end{aligned}$$

Global maxima If $\|\boldsymbol{\sigma}\|_0 = 0$, then the objective remains constant and achieves the global maximum.

■

Lemma 2.2 (Continuity of the Gradient of $\widehat{\varphi}$) *$\nabla\widehat{\varphi}$ is a continuous function of $\boldsymbol{a}$.*

Proof Recall that for a given $\boldsymbol{\sigma}$, the gradient

$$\begin{aligned}
\nabla\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) &= -\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_I\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} + \lambda\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_I\boldsymbol{\sigma} && (2.56)\\
&= -\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_I\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} - \lambda\boldsymbol{\sigma}\right) && (2.57)
\end{aligned}$$

This is a continuous function within the relative interior of $R_{\boldsymbol{\sigma}}$. Next, we show this function is continuous at the relative boundary of $R_{\boldsymbol{\sigma}}$. Let $\boldsymbol{a}' \in \operatorname{relbdy}(R_{\boldsymbol{\sigma}})$, and $\boldsymbol{\sigma}' = \operatorname{sign}(\boldsymbol{a}')$, $I = \operatorname{supp}(\boldsymbol{\sigma}')$ are the corresponding sign and support. Without loss of generality, we assume $\boldsymbol{\sigma}' \trianglelefteq \boldsymbol{\sigma}$, denote $\boldsymbol{a} = \boldsymbol{a}' + \varepsilon\boldsymbol{\delta}$ ($\|\boldsymbol{\delta}\|_2 = 1$) and $I_{\boldsymbol{\delta}} = I \setminus I'$, then

$$\begin{aligned}
\nabla\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) - \nabla\widehat{\varphi}_{\boldsymbol{\sigma}'}(\boldsymbol{a}') &= -\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_I\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} - \lambda\boldsymbol{\sigma}\right) + \boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_{I'}\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a}' - \lambda\boldsymbol{\sigma}'\right) && (2.58)\\
&= -\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}(\boldsymbol{P}_{I'} + \boldsymbol{P}_{I_\delta})\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} - \lambda\boldsymbol{\sigma}\right) + \boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_{I'}\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a}' - \lambda\boldsymbol{\sigma}'\right) && (2.59)\\
&= -\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_{I_\delta}\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} - \lambda\boldsymbol{\sigma}\right) - \varepsilon\boldsymbol{\iota}^*\check{\boldsymbol{C}}_{\boldsymbol{a}_0}\boldsymbol{P}_{I'}\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{\delta} && (2.60)
\end{aligned}$$

Since $\left\|\boldsymbol{P}_{I_\delta}\left(\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{a} - \lambda\boldsymbol{\sigma}\right)\right\|_\infty = \varepsilon\left\|\boldsymbol{P}_{I_\delta}\check{\boldsymbol{C}}_{\boldsymbol{a}_0}^*\boldsymbol{\iota}\boldsymbol{\delta}\right\|_\infty$, we have $\|\nabla\widehat{\varphi}_{\boldsymbol{\sigma}}(\boldsymbol{a}) - \nabla\widehat{\varphi}_{\boldsymbol{\sigma}'}(\boldsymbol{a}')\|_\infty \le \mathcal{O}(\varepsilon)$. ■

3 Proof of Lemma 3.1

Lemma 3.1 *Let $\lambda_{rel} = \lambda/\|\boldsymbol{x}_0\|_\infty$, suppose the ground truth $\boldsymbol{a}_0$ satisfies*

$$\left|\left\langle \boldsymbol{a}_0, \boldsymbol{\iota} s_{\tau\neq 0}\left[\widetilde{\boldsymbol{a}_0}\right]\right\rangle\right| < \lambda_{rel}^2 - \left(2 + 1/\lambda_{rel}^2\right)\sqrt{1 - \lambda_{rel}^2} \tag{3.1}$$

for any nonzero shift τ, and $\boldsymbol{x}_0$ is separated enough such that any two nonzero components are at least $2k$ entries away from each other. If initialized at some $\boldsymbol{a} \in \mathbb{S}^{k-1}$ that $|\langle \boldsymbol{a}, \boldsymbol{a}_0\rangle| > \lambda/\|\boldsymbol{x}_0\|_\infty$, a gradient descent algorithm minimizing $\varphi(\boldsymbol{a})$ recovers the signed ground truth $\pm\boldsymbol{a}_0$.

[3]The other boundary point is $\left\langle \sigma_i \boldsymbol{\iota}^* s_{-i}\left[\widetilde{\boldsymbol{a}_0}\right], \boldsymbol{a}\right\rangle = \lambda$, which achieves a smaller objective value.

Proof Without loss of generality, we are going to assume $\|\boldsymbol{x}_0\|_\infty = 1$ for simplicity. Given that

$$|\langle \boldsymbol{a}_0, \iota s_{\tau \neq 0} [\widetilde{\boldsymbol{a}_0}] \rangle| < \lambda^2 - \sqrt{1-\lambda^2} \left(2 + 1/\lambda^2\right) \tag{3.2}$$

and $\boldsymbol{a} = \langle \boldsymbol{a}, \boldsymbol{a}_0 \rangle \boldsymbol{a}_0 + \boldsymbol{\delta}$ with $\|\boldsymbol{\delta}\|_2 \leq \sqrt{1-\lambda^2}$, therefore

$$\begin{aligned} |\langle \boldsymbol{a}, \iota s_\tau [\boldsymbol{a}] \rangle| &= |\langle \langle \boldsymbol{a}, \boldsymbol{a}_0 \rangle \boldsymbol{a}_0 + \boldsymbol{\delta}, \iota s_\tau [\langle \boldsymbol{a}, \boldsymbol{a}_0 \rangle \boldsymbol{a}_0 + \boldsymbol{\delta}] \rangle| \\ &\leq \langle \boldsymbol{a}, \boldsymbol{a}_0 \rangle^2 |\langle \boldsymbol{a}_0, \iota s_\tau [\boldsymbol{a}_0] \rangle| + 2 \langle \boldsymbol{a}, \boldsymbol{a}_0 \rangle \|\boldsymbol{\delta}\|_2 + \|\boldsymbol{\delta}\|_2^2 \\ &< 1 - \sqrt{1-\lambda^2}/\lambda^2 \end{aligned}$$

Moreover, as $\boldsymbol{x}_0$ is sufficiently separated, we have

$$\begin{aligned} |\langle \boldsymbol{a}, \iota s_\tau [\boldsymbol{a}] \rangle \|\boldsymbol{x}^\star\|_\infty - \langle \boldsymbol{a}, \iota s_\tau [\boldsymbol{a}_0] \rangle \|\boldsymbol{x}_0\|_\infty| &\leq |\langle \boldsymbol{a}, \iota s_\tau [\boldsymbol{a}] \rangle| \|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|_\infty + |\langle \boldsymbol{a}, \iota s_\tau [\boldsymbol{a}_0 - \boldsymbol{a}] \rangle| \|\boldsymbol{x}_0\|_\infty \\ &< \lambda |\langle \boldsymbol{a}, \iota s_\tau [\boldsymbol{a}] \rangle| + \|\boldsymbol{a}_0 - \boldsymbol{a}\|_2 \|\boldsymbol{x}_0\|_\infty \\ &< \lambda. \end{aligned}$$

Hence, there exists a unique nonzero minimizer satisfying $\operatorname{supp} \boldsymbol{x}^\star \subset \operatorname{supp} \boldsymbol{x}_0$, and the optimality condition for $\boldsymbol{x}^\star$ implies

$$\boldsymbol{x}^\star = \mathrm{SOFT}_\lambda [\langle \boldsymbol{a}, \boldsymbol{a}_0 \rangle \boldsymbol{x}_0], \tag{3.3}$$

In this case, we can calculate the Euclidean gradient

$$\begin{aligned} \nabla \varphi(\boldsymbol{a}) &= \iota \boldsymbol{C}_{\boldsymbol{x}^\star}^* (\boldsymbol{a} \circledast \boldsymbol{x}^\star - \boldsymbol{a}_0 \circledast \boldsymbol{x}_0) &\quad (3.4) \\ &= \|\boldsymbol{x}^\star\|_2^2 \boldsymbol{a} - \langle \boldsymbol{x}^\star, \boldsymbol{x}_0 \rangle \boldsymbol{a}_0, &\quad (3.5) \end{aligned}$$

and the Riemannian gradient

$$\begin{aligned} \operatorname{grad} [\varphi] (\boldsymbol{a}) &= (\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^*) \nabla \varphi(\boldsymbol{a}) &\quad (3.6) \\ &= - \langle \boldsymbol{x}^\star, \boldsymbol{x}_0 \rangle (\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^*) \boldsymbol{a}_0. &\quad (3.7) \end{aligned}$$

It's easy to check that at any point along the geodesic curve between $\boldsymbol{a}_0$ and $\boldsymbol{a}$, support recovery of $\boldsymbol{x}^\star$ is achieved and a gradient descent algorithm moves towards the signed ground truth $\pm \boldsymbol{a}_0$ as desired. ■

4 Auxiliary Lemmas

Lemma 4.1 *(Lemma B.2 of [SQW15]) Suppose that $\boldsymbol{A} \succ \boldsymbol{0}$ is a positive definite matrix. For any symmetic matrix $\boldsymbol{\Delta}$ with $\|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^2} \leq \sigma_{\min}(\boldsymbol{A})/2$,*

$$\left\| (\boldsymbol{A} + \boldsymbol{\Delta})^{-1/2} - \boldsymbol{A}^{-1/2} \right\|_{\ell^2 \to \ell^2} \leq \frac{2 \|\boldsymbol{A}\|_{\ell^2 \to \ell^2}^{1/2} \|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^2}}{\sigma_{\min}(\boldsymbol{A})^2}. \tag{4.1}$$

Lemma 4.2 *Let $\boldsymbol{U}$ be a matrix such that $\boldsymbol{U}^* \boldsymbol{U} = \boldsymbol{I} + \boldsymbol{\Delta}$, with $\|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^2} < 1/2$. Then $\boldsymbol{U}^* \boldsymbol{U}$ is invertible,*

$$\|\boldsymbol{U}\|_{\ell^2 \to \ell^2} < \sqrt{3/2}, \tag{4.2}$$

and

$$\left\| \boldsymbol{U} - \boldsymbol{U} (\boldsymbol{U}^* \boldsymbol{U})^{-1/2} \right\|_{\ell^2 \to \ell^2} < 3 \|\boldsymbol{\Delta}\|_{\ell^2 \to \ell^2}. \tag{4.3}$$

Proof Upper bound for the first quantity can be derived

$$\begin{aligned}
\|\boldsymbol{U}\|_{\ell^2\to\ell^2} &= \sqrt{\|\boldsymbol{U}^*\boldsymbol{U}\|_{\ell^2\to\ell^2}} &(4.4)\\
&\le \sqrt{\|\boldsymbol{I}\|_{\ell^2\to\ell^2} + \|\boldsymbol{\Delta}\|_{\ell^2\to\ell^2}} &(4.5)\\
&< \sqrt{3/2}. &(4.6)
\end{aligned}$$

Applying Lemma 4.1 for the second term

$$\begin{aligned}
\left\|\boldsymbol{U} - \boldsymbol{U}(\boldsymbol{U}^*\boldsymbol{U})^{-1/2}\right\|_{\ell^2\to\ell^2} &\le \|\boldsymbol{U}\|_{\ell^2\to\ell^2}\left\|\boldsymbol{I} - (\boldsymbol{U}^*\boldsymbol{U})^{-1/2}\right\|_{\ell^2\to\ell^2} &(4.7)\\
&= \|\boldsymbol{U}\|_{\ell^2\to\ell^2}\left\|\boldsymbol{I}^{-1/2} - (\boldsymbol{I}+\boldsymbol{\Delta})^{-1/2}\right\|_{\ell^2\to\ell^2} &(4.8)\\
&\le \sqrt{3/2}\times 2\,\|\boldsymbol{\Delta}\|_{\ell^2\to\ell^2}\,, &(4.9)
\end{aligned}$$

Hence, we can obtain the claim by using $2\sqrt{3/2} < 3$ to simplify the constant.

■

References

[AMS07] P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, USA, 2007.

[SQW15] J. Sun, Q. Qu, and J. Wright. Complete dictionary recovery over the sphere. Preprint, http://arxiv.org/abs/1504.06785, 2015.