中国科学: 数学 2013 年 第 43 卷 第 2 期: 121～136
www.springerlink.com math.scichina.com

《中国科学》杂志社
SCIENCE CHINA PRESS

奇异值函数的二阶方向导数

张宁, 张立卫*, 肖现涛

大连理工大学数学科学学院, 大连 116024
E-mail: ningzhang_2008@yeah.net, lwzhang@dlut.edu.cn, xtxiao@dlut.edu.cn

收稿日期: 2011-09-05; 接受日期: 2012-07-26; * 通信作者
国家自然科学基金 (批准号: 11071029, 91130007 和 11101064) 资助项目

摘要 本文用一个直接的方法给出了奇异值函数的二阶方向导数公式. 作为应用, 利用这一公式建立了谱范数的上图集合与核范数的上图集合的切锥和二阶切集的具体表达式, 这些表达式在矩阵优化的一阶和二阶最优条件的研究中起着重要作用.

关键词 奇异值 二阶方向导数 切锥 二阶切集 谱范数 核范数

MSC (2010) 主题分类 90C30

1 引言

令 $\Re^{m\times n}$ 为所有 $m\times n$ 维实对称矩阵构成的线性空间. 不失一般性, 本文假设 $m\leqslant n$. 对任意给定矩阵 $X\in\Re^{m\times n}$, 存在正交矩阵 $U\in\Re^{m\times m}$ 和 $V\in\Re^{n\times n}$ 使得

$$X=U[\Sigma(X)\ \ 0]V^{\mathrm{T}}, \tag{1.1}$$

其中

$$\Sigma(X):=\operatorname{diag}(\sigma_1(X),\sigma_2(X),\ldots,\sigma_m(X)),$$

并且 $\sigma_1(X)\geqslant\sigma_2(X)\geqslant\cdots\geqslant\sigma_m(X)$. (1.1) 称为矩阵 X 的奇异值分解 (SVD), $\sigma_i(X)$, $i=1,\ldots,m$ 称为 X 的奇异值.

任意矩阵的奇异值可以由对称矩阵的特征值表示. 早在 1873 年, Beltrami 提出 X 的第 i 个奇异值可以表示成 XX^{T} (或者 $X^{\mathrm{T}}X$) 的第 i 个特征值的算术平方根. 紧随其后, Jordan 于 1874 年从另外的角度建立了奇异值与特征值之间的关系, 证明了如下矩阵的特征值中前 m 个最大的特征值即为矩阵 X 的所有奇异值:

$$\begin{bmatrix} 0 & X \\ X^{\mathrm{T}} & 0 \end{bmatrix}\in\Re^{(m+n)\times(m+n)}.$$

在接下来的讨论中, Beltrami 和 Jordan 的结论都起到了至关重要的作用.

奇异值在统计计算和低秩优化中扮演着很重要的角色 [1]. 此外, 控制系统中的许多重要的结构特性, 如鲁棒性和噪声的敏感度, 都可以由包含转移矩阵的奇异值的不等式表示 [2–5].

英文引用格式: Zhang N, Zhang L W, Xiao X T. The second-order directional derivatives of singular values (in Chinese). Sci Sin Math, 2013, 43: 121–136, doi: 10.1360/012011-668

众所周知, 一阶和二阶最优性条件是最优化的核心问题. 最近, Ding 等人[6] 介绍了一系列有着广泛应用背景的矩阵锥优化问题, 这些优化问题都与 k 范数的上图有着密切的关联. 若任意给定矩阵 $X \in \Re^{m\times n}$ 满足奇异值分解 (1.1), 则 X 的前 k 个最大的奇异值之和称为矩阵 X 的 k 范数, 记作 $\|X\|_{(k)}$, 即

$$\|X\|_{(k)} := \sum_{i=1}^{k} \sigma_i(X).$$

当 $k=1$ 时, 矩阵的 k 范数退化为谱范数 $\|\cdot\|_2$; 当 $k=m$ 时, 矩阵的 k 范数退化为核范数 $\|\cdot\|_*$. 由核范数上图 epi $\|\cdot\|_2$ 和谱范数的上图 epi $\|\cdot\|_*$ 的凸性可知, 在这些上图集合的二阶切集可以由奇异值函数的二阶方向导数表示, 而二阶切集在优化问题的二阶充分和必要条件的研究中有着至关重要的作用, 参见文献 [7,8].

本文给出了奇异值函数的一阶和二阶方向导数的公式. 基于这些公式, 分别给出了谱范数上图和核范数的上图的切锥和二阶切集的表达式.

本文的结构如下: 在第 2 节, 给出关于对称矩阵特征向量扰动分析的结论和非对称矩阵的奇异值分解中直交矩阵的扰动分析结论, 为研究奇异值的二阶方向导数打下基础. 本文关于任意奇异值函数的一阶和二阶方向导数的结果在第 3 节给出. 第 4 节给出了 epi $\|\cdot\|_2$ 和 epi $\|\cdot\|_*$ 的切锥和二阶切集计算公式.

本文使用以下符号: 任意给定 $m\times n$ 矩阵 Z 和指标集合 $\alpha \subseteq \{1,2,\ldots,m\}$, $\beta \subseteq \{1,2,\ldots,n\}$, $Z_{\alpha\beta}$ 表示 Z 的所有行指标在 α 中且列指标在 β 中的那些分量构成的子矩阵, Z_α 表示 Z 的所有指标在 α 中的那些列构成的子矩阵. 定义 $\|\cdot\|$ 为 Frobenius 范数, 即 $\|Z\| := (\sum_{i=1}^{m}\sum_{j=1}^{n}|Z_{ij}|^2)^{1/2}$. 同时, $Z=\mathrm{O}(t)$ 意味着当 $t\to 0$ 时, $\|Z\|/|t|$ 是一致有界的.

2 特征向量的扰动分析

给定矩阵 $X\in\Re^{m\times n}$, 本节讨论当 X 发生形如 $tH+\frac{t^2}{2}W$ (t 为变量) 的扰动时对应的特征向量的一阶展开式. 这个展开式在研究奇异值函数的二阶方向导数时扮演着重要的角色.

2.1 对称的情形

令 $\mathcal{S}^n$ 为所有 $n\times n$ 实对称矩阵构成的线性空间. 任意给定 $X\in\mathcal{S}^n$, 定义 $\lambda_1(X)\geqslant\lambda_2(X)\geqslant\cdots\geqslant\lambda_n(X)$ 为 X 的实特征值. 令 $\Lambda(X)\in\mathcal{S}^n$ 为第 i 个对角元为 $\lambda_i(X)$ 的对角矩阵, $i=1,\ldots,n$, 即 $\Lambda(X)=\operatorname{diag}(\lambda(X))$. 定义 $\mathcal{O}^n$ 为 $\Re^{n\times n}$ 中所有 $n\times n$ 的正交矩阵构成的集合.

给定 $X\in\mathcal{S}^n$, 则存在一个正交矩阵 $P\in\mathcal{O}^n$ 使得

$$X = P\operatorname{diag}(\lambda(X))P^{\mathrm{T}}. \tag{2.1}$$

记所有满足上述特征值分解 (2.1) 的矩阵 P 的全体为 $\mathcal{O}^n(X)$. 令 $\mu_1>\mu_2>\cdots>\mu_r$ 为 X 的不同特征值. 定义 $\{1,\ldots,n\}$ 的子集合

$$\alpha_k := \{i \mid \lambda_i(X)=\mu_k\}, \quad k=1,\ldots,r. \tag{2.2}$$

任意给定 $X,H,W\in\mathcal{S}^n$, 定义 $Y(t):=X+tH+\frac{t^2}{2}W$. 令 $Y(t)$ 具有如下的特征值分解:

$$Y(t)=U(t)\Xi(t)U(t)^{\mathrm{T}}, \quad U(t)\in\mathcal{O}^n(Y(t)), \tag{2.3}$$

其中 $\Xi(t) := \Lambda(Y(t))$.

本节只考虑 $X = \Lambda(X)$ 的情形, 即 X 是 $\mathcal{S}^n$ 中第 i 个对角元为 $\lambda_i(X)$ 的对角矩阵, $i = 1, \ldots, n$. 显然, 单位矩阵 $I \in \mathcal{O}^n(X)$. 不失一般性, 假设 $P = I$, 则

$$P_{\alpha_i} = E_{\alpha_i} = (e_{\kappa_{i-1}+1} \cdots e_{\kappa_i}) \in \Re^{n \times |\alpha_i|},$$

其中, e_j 是 $\Re^n$ 中的第 j 个单位向量, 并且 $\kappa_i := \sum_{j=1}^{i} |\alpha_j|$. 因此, 有 $P_{\alpha_k}^{\mathrm{T}} H P_{\alpha_l} = H_{\alpha_k \alpha_l}$.

对于每一个 $i \in \{1, \ldots, n\}$, $\lambda_i(X)$ 表示矩阵 X 的第 i 大的特征值, $l_i(X)$ 和 $s_i(X)$ 满足

$$\begin{aligned} \lambda_1(X) \geqslant \cdots \geqslant \lambda_{i-l_i(X)}(X) > \lambda_{i-l_i(X)+1}(X) = \cdots = \lambda_i(X) = \cdots = \lambda_{i+s_i(X)}(X) \\ > \lambda_{i+s_i(X)+1}(X) \geqslant \cdots \geqslant \lambda_n(X). \end{aligned}$$

在根据上下文不引起歧义的情况下, 我们将 $l_i(X)$ 和 $s_i(X)$ 分别简记为 l_i 和 s_i.

引理 2.1[6, 命题 1 和 2] 任给 $H \in \mathcal{S}^n$, 令 $U \in \Re^{m \times n}$ 为正交矩阵且满足

$$U^{\mathrm{T}}(\Lambda(X) + H)U = \Lambda(\Lambda(X) + H).$$

则任给 $H \to 0$, 有

$$\begin{cases} U_{\alpha_k \alpha_l} = \mathrm{O}(\|H\|), & k, l = 1, \ldots, r, \quad k \neq l, \\ U_{\alpha_k \alpha_k} U_{\alpha_k \alpha_k}^{\mathrm{T}} = I_{|\alpha_k|} + \mathrm{O}(\|H\|^2), & k = 1, \ldots, r, \\ \mathrm{dist}(U_{\alpha_k \alpha_k}, \mathcal{O}^{|\alpha_k|}) = \mathrm{O}(\|H\|^2), & k = 1, \ldots, r, \end{cases}$$

并且

$$\lambda_i(\Lambda(X) + H) - \lambda_i(X) - \lambda_{l_i}(H_{\alpha_k \alpha_k}) = \mathrm{O}(\|H\|^2), \quad i \in \alpha_k, \quad k = 1, \ldots, r.$$

因此, 对于任意给定的方向 $H \in \mathcal{S}^n$, 特征值函数 $\lambda_i(\cdot)$ 在 X 处方向可微且 $\lambda_i'(\Lambda(X); H) = \lambda_{l_i}(H_{\alpha_k \alpha_k})$, $i \in \alpha_k$, $k = 1, \ldots, r$.

引理 2.1 中的特征值函数 $\lambda(\cdot)$ 的方向导数可以在很多文献中找到, 例如 [9, 定理 7] 和 [10, 命题 1.4].

由 (2.3) 可知,

$$\Xi(t) = U(t)^{\mathrm{T}} Y(t) U(t) = U(t)^{\mathrm{T}} \left(X + t \left(H + \frac{t}{2} W \right) \right) U(t). \tag{2.4}$$

又由引理 2.1 有对于任意 $k, l \in \{1, \ldots, r\}$, $k \neq l$, 存在 $Q^k(t) \in \mathcal{O}^{|\alpha_k|}$ 满足

$$U_{\alpha_k \alpha_k}(t) = Q^k(t) + \mathrm{O}(t^2), \quad U_{\alpha_k \alpha_l}(t) = O(t).$$

结合 (2.4), 可得

$$\Xi_{\alpha_k}(t) = \mu_k I_{|\alpha_k|} + \mathrm{O}(t).$$

因此

$$(\Xi_{\alpha_k}(t) - \mu_l I_{|\alpha_k|})^{-1} = \frac{1}{\mu_k - \mu_l} I_{|\alpha_k|} + \mathrm{O}(t), \quad l \neq k. \tag{2.5}$$

命题 2.1 令 $Y(t)\in\mathcal{S}^n$ 满足特征值分解 (2.3). 任给 $k,l\in\{1,\dots,r\}$, $k\neq l$, 存在 $Q^l(t)\in\mathcal{O}^{|\alpha_l|}$ 满足

$$U_{\alpha_k\alpha_l}(t)=t\frac{H_{\alpha_k\alpha_l}Q^l(t)}{\mu_l-\mu_k}+\mathrm{O}(t^2), \tag{2.6}$$

$$U_{\alpha_l\alpha_l}(t)^{\mathrm{T}}U_{\alpha_l\alpha_l}(t)=I_{|\alpha_l|}-t^2\sum_{\substack{j=1\\j\neq l}}^{r}\frac{(Q^l(t))^{\mathrm{T}}H^{\mathrm{T}}_{\alpha_j\alpha_l}H_{\alpha_j\alpha_l}Q^l(t)}{(\mu_j-\mu_l)^2}+\mathrm{O}(t^3). \tag{2.7}$$

证明 由 (2.3) 可知 $Y(t)U(t)=U(t)\Xi(t)$, 则

$$E^{\mathrm{T}}_{\alpha_k}Y(t)U(t)E_{\alpha_l}=E^{\mathrm{T}}_{\alpha_k}U(t)\Xi(t)E_{\alpha_l},$$

即

$$\begin{aligned}&E^{\mathrm{T}}_{\alpha_k}\left[\begin{pmatrix}\mu_1I_{|\alpha_1|}&&\\&\ddots&\\&&\mu_rI_{|\alpha_r|}\end{pmatrix}+tH+\frac{t^2}{2}W\right]\begin{pmatrix}U_{\alpha_1\alpha_1}(t)&\cdots&U_{\alpha_1\alpha_r}(t)\\\vdots&\ddots&\vdots\\U_{\alpha_r\alpha_1}(t)&\cdots&U_{\alpha_r\alpha_r}(t)\end{pmatrix}E_{\alpha_l}\\&=E^{\mathrm{T}}_{\alpha_k}\begin{pmatrix}U_{\alpha_1\alpha_1}(t)&\cdots&U_{\alpha_1\alpha_r}(t)\\\vdots&\ddots&\vdots\\U_{\alpha_r\alpha_1}(t)&\cdots&U_{\alpha_r\alpha_r}(t)\end{pmatrix}\begin{pmatrix}\Xi_{\alpha_1}(t)&&\\&\ddots&\\&&\Xi_{\alpha_r}(t)\end{pmatrix}E_{\alpha_l}.\end{aligned}$$

因此,

$$\mu_kU_{\alpha_k\alpha_l}(t)+tE^{\mathrm{T}}_{\alpha_k}HU(t)E_{\alpha_l}+\frac{t^2}{2}E^{\mathrm{T}}_{\alpha_k}WU(t)E_{\alpha_l}=U_{\alpha_k\alpha_l}(t)\Xi_{\alpha_l}(t).$$

由 E_{α_k} 的定义可得

$$E^{\mathrm{T}}_{\alpha_k}H=[H_{\alpha_k\alpha_1}\cdots H_{\alpha_k\alpha_l}\cdots H_{\alpha_k\alpha_r}]\quad 且\quad U(t)E_{\alpha_l}=\begin{pmatrix}U_{\alpha_1\alpha_l}(t)\\\vdots\\U_{\alpha_r\alpha_l}(t)\end{pmatrix},$$

于是有

$$\mu_kU_{\alpha_k\alpha_l}(t)+t\sum_{j=1}^{r}H_{\alpha_k\alpha_j}(t)U_{\alpha_j\alpha_l}(t)+\frac{t^2}{2}E^{\mathrm{T}}_{\alpha_k}WU(t)E_{\alpha_l}=U_{\alpha_k\alpha_l}(t)\Xi_{\alpha_l}(t).$$

结合引理 2.1 和 (2.5), 得到

$$\begin{aligned}U_{\alpha_k\alpha_l}(t)&=\left(t\sum_{j=1}^{r}H_{\alpha_k\alpha_j}(t)U_{\alpha_j\alpha_l}(t)+\frac{t^2}{2}E^{\mathrm{T}}_{\alpha_k}WU(t)E_{\alpha_l}\right)(\Xi_{\alpha_l}(t)-\mu_kI_{|\alpha_l|})^{-1}\\&=\left(t[H_{\alpha_k\alpha_1}\cdots H_{\alpha_k\alpha_l}\cdots H_{\alpha_k\alpha_r}]\begin{pmatrix}\mathrm{O}(t)\\\vdots\\Q^l(t)+\mathrm{O}(t^2)\\\vdots\\\mathrm{O}(t)\end{pmatrix}+\frac{t^2}{2}E^{\mathrm{T}}_{\alpha_k}WU(t)E_{\alpha_l}\right)(\Xi_{\alpha_l}(t)-\mu_kI_{|\alpha_l|})^{-1}\end{aligned}$$

$$
\begin{aligned}
&= (tH_{\alpha_k\alpha_l}Q^l(t) + \mathrm{O}(t^2))\left(\frac{1}{\mu_l - \mu_k} I_{|\alpha_l|} + \mathrm{O}(t)\right) \\
&= \frac{tH_{\alpha_k\alpha_l}Q^l(t)}{\mu_l - \mu_k} + \mathrm{O}(t^2),
\end{aligned}
$$

公式 (2.6) 得证. 注意到 $U_{\alpha_l}^{\mathrm{T}}(t)U_{\alpha_l}(t) = I_{|\alpha_l|}$, 公式 (2.7) 得证. □

2.2 非对称的情形

从本节开始, 我们考虑非对称的情况. 给定 $X \in \Re^{m\times n}$, 定义 $\sigma_1(X) \geqslant \sigma_2(X) \geqslant \cdots \geqslant \sigma_m(X)$ 为矩阵 X 按非增顺序排列的奇异值. 令

$$
\sigma(X) := (\sigma_1(X), \sigma_2(X), \ldots, \sigma_m(X))^{\mathrm{T}} \in \Re^m \quad 且 \quad \Sigma(X) := \mathrm{diag}\,(\sigma(X)).
$$

令 $X \in \Re^{m\times n}$ 具有如下的奇异值分解:

$$
X = \overline{U}[\Sigma(X)\ \ 0]\overline{V}^{\mathrm{T}}, \tag{2.8}
$$

其中 $\overline{U} \in \mathcal{O}^m$, $\overline{V} \in \mathcal{O}^n$. 将所有满足奇异值分解 (2.8) 的矩阵对 $(\overline{U}, \overline{V})$ 的全体记作 $\mathcal{O}^{m,n}(X)$, 即

$$
\mathcal{O}^{m,n}(X) = \{(U, V) \in \mathcal{O}^m \times \mathcal{O}^n \mid X = U[\Sigma(X)\ \ 0]V^{\mathrm{T}}\}.
$$

定义指标集合 α, β 和 β_0:

$$
\alpha := \{i \mid \sigma_i(X) > 0, 1 \leqslant i \leqslant m\}, \quad \beta := \{i \mid \sigma_i(X) = 0, 1 \leqslant i \leqslant m\}, \quad \beta_0 := \{m+1, \ldots, n\}.
$$

令 $\mu_1 > \mu_2 > \cdots > \mu_r > \mu_{r+1} = 0$ 为 X 的不同奇异值. 定义如下的指标集合:

$$
\alpha_k = \{i \mid \sigma_i(X) = \mu_k, 1 \leqslant i \leqslant m\}, \quad k = 1, \ldots, r+1, \quad \widehat{\beta} = \beta \cup \beta_0. \tag{2.9}
$$

显然, $\alpha = \bigcup_{k=1}^r \alpha_k$, $\beta = \alpha_{r+1}$.

将矩阵 $\overline{U}$ 划分为 $\overline{U} = [\overline{U}_{\alpha_1}\, \overline{U}_{\alpha_2}\, \cdots\, \overline{U}_{\alpha_r}\, \overline{U}_{\alpha_{r+1}}]$, 其中, $\overline{U}_{\alpha_k} \in \Re^{m\times|\alpha_k|}$, $k = 1, \ldots, r+1$. 类似地, $\overline{V} = [\overline{V}_{\alpha_1}\, \overline{V}_{\alpha_2}\, \cdots\, \overline{V}_{\alpha_r}\, \overline{V}_{\widehat{\beta}}]$, 其中, $\overline{V}_{\alpha_k} \in \Re^{n\times|\alpha_k|}$, $\overline{V}_{\widehat{\beta}} \in \Re^{n\times|\widehat{\beta}|}$, $k = 1, \ldots, r$.

为了符号的简便, 任给 $H, W \in \Re^{m\times n}$, 定义

$$
\overline{H} := \overline{U}^{\mathrm{T}} H \overline{V}, \quad \overline{W} := \overline{U}^{\mathrm{T}} W \overline{V}, \quad G_t := H + \frac{t}{2}W, \quad \overline{G}_t := [\overline{G}_t^1\ \ \overline{G}_t^2] = \overline{H} + \frac{t}{2}\overline{W},
$$

其中, $\overline{G}_t^1 \in \Re^{m\times m}$, $\overline{G}_t^2 \in \Re^{m\times(n-m)}$. 按如下方式定义线性算子 $\mathscr{S} : \Re^{p\times p} \to \mathcal{S}^p$:

$$
\mathscr{S}(A) := \frac{1}{2}(A + A^{\mathrm{T}}), \quad \forall\, A \in \Re^{p\times p}.
$$

对于 $k \in \{1, \ldots, r\}$, 假设 $\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}) \in \mathcal{S}^{|\alpha_k|}$ 具有如下的奇异值分解:

$$
(Q^k)^{\mathrm{T}} \mathscr{S}(\overline{H}_{\alpha_k\alpha_k}) Q^k = \mathrm{diag}\,(\xi_1^k, \ldots, \xi_{|\alpha_k|}^k),
$$

其中, $Q^k \in \mathcal{O}^{|\alpha_k|}$ 且 $\xi_i^k = \lambda_i(\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}))$, $i = 1, \ldots, |\alpha_k|$. 令 $\eta_1^k, \ldots, \eta_{N_k}^k$ 为 $\mathscr{S}(\overline{H}_{\alpha_k\alpha_k})$ 的不同的奇异值. 定义 $\{1, \ldots, |\alpha_k|\}$ 的子集合

$$
\beta_j^k := \{i \mid \xi_i^k = \eta_j^k, i = 1, \ldots, |\alpha_k|\}, \quad j = 1, \ldots, N_k. \tag{2.10}
$$

对于 $i\in\beta$, 令 $\eta_1^{r+1}>\eta_2^{r+1}>\cdots>\eta_{N_{r+1}}^{r+1}>\eta_{N_{r+1}+1}^{r+1}=0$ 为 $\overline{H}_{\beta\widehat{\beta}}$ 的不同奇异值, 定义

$$\beta_j^{r+1}=\{i\mid\sigma_i(\overline{H}_{\beta\widehat{\beta}})=\eta_j^{r+1},\ i=1,\ldots,|\beta|\},\quad j=1,\ldots,N_{r+1}+1.$$

令

$$\kappa_i:=\sum_{j=1}^{i}|\alpha_j|(\kappa_0:=0),\quad \kappa_i^{(k)}:=\sum_{j=1}^{i}|\beta_j^k|(\kappa_0^{(k)}:=0),$$

为了符号的简便, 定义如下映射:

$$\begin{aligned}
&q_a:\{1,\ldots,m\}\to\{1,\ldots,r+1\},\quad q_a(i)=k,\quad \text{若 } i\in\alpha_k,\\
&l:\{1,\ldots,m\}\to\mathbb{N},\quad l(i)=i-\kappa_{q_a(i)-1},\\
&q_b:\{1,\ldots,m\}\to\mathbb{N},\quad q_b(i)=p,\quad \text{若 } l(i)\in\beta_p^{q_a(i)},\\
&l':\{1,\ldots,m\}\to\mathbb{N},\quad l'(i)=l(i)-\kappa_{q_b(i)-1}^{(q_a(i))}.
\end{aligned}$$

例 2.1 令 $X=[\Sigma(X)\ 0]\in\Re^{8\times9}$, 其中, $\Sigma(X)=\mathrm{diag}\,(9,9,8,8,8,0,0,0)$. 假设 $U=I_8\in\mathcal{O}^8$, $V=I_9\in\mathcal{O}^9$, 则 $\alpha_1=\{1,2\}$, $\alpha_2=\{3,4,5\}$, $\beta=\{6,7,8\}$, $\widehat{\beta}=\{6,7,8,9\}$. 令 $H\in\Re^{8\times9}$ 满足 $\mathscr{S}(U_{\alpha_1}^{\mathrm{T}}HV_{\alpha_1})=\mathrm{diag}\,(4,4)$, $\mathscr{S}(U_{\alpha_2}^{\mathrm{T}}HV_{\alpha_2})=\mathrm{diag}\,(3,2,2)$, $U_\beta^{\mathrm{T}}HV_{\widehat{\beta}}=[\Sigma\ 0]\in\Re^{3\times4}$, 且 $\Sigma=\mathrm{diag}\,(1,1,0)$, 则 $\beta_1^1=\{1,2\}$, $\beta_1^2=\{1\}$, $\beta_2^2=\{2,3\}$, $\beta_1^3=\{1,2\}$, $\beta_2^3=\{3\}$. 若 $Q\in\mathcal{O}^3(\mathscr{S}(U_{\alpha_2}^{\mathrm{T}}HV_{\alpha_2}))$, 则 $Q_{\beta_1^2}^{\mathrm{T}}WQ_{\beta_1^2}=5$, $Q_{\beta_2^2}^{\mathrm{T}}WQ_{\beta_2^2}=\mathrm{diag}\,(4,3)$, 存在 $W\in\mathcal{S}^3$. 对于 $i=5$, 有 $i\in\alpha_2$, 则 $q_a(i)=2$, $l(i)=5-\kappa_1=3\in\beta_2^2$, $q_b(i)=2$ 且 $l'(i)=3-\kappa_1^{(2)}=2$. 因此, $\lambda_i(X)=8$, $\lambda_{l(i)}(\mathscr{S}(U_{\alpha_2}^{\mathrm{T}}HV_{\alpha_2}))=2$, $\lambda_{l'(i)}(Q_{\beta_2^2}^{\mathrm{T}}WQ_{\beta_2^2})=3$.

令 $Y(t)=X+tG_t$, 则 $\overline{Y}(t):=\overline{U}^{\mathrm{T}}Y(t)\overline{V}=[\Sigma(X)\ \ 0]+t\overline{G}_t$. 因为 $Y(t)$ 与 $\overline{Y}(t)$ 具有相同的奇异值, 故在接下来的讨论中, 我们只需考虑 $\overline{Y}(t)$. 设 $(U(t),V(t))\in\mathcal{O}^{m,n}(\overline{Y}(t))$, 由 [11, 定理 7.3.5] 可知, $U(t)\in\mathcal{O}^m(G_U(t))$ 且 $V(t)\in\mathcal{O}^n(G_V(t))$, 其中, $G_U(t):=\overline{Y}(t)\overline{Y}(t)^{\mathrm{T}}$, $G_V(t):=\overline{Y}(t)^{\mathrm{T}}\overline{Y}(t)$.

对于每一个 $i\in\{1,\ldots,m\}$ 都有 $\lambda_i(G_U(t))=\sigma_i^2(X)$ 和

$$G_U(t)=\Sigma(X)^2+t[\Sigma(X)(\overline{G}_t^1)^{\mathrm{T}}+\overline{G}_t^1\Sigma(X)+t\overline{G}_t^1(\overline{G}_t^1)^{\mathrm{T}}].\tag{2.11}$$

对于 $k,l\in\{1,\ldots,r\}$, $k\neq l$, 由命题 2.1 和 $\overline{G}_t^1$ 的定义可得

$$U_{\alpha_k\alpha_l}(t)=\frac{t}{\mu_l^2-\mu_k^2}(\mu_k\overline{H}_{\alpha_l\alpha_k}^{\mathrm{T}}+\mu_l\overline{H}_{\alpha_k\alpha_l})U_{\alpha_l\alpha_l}(t)+\mathrm{O}(t^2).\tag{2.12}$$

类似地, 对 $G_V(t)$ 运用命题 2.1 可得

$$V_{\alpha_k\alpha_l}(t)=\frac{t}{\mu_l^2-\mu_k^2}(\mu_k\overline{H}_{\alpha_k\alpha_l}+\mu_l\overline{H}_{\alpha_l\alpha_k}^{\mathrm{T}})V_{\alpha_l\alpha_l}(t)+\mathrm{O}(t^2).\tag{2.13}$$

由引理 2.1 可知, 存在 $Q_\beta\in\mathcal{O}^{|\beta|}$, $\widehat{Q}_{\widehat{\beta}\widehat{\beta}}\in\mathcal{O}^{|\widehat{\beta}|}$ 分别满足

$$U_{\beta\beta}(t)=Q_{\beta\beta}(t)+\mathrm{O}(t^2)\quad\text{和}\quad V_{\widehat{\beta}\widehat{\beta}}(t)=\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\mathrm{O}(t^2),\tag{2.14}$$

于是, 任给 $k\leqslant r$ 有

$$U_{\alpha_k\beta}(t)=-\frac{t}{\mu_k}\overline{H}_{\beta\alpha_k}^{\mathrm{T}}Q_{\beta\beta}(t)+\mathrm{O}(t^2),\tag{2.15}$$

$$V_{\alpha_k\widehat{\beta}}(t)=-\frac{t}{\mu_k}\overline{H}_{\alpha_k\widehat{\beta}}\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\mathrm{O}(t^2).\tag{2.16}$$

3 奇异值函数的二阶方向导数

令 $\mathcal{X}$ 和 $\mathcal{Y}$ 为有限维的 Hilbert 空间, 称映射 $G:\mathcal{X}\to\mathcal{Y}$ 在 $\overline{x}\in\mathcal{X}$ 处是二阶方向可微的 (参见 [12] 或者 [8]), 若 G 在 $\overline{x}$ 处是方向可微的, 并且对任意 $h,w\in\mathcal{X}$, 下述极限存在:

$$\lim_{t\downarrow 0}\frac{G(\overline{x}+th+\frac{1}{2}t^2w)-G(\overline{x})-tG'(\overline{x};h)}{\frac{1}{2}t^2}.$$

上述极限称为二阶方向导数, 记 $G''(\overline{x};h,w)$.

我们首先讨论奇异值函数在非零点处的一阶和二阶方向导数. 令 $\mathcal{B}(\cdot):\Re^{m\times n}\to\mathcal{S}^{m+n}$ 为如下定义的线性算子:

$$\mathcal{B}(X):=\begin{bmatrix}0 & X\\ X^{\mathrm{T}} & 0\end{bmatrix},\quad X\in\Re^{m\times n}.$$

定义 $(m+n)\times(m+n)$ 矩阵 $\overline{P}$,

$$\overline{P}=\frac{1}{\sqrt{2}}\begin{bmatrix}I_{|\alpha|} & 0 & 0 & 0 & I_{|\alpha|}^{\uparrow}\\ 0 & I_{|\beta|} & 0 & I_{|\beta|}^{\uparrow} & 0\\ I_{|\alpha|} & 0 & 0 & 0 & -I_{|\alpha|}^{\uparrow}\\ 0 & I_{|\beta|} & 0 & -I_{|\beta|}^{\uparrow} & 0\\ 0 & 0 & \sqrt{2}I_{n-m} & 0 & 0\end{bmatrix},$$

其中

$$I_p^{\uparrow}=\begin{bmatrix}0 & 0 & \dots & 0 & 1\\ 0 & 0 & \dots & 1 & 0\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 1 & \dots & 0 & 0\\ 1 & 0 & \dots & 0 & 0\end{bmatrix}\in\Re^{p\times p}.$$

于是 $\overline{P}\in\mathcal{O}^{m+n}(\mathcal{B}(\overline{X}))$, 即

$$\overline{P}^{\mathrm{T}}\mathcal{B}(\overline{X})\overline{P}=\begin{bmatrix}\Sigma_\alpha(\overline{X}) & & & & \\ & 0_\beta & & & \\ & & 0_{n-m} & & \\ & & & 0_\beta & \\ & & & & -\Sigma_\alpha(\overline{X})^{\uparrow}\end{bmatrix},$$

其中, $\overline{X}:=\overline{U}^{\mathrm{T}}X\overline{V}=[\Sigma(X)\ \ 0]$, $\Sigma_\alpha(\overline{X})^{\uparrow}=\mathrm{diag}\,(\mu_rI_{|\alpha_r|},\dots,\mu_1I_{|\alpha_1|})$.

令 $\overline{g}_t := \overline{P}^{\mathrm{T}}\mathcal{B}(\overline{G}_t)\overline{P}$, 则有

$$\overline{P}^{\mathrm{T}}\mathcal{B}(\overline{X}+t\overline{G}_t)\overline{P} = \begin{bmatrix} \Sigma_\alpha(\overline{X}) & & & & \\ & 0_\beta & & & \\ & & 0_{n-m} & & \\ & & & 0_\beta & \\ & & & & -\Sigma_\alpha(\overline{X})^{\uparrow} \end{bmatrix} + t\overline{g}_t\,.$$

令 $S(t)\in\mathcal{O}^{m+n}(\overline{P}^{\mathrm{T}}\mathcal{B}(\overline{X}+t\overline{G}_t)\overline{P})$, 即

$$\overline{P}^{\mathrm{T}}\mathcal{B}(\overline{X}+t\overline{G}_t)\overline{P} = S(t)\Xi(t)S(t)^{\mathrm{T}}, \quad \Xi(t) := \Lambda(\overline{P}^{\mathrm{T}}\mathcal{B}(\overline{X}+t\overline{G}_t)\overline{P}). \tag{3.1}$$

令 $\beta^+ = \{n+1,\ldots,n+|\beta|\}$, 定义 $\overline{\beta} = \widehat{\beta}\cup\beta^+$. 为了方便接下来的讨论, 令

$$S(t) = [S_{\alpha_1}(t)\ \ S_{\alpha_2}(t)\cdots S_{\alpha_r}(t)\ \ S_{\overline{\beta}}(t)\ \ S_{\alpha_r^+}(t)\ \ S_{\alpha_{r-1}^+}(t)\cdots S_{\alpha_1^+}(t)],$$

其中

$$\alpha_k^+ = \left\{m+n-\sum_{l=0}^{k}|\alpha_l|+1,\ldots,m+n-\sum_{l=0}^{k-1}|\alpha_l|\right\},$$

$\alpha_0 := 0$, 显然, $|\alpha_k^+| = |\alpha_k|$, $k=1,\ldots,r$. 由引理 2.1 可知, 任意给定 $k=1,\ldots,r$, 存在

$$Q^k(t)\in\mathcal{O}^{|\alpha_k|}, \quad Q_+^k(t)\in\mathcal{O}^{|\alpha_k|}, \quad Q^{\overline{\beta}}(t)\in\mathcal{O}^{|\overline{\beta}|}$$

满足

$$S_{\alpha_k\alpha_k}(t) = Q^k(t)+\mathrm{O}(t^2), \quad S_{\alpha_k^+\alpha_k^+}(t) = Q_+^k(t)+\mathrm{O}(t^2), \quad S_{\overline{\beta}\overline{\beta}}(t) = Q^{\overline{\beta}}(t)+\mathrm{O}(t^2).$$

任给 $l\leqslant r$, 由命题 2.1 知, 存在 $Q^l(t)\in\mathcal{O}^{|\alpha_l|}$ 满足

$$\begin{aligned} S_{\alpha_k\alpha_l}(t) &= t\frac{(\overline{g}_t)_{\alpha_k\alpha_l}Q^l(t)}{\mu_l-\mu_k}+\mathrm{O}(t^2), \quad k\neq l, \quad k\leqslant r,\\ S_{\overline{\beta}\alpha_l}(t) &= t\frac{(\overline{g}_t)_{\overline{\beta}\alpha_l}Q^l(t)}{\mu_l}+\mathrm{O}(t^2),\\ S_{\alpha_k^+\alpha_l}(t) &= t\frac{(\overline{g}_t)_{\alpha_k^+\alpha_l}Q^l(t)}{\mu_l+\mu_k}+\mathrm{O}(t^2), \quad k=r,r-1,\ldots,1, \quad k\neq l,\\ S_{\alpha_l\alpha_l}(t)^{\mathrm{T}}S_{\alpha_l\alpha_l}(t) &= I_{|\alpha_l|}-t^2\sum_{\substack{j=1\\ j\neq l}}^{r+1}\frac{(Q^l(t))^{\mathrm{T}}(\overline{g}_t)_{\alpha_j\alpha_l}^{\mathrm{T}}(\overline{g}_t)_{\alpha_j\alpha_l}Q^l(t)}{(\mu_j-\mu_l)^2}\\ &\quad -t^2\sum_{j=1}^{r}\frac{(Q^l(t))^{\mathrm{T}}(\overline{g}_t)_{\alpha_j^+\alpha_l}^{\mathrm{T}}(\overline{g}_t)_{\alpha_j^+\alpha_l}Q^l(t)}{(\mu_j+\mu_l)^2}+\mathrm{O}(t^3). \end{aligned}$$

于是, 存在 $Q^k(t)\in\mathcal{O}^{|\alpha_k|}$ 使得

$$S_{\alpha_k}(t)^{\mathrm{T}}\begin{bmatrix} \Sigma_\alpha(\overline{X}) & & \\ & 0_{2\beta+n-m} & \\ & & -\Sigma_\alpha(\overline{X})^{\uparrow} \end{bmatrix}S_{\alpha_k}(t)$$

$$
\begin{aligned}
&= \sum_{j=1}^{r} \mu_j S_{\alpha_j\alpha_k}(t)^{\mathrm{T}} S_{\alpha_j\alpha_k}(t) - \sum_{j=1}^{r} \mu_j S_{\alpha_j^+\alpha_k}(t)^{\mathrm{T}} S_{\alpha_j^+\alpha_k}(t)\\
&= \mu_k S_{\alpha_k\alpha_k}(t)^{\mathrm{T}} S_{\alpha_k\alpha_k}(t) + \sum_{\substack{j=1\\ j\neq k}}^{r} \mu_j S_{\alpha_j\alpha_k}(t)^{\mathrm{T}} S_{\alpha_j\alpha_k}(t) - \sum_{j=1}^{r} \mu_j S_{\alpha_j^+\alpha_k}(t)^{\mathrm{T}} S_{\alpha_j^+\alpha_k}(t)\\
&= \mu_k I_{|\alpha_k|} - \mu_k t^2 Q^k(t)^{\mathrm{T}} \Bigg[\sum_{\substack{j\neq k\\ j\leqslant r+1}} \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j\alpha_k}(\overline{g}_t)_{\alpha_j\alpha_k}}{(\mu_j-\mu_k)^2} + \sum_{j\leqslant r} \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j^+\alpha_k}(\overline{g}_t)_{\alpha_j^+\alpha_k}}{(\mu_j+\mu_k)^2} \Bigg] Q^k(t)\\
&\quad + t^2 Q^k(t)^{\mathrm{T}} \Bigg[\sum_{\substack{j\neq k\\ j\leqslant r}} \mu_j \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j\alpha_k}(\overline{g}_t)_{\alpha_j\alpha_k}}{(\mu_j-\mu_k)^2} - \sum_{j\leqslant r} \mu_j \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j^+\alpha_k}(\overline{g}_t)_{\alpha_j^+\alpha_k}}{(\mu_j+\mu_k)^2} \Bigg] Q^k(t) + \mathrm{O}(t^3)\\
&= \mu_k I_{|\alpha_k|} + t^2 Q^k(t)^{\mathrm{T}} \Bigg[\sum_{\substack{j\neq k\\ j\leqslant r+1}} \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j\alpha_k}(\overline{g}_t)_{\alpha_j\alpha_k}}{\mu_j-\mu_k} - \sum_{j\leqslant r} \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j^+\alpha_k}(\overline{g}_t)_{\alpha_j^+\alpha_k}}{\mu_j+\mu_k} \Bigg] Q^k(t) + \mathrm{O}(t^3). \qquad (3.2)
\end{aligned}
$$

同时还有

$$
\begin{aligned}
S_{\alpha_k\alpha_k}(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_k\alpha_k}S_{\alpha_k\alpha_k}(t) &= Q^k(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_k\alpha_k}Q^k(t) + \mathrm{O}(t^2)\\
&= \frac{1}{2}Q^k(t)\left(\overline{H}_{\alpha_k\alpha_k} + \overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k} + \frac{t}{2}(\overline{W}_{\alpha_k\alpha_k} + \overline{W}^{\mathrm{T}}_{\alpha_k\alpha_k})\right) Q^k(t) + \mathrm{O}(t^2), \qquad (3.3)
\end{aligned}
$$

$$
\begin{aligned}
&S_{\alpha_k\alpha_k}(t)^{\mathrm{T}}\Bigg(\sum_{\substack{j\neq k\\ j\leqslant r+1}} (\overline{g}_t)_{\alpha_k\alpha_j}S_{\alpha_j\alpha_k}(t) + \sum_{j\leqslant r}(\overline{g}_t)_{\alpha_k\alpha_j^+}S_{\alpha_j^+\alpha_k}(t)\Bigg)\\
&\quad = tQ^k(t)^{\mathrm{T}}\Bigg(\sum_{\substack{j\neq k\\ j\leqslant r+1}} \frac{(\overline{g}_t)_{\alpha_k\alpha_j}(\overline{g}_t)_{\alpha_j\alpha_k}}{\mu_k-\mu_j} + \sum_{j\leqslant r}\frac{(\overline{g}_t)_{\alpha_k\alpha_j^+}(\overline{g}_t)_{\alpha_j^+\alpha_k}}{\mu_j+\mu_k}\Bigg)Q^k(t) + \mathrm{O}(t^2), \qquad (3.4)
\end{aligned}
$$

$$
\begin{aligned}
&\Bigg(\sum_{\substack{j\neq k\\ j\leqslant r+1}} S_{\alpha_j\alpha_k}(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_j\alpha_k} + \sum_{j\leqslant r} S_{\alpha_j^+\alpha_k}(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_j^+\alpha_k}\Bigg)S_{\alpha_k\alpha_k}(t)\\
&\quad = tQ^k(t)^{\mathrm{T}}\Bigg(\sum_{\substack{j\neq k\\ j\leqslant r+1}} \frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j\alpha_k}(\overline{g}_t)_{\alpha_j\alpha_k}}{\mu_k-\mu_j} + \sum_{j\leqslant r}\frac{(\overline{g}_t)^{\mathrm{T}}_{\alpha_j\alpha_k^+}(\overline{g}_t)_{\alpha_j^+\alpha_k}}{\mu_j+\mu_k}\Bigg)Q^k(t) + \mathrm{O}(t^2). \qquad (3.5)
\end{aligned}
$$

通过简单的计算可以得到

$$
\begin{aligned}
tS_{\alpha_k}(t)^{\mathrm{T}}\overline{g}_tS_{\alpha_k}(t) &= tS_{\alpha_k\alpha_k}(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_k\alpha_k}S_{\alpha_k\alpha_k}(t)\\
&\quad + tS_{\alpha_k\alpha_k}(t)^{\mathrm{T}}\Bigg(\sum_{\substack{j\neq k\\ j\leqslant r+1}} (\overline{g}_t)_{\alpha_k\alpha_j}S_{\alpha_j\alpha_k}(t) + \sum_{j\leqslant r}(\overline{g}_t)_{\alpha_k\alpha_j^+}S_{\alpha_j^+\alpha_k}(t)\Bigg)\\
&\quad + t\Bigg(\sum_{\substack{j\neq k\\ j\leqslant r+1}} S_{\alpha_j\alpha_k}(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_j\alpha_k} + \sum_{j\leqslant r} S_{\alpha_j^+\alpha_k}(t)^{\mathrm{T}}(\overline{g}_t)_{\alpha_j^+\alpha_k}\Bigg)S_{\alpha_k\alpha_k}(t) + \mathrm{O}(t^2). \qquad (3.6)
\end{aligned}
$$

注意到 $\overline{g}_t \in \mathcal{S}^{m+n}$, 即 $(\overline{g}_t)^{\mathrm{T}}_{\alpha_j\alpha_k} = (\overline{g}_t)_{\alpha_k\alpha_j}$, $j \leqslant r+1$, 且 $(\overline{g}_t)^{\mathrm{T}}_{\alpha_i^+\alpha_k} = (\overline{g}_t)_{\alpha_k\alpha_i^+}$, $i \leqslant r$. 因为 (3.2)–(3.6) 成

立, 于是有

$$
\begin{aligned}
&S_{\alpha_k}(t)^{\mathrm{T}}\left(\begin{bmatrix}\Sigma_\alpha(\overline{X}) & & \\ & 0_{2\beta+n-m} & \\ & & -\Sigma_\alpha(\overline{X})^{\uparrow}\end{bmatrix}+t\overline{g}_t\right)S_{\alpha_k}(t)\\
&\quad=\mu_k I_{|\alpha_k|}+\frac{t}{2}Q^k(t)^{\mathrm{T}}(\overline{H}_{\alpha_k\alpha_k}+\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k})Q^k(t)+\frac{t^2}{4}Q^k(t)^{\mathrm{T}}(\overline{W}_{\alpha_k\alpha_k}+\overline{W}^{\mathrm{T}}_{\alpha_k\alpha_k})Q^k(t)\\
&\qquad+t^2Q^k(t)^{\mathrm{T}}\Bigg[\sum_{\substack{j\neq k\\ j\leqslant r+1}}\frac{(\overline{g}_t)_{\alpha_k\alpha_j}(\overline{g}_t)_{\alpha_j\alpha_k}}{\mu_k-\mu_j}+\sum_{j\leqslant r}\frac{(\overline{g}_t)_{\alpha_k\alpha_j^+}(\overline{g}_t)_{\alpha_j^+\alpha_k}}{\mu_j+\mu_k}\Bigg]Q^k(t)+\mathrm{O}(t^3)\\
&\quad=\mu_k I_{|\alpha_k|}+\frac{t}{2}Q^k(t)^{\mathrm{T}}(\overline{H}_{\alpha_k\alpha_k}+\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k})Q^k(t)+\frac{t^2}{4}Q^k(t)^{\mathrm{T}}(\overline{W}_{\alpha_k\alpha_k}+\overline{W}^{\mathrm{T}}_{\alpha_k\alpha_k})Q^k(t)\\
&\qquad+t^2(Q^k(t))^{\mathrm{T}}\Bigg[\sum_{\substack{j\neq k\\ j\leqslant r}}\Bigg\{\frac{1}{4(\mu_k-\mu_j)}(\overline{H}_{\alpha_k\alpha_j}\overline{H}_{\alpha_j\alpha_k}+\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}_{\alpha_j\alpha_k}+\overline{H}_{\alpha_k\alpha_j}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j}+\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j})\\
&\qquad+\frac{1}{4(\mu_k+\mu_j)}(-\overline{H}_{\alpha_k\alpha_j}\overline{H}_{\alpha_j\alpha_k}+\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}_{\alpha_j\alpha_k}+\overline{H}_{\alpha_k\alpha_j}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j}-\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j})\Bigg\}\\
&\qquad+\frac{1}{\mu_k}(\overline{g}_t)_{\alpha_k\overline{\beta}}(\overline{g}_t)_{\overline{\beta}\alpha_k}+\frac{1}{8\mu_k}(\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k}-\overline{H}_{\alpha_k\alpha_k})(\overline{H}_{\alpha_k\alpha_k}-\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k})\Bigg]Q^k(t)+\mathrm{O}(t^3)\\
&\quad=\mu_k I_{|\alpha_k|}+\frac{t}{2}Q^k(t)^{\mathrm{T}}(\overline{H}_{\alpha_k\alpha_k}+\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k})Q^k(t)+\frac{t^2}{4}Q^k(t)^{\mathrm{T}}(\overline{W}_{\alpha_k\alpha_k}+\overline{W}^{\mathrm{T}}_{\alpha_k\alpha_k})Q^k(t)\\
&\qquad+t^2(Q^k(t))^{\mathrm{T}}\Bigg[\sum_{\substack{j\neq k\\ j\leqslant r}}\frac{\mu_j\overline{H}_{\alpha_k\alpha_j}\overline{H}_{\alpha_j\alpha_k}+\mu_k\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}_{\alpha_j\alpha_k}+\mu_k\overline{H}_{\alpha_k\alpha_j}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j}+\mu_j\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j}}{2(\mu_k^2-\mu_j^2)}\\
&\qquad+\frac{1}{\mu_k}(\overline{g}_t)_{\alpha_k\overline{\beta}}(\overline{g}_t)_{\overline{\beta}\alpha_k}+\frac{1}{8\mu_k}(\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k}-\overline{H}_{\alpha_k\alpha_k})(\overline{H}_{\alpha_k\alpha_k}-\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k})\Bigg]Q^k(t)+\mathrm{O}(t^3).
\end{aligned}\tag{3.7}
$$

对每一个 $k\leqslant r$, 定义

$$
\begin{aligned}
V_k(H,W)=&\frac{\overline{W}_{\alpha_k\alpha_k}+\overline{W}^{\mathrm{T}}_{\alpha_k\alpha_k}}{2}+\frac{1}{\mu_k}\left(\frac{\overline{H}_{\alpha_k\alpha_k}-\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k}}{2}\right)^{\mathrm{T}}\left(\frac{\overline{H}_{\alpha_k\alpha_k}-\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_k}}{2}\right)\\
&+\frac{1}{\mu_k}(\overline{H}^{\mathrm{T}}_{\beta\alpha_k}\overline{H}_{\beta\alpha_k}+\overline{H}_{\alpha_k\beta}\overline{H}^{\mathrm{T}}_{\alpha_k\beta}+\overline{H}_{\alpha_k\beta_0}\overline{H}^{\mathrm{T}}_{\alpha_k\beta_0})\\
&+\sum_{\substack{j\neq k\\ j\leqslant r}}\frac{\mu_j\overline{H}_{\alpha_k\alpha_j}\overline{H}_{\alpha_j\alpha_k}+\mu_k\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}_{\alpha_j\alpha_k}+\mu_k\overline{H}_{\alpha_k\alpha_j}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j}+\mu_j\overline{H}^{\mathrm{T}}_{\alpha_j\alpha_k}\overline{H}^{\mathrm{T}}_{\alpha_k\alpha_j}}{\mu_k^2-\mu_j^2}.
\end{aligned}\tag{3.8}
$$

由 $\overline{g}_t$ 的定义知,

$$
(\overline{g}_t)_{\alpha_k\overline{\beta}}(\overline{g}_t)_{\overline{\beta}\alpha_k}=\frac{1}{2}(\overline{H}^{\mathrm{T}}_{\beta\alpha_k}\overline{H}_{\beta\alpha_k}+\overline{H}_{\alpha_k\beta}\overline{H}^{\mathrm{T}}_{\alpha_k\beta}+\overline{H}_{\alpha_k\beta_0}\overline{H}^{\mathrm{T}}_{\alpha_k\beta_0})+\mathrm{O}(t).
$$

结合 (3.1) 和 (3.7) 可知,

$$
\Xi_{\alpha_k}(t)=\mu_k I_{|\alpha_k|}+t(Q^k(t))^{\mathrm{T}}\mathscr{S}(\overline{H}_{\alpha_k\alpha_k})Q^k(t)+\frac{t^2}{2}(Q^k(t))^{\mathrm{T}}V_k(H,W)Q^k(t)+\mathrm{O}(t^3).\tag{3.9}
$$

于是,

$$
\mathscr{S}(\overline{H}_{\alpha_k\alpha_k})=\frac{1}{t}Q^k(t)\big(\Xi_{\alpha_k}(t)-\mu_k I_{|\alpha_k|}\big)(Q^k(t))^{\mathrm{T}}+\mathrm{O}(t).\tag{3.10}
$$

因为对于任意 $t>0$, $Q^k(t)$ 是一致有界的, 当 $t\searrow 0$ 时, 设 $\{Q^k(t)\}$ 收敛到正交矩阵 Q^k. 于是有

$$\begin{aligned}\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}) &= Q^k \operatorname{diag}(\sigma_i'(X,H): i\in\alpha_k)(Q^k)^{\mathrm{T}} \\ &= Q^k \operatorname{diag}(\eta_1^k I_{|\beta_1^k|},\ldots,\eta_{N_k}^k I_{|\beta_{N_k}^k|})(Q^k)^{\mathrm{T}}.\end{aligned}\tag{3.11}$$

事实上, 当 $t\searrow 0$, $\{Q^k(t)\}$ 的任意聚点 Q^k 都满足 (3.11), 换而言之, $Q^k\in\mathcal{O}^{|\alpha_k|}(\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}))$. 由 (3.9) 可以得到

$$\begin{aligned}&(Q^k)^{\mathrm{T}}Q^k(t)\Xi_{\alpha_k}(t)(Q^k(t))^{\mathrm{T}}Q^k-\mu_k I_{|\alpha_k|}-t\operatorname{diag}(\eta_1^k I_{|\beta_1^k|},\ldots,\eta_{N_k}^k I_{|\beta_{N_k}^k|})\\ &\quad=\frac{t^2}{2}(Q^k)^{\mathrm{T}}V_k(H,W)Q^k+\mathrm{O}(t^3).\end{aligned}\tag{3.12}$$

进一步, 有

$$\Xi_{\alpha_k}(t)=\mu_k I_{|\alpha_k|}+t\Lambda\bigg(\operatorname{diag}(\eta_1^k I_{|\beta_1^k|},\ldots,\eta_{N_k}^k I_{|\beta_{N_k}^k|})+\frac{t}{2}(Q^k)^{\mathrm{T}}V_k(H,W)Q^k\bigg)+\mathrm{O}(t^3).\tag{3.13}$$

利用特征值函数的 Lipschitz 连续性质, 通过 (3.13) 可直接得到以下定理.

定理 3.1 对于任意给定的矩阵 $X\in\Re^{m\times n}$ 满足奇异值分解 (2.8). 令 α_k, $k=1,\ldots,r+1$ 为 (2.9) 定义的指标集合. 任给 $H,W\in\Re^{m\times n}$, 于是, 对于每个 $i\in\alpha_k$, 有

$$\begin{aligned}\sigma_i(Y(t))=&\sigma_i(X)+t\lambda_{l(i)}(\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}))\\ &+\frac{t^2}{2}\lambda_{l'(i)}((Q^k_{\beta^k_{q_b(i)}})^{\mathrm{T}}V_k(H,W)Q^k_{\beta^k_{q_b(i)}})+\mathrm{O}(t^3).\end{aligned}\tag{3.14}$$

定理 3.2 任给 H, $W\in\Re^{m\times n}$, 定义

$$V_{\beta\widehat{\beta}}(H,W)=\overline{W}_{\beta\widehat{\beta}}-2\sum_{k=1}^{r}\mu_k^{-1}\overline{H}_{\beta\alpha_k}\overline{H}_{\alpha_k\widehat{\beta}},$$

于是对于每一个 $i\in\beta$, $\sigma_i(\cdot)$ 在 X 处的一阶方向导数可以表示为

$$\sigma_i'(X;H)=\sigma_{l(i)}(\overline{H}_{\beta\widehat{\beta}}).\tag{3.15}$$

二阶方向导数为, 若 $q_b(i)\in\{1,\ldots,N_{r+1}\}$,

$$\sigma_i''(X;H,W)=\lambda_{l'(i)}(\mathscr{S}(Q^{\mathrm{T}}_{\beta\beta^{r+1}_{q_b(i)}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\beta^{r+1}_{q_b(i)}}));$$

若 $q_b(i)=N_{r+1}+1$,

$$\sigma_i''(X;H,W)=\sigma_{l'(i)}(Q^{\mathrm{T}}_{\beta\beta^{r+1}_{q_b(i)}}V_{\beta\widehat{\beta}}(H,W)[\widehat{Q}_{\widehat{\beta}\beta^{r+1}_{q_b(i)}}\ \ \widehat{Q}_{\widehat{\beta}\beta_0}]).\tag{3.16}$$

证明 任给方向 $H,W\in\Re^{m\times n}$ 和 $t>0$, 令 $Y(t)=X+tG_t$ 具有如下的奇异值分解:

$$Y(t)=\widetilde{U}(t)[\Sigma(X+tG_t)\ \ 0]\widetilde{V}(t)^{\mathrm{T}},\quad (\widetilde{U}(t),\widetilde{V}(t))\in\mathcal{O}^{m,n}(Y(t)).\tag{3.17}$$

事实上, 任给 $X\in\Re^{m\times n}$, (3.17) 可以写成

$$U(t)[\Xi(t)\ \ 0]V(t)^{\mathrm{T}}=[\Sigma(X)\ \ 0]+t\overline{G}_t,$$

其中 $\Xi(t)=\Sigma(X+tG_t)$, $U(t)=\overline{U}^{\mathrm{T}}\widetilde{U}(t)$, $V(t)=\overline{V}\widetilde{V}(t)^{\mathrm{T}}$. 于是

$$[\Xi_\beta(t)\ \ 0_{|\beta|\times|\beta_0|}]=U_\beta(t)^{\mathrm{T}}([\Sigma(X)\ \ 0]+t\overline{G}_t)V_{\widehat{\beta}}(t). \tag{3.18}$$

通过 (2.14), (2.15) 和 (2.16) 可知,

$$\begin{aligned}U_\beta(t)^{\mathrm{T}}[\Sigma(X)\ \ 0]V_{\widehat{\beta}}(t)&=U_{\alpha\beta}(t)^{\mathrm{T}}\Sigma_\alpha(X)V_{\alpha\widehat{\beta}}(t)\\&=t^2Q_{\beta\beta}(t)^{\mathrm{T}}\left(\sum_{k=1}^r\mu_k^{-1}\overline{H}_{\beta\alpha_k}\overline{H}_{\alpha_k\widehat{\beta}}\right)\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\mathrm{O}(t^3),\\tU_\beta(t)^{\mathrm{T}}\overline{G}_tV_{\widehat{\beta}}(t)&=tQ_{\beta\beta}(t)^{\mathrm{T}}\overline{H}_{\beta\widehat{\beta}}\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\frac{t^2}{2}Q_{\beta\beta}(t)^{\mathrm{T}}\overline{W}_{\beta\widehat{\beta}}\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)\\&\quad-2t^2Q_{\beta\beta}(t)^{\mathrm{T}}\left(\sum_{i=1}^r\mu_k^{-1}\overline{H}_{\beta\alpha_k}\overline{H}_{\alpha_k\widehat{\beta}}\right)\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\mathrm{O}(t^3).\end{aligned}$$

于是 (3.18) 可以写成如下形式:

$$\begin{aligned}[\Xi_\beta(t)\ \ 0_{|\beta|\times|\beta_0|}]&=tQ_{\beta\beta}(t)^{\mathrm{T}}\overline{H}_{\beta\widehat{\beta}}\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)\\&\quad+\frac{t^2}{2}Q_{\beta\beta}(t)^{\mathrm{T}}\left[\overline{W}_{\beta\widehat{\beta}}-2\sum_{k=1}^r\mu_k^{-1}\overline{H}_{\beta\alpha_k}\overline{H}_{\alpha_k\widehat{\beta}}\right]\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\mathrm{O}(t^3).\end{aligned} \tag{3.19}$$

进一步, 显然有

$$\frac{1}{t}[\Xi_\beta(t)-\Xi(0)\ \ 0_{|\beta|\times|\beta_0|}]=Q_{\beta\beta}(t)^{\mathrm{T}}\overline{H}_{\beta\widehat{\beta}}\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)+\mathrm{O}(t). \tag{3.20}$$

假设当 $t\searrow 0$ 时, $(Q_{\beta\beta},\widehat{Q}_{\widehat{\beta}\widehat{\beta}})$ 是 $(Q_{\beta\beta}(t),\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t))$ 的一个聚点, 则 $(Q_{\beta\beta},\widehat{Q}_{\widehat{\beta}\widehat{\beta}})\in\mathcal{O}^{|\beta|\times|\widehat{\beta}|}(\overline{H}_{\beta\widehat{\beta}})$. 因此, 由 (3.20) 可知,

$$[\operatorname{diag}(\sigma_\beta'(X,H))\ \ 0_{|\beta|\times|\beta_0|}]=Q_{\beta\beta}^{\mathrm{T}}\overline{H}_{\beta\widehat{\beta}}\widehat{Q}_{\widehat{\beta}\widehat{\beta}},$$

即 (3.15) 成立.

考虑 (3.19) 有

$$\begin{aligned}[\Xi_\beta(t)\,0_{|\beta|\times|\beta_0|}]&=[\Sigma(Q_{\beta\beta}(t)[\Xi_\beta(t)\,0_{|\beta|\times|\beta_0|}]\widehat{Q}_{\widehat{\beta}\widehat{\beta}}(t)^{\mathrm{T}})\,0_{|\beta|\times|\beta_0|}]\\&=\left[\Sigma\left(t\overline{H}_{\beta\widehat{\beta}}+\frac{t^2}{2}V_{\beta\widehat{\beta}}(H,W)+\mathrm{O}(t^3)\right)\ 0_{|\beta|\times|\beta_0|}\right]\\&=\left[t\Sigma\left(\overline{H}_{\beta\widehat{\beta}}+\frac{t}{2}V_{\beta\widehat{\beta}}(H,W)\right)+\mathrm{O}(t^3)\ 0_{|\beta|\times|\beta_0|}\right]\\&=\left[t\Sigma\left(Q_{\beta\beta}[\operatorname{diag}(\sigma_\beta'(X,H))\ 0_{|\beta|\times|\beta_0|}]\widehat{Q}_{\widehat{\beta}\widehat{\beta}}^{\mathrm{T}}+\frac{t}{2}V_{\beta\widehat{\beta}}(H,W)\right)+\mathrm{O}(t^3)\ 0_{|\beta|\times|\beta_0|}\right]\\&=\left[t\Sigma\left([\operatorname{diag}(\sigma_\beta'(X,H))\ 0_{|\beta|\times|\beta_0|}]+\frac{t}{2}Q_{\beta\beta}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}}\right)+\mathrm{O}(t^3)\ 0_{|\beta|\times|\beta_0|}\right].\end{aligned}$$

于是,

$$\Xi_\beta(t)=t\Sigma\left([\operatorname{diag}(\sigma_\beta'(X,H))\ \ 0_{|\beta|\times|\beta_0|}]+\frac{t}{2}Q_{\beta\beta}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}}\right)+\mathrm{O}(t^3).$$

对于 $j\in\{1,\ldots,|\beta|\}$, 我们考虑如下两个情形.

情形 I 若 $j\in\beta_p^{r+1}$, $p\in\{1,\ldots,N_{r+1}\}$, 由定理 3 可知,

$$\sigma_j\left(\left[\operatorname{diag}(\sigma'_\beta(X,H))\ \ 0_{|\beta|\times|\beta_0|}\right]+\frac{t}{2}Q_{\beta\beta}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}}\right)$$
$$=[\sigma'_\beta(X;H)]_j+\frac{t}{2}\lambda_{l(j)}(\mathscr{S}(Q_{\beta\beta_p^{r+1}}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\beta_p^{r+1}}))+\mathrm{O}(t^2).$$

因此, 若 $p:=q_b(i)\leqslant N_{r+1}$, 对于 $i\in\beta$ 满足 $l(i)\in\beta_p^{r+1}$, 有

$$\sigma_i(Y(t))=t\sigma_{l(i)}(\overline{H}_{\beta\widehat{\beta}})+\frac{t^2}{2}\lambda_{l'(i)}(\mathscr{S}(Q_{\beta\beta_p^{r+1}}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\beta_p^{r+1}}))+\mathrm{O}(t^3).\tag{3.21}$$

情形 II 若 $j\in\beta_{N_{r+1}+1}^{r+1}$, 由 (3.15) 可知,

$$\sigma_j\left([\operatorname{diag}(\sigma'_\beta(X,H))\ \ 0_{|\beta|\times|\beta_0|}]+\frac{t}{2}Q_{\beta\beta}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}}\right)$$
$$=\frac{t}{2}\sigma_{l(j)}(Q_{\beta\beta_{N_{r+1}+1}^{r+1}}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)[\widehat{Q}_{\widehat{\beta}\beta_{N_{r+1}+1}^{r+1}}\ \ \widehat{Q}_{\widehat{\beta}\beta_0}])+\mathrm{O}(t^2).$$

因此, 对于 $i\in\beta$ 满足 $l(i)\in\beta_{N_{r+1}+1}^{r+1}$, 有

$$\sigma_i(Y(t))=\frac{t^2}{2}\sigma_{l'(i)}(Q_{\beta\beta_{N_{r+1}+1}^{r+1}}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)[\widehat{Q}_{\widehat{\beta}\beta_{N_{r+1}+1}^{r+1}}\ \ \widehat{Q}_{\widehat{\beta}\beta_0}])+\mathrm{O}(t^3).\tag{3.22}$$

由 (3.21) 和 (3.22) 可知定理成立. □

4 应用

在这一节, 我们利用奇异值函数的一阶和二阶方向导数分别刻画出核范数上图和谱范数上图切锥和二阶切集. 分别把 $f:=\|\cdot\|_*$ 和 $g:=\|\cdot\|_2$ 的上图定义为凸锥 $\mathcal{K}_*$ 和 $\mathcal{K}_2$, 即

$$\mathcal{K}_*:=\operatorname{epi}f=\{(X;t)\in\Re^{m\times n}\times\Re:\ f(X)\leqslant t\},\tag{4.1}$$
$$\mathcal{K}_2:=\operatorname{epi}g=\{(X;t)\in\Re^{m\times n}\times\Re:\ g(X)\leqslant t\}.\tag{4.2}$$

接下来, 我们利用定理 3.1 和 3.2 得到 $\mathcal{K}_*$ 和 $\mathcal{K}_2$ 的切锥和二阶切集的表示式. 对于 $k=1,\ldots,r+1$, $i=1,\ldots,|\alpha_k|$, 令 α_k, β_i^k, β 和 β_0 为 (2.9) 中相应的集合. 首先回顾一下在凸分析意义下的切锥和二阶切集中的表达式[8]. 设 $\mathcal{Z}$ 为有限维的 Hilbert 空间, $\mathcal{K}$ 为 $\mathcal{Z}$ 中的凸锥. 于是, 可定义 $\mathcal{K}$ 在 $z\in\mathcal{Z}$ 处的切锥 $\mathcal{T}_{\mathcal{K}}(z)$ 为

$$\mathcal{T}_{\mathcal{K}}(z):=\{d\in\mathcal{Z}:\ \exists t_k\downarrow 0\ \text{使得}\ \operatorname{dist}(z+t_kd,\mathcal{K})=o(t_k)\}.\tag{4.3}$$

凸锥 $\mathcal{K}$ 在点 z 沿方向 d 的二阶切集 $\mathcal{T}_{\mathcal{K}}^2(z,d)$ 为

$$\mathcal{T}_{\mathcal{K}}^2(z,d):=\left\{w\in\mathcal{Z}:\ \exists t_k\downarrow 0\ \text{使得}\ \operatorname{dist}\left(z+t_kd+\frac{1}{2}t_k^2w,\mathcal{K}\right)=o(t_k^2)\right\}.\tag{4.4}$$

不失一般性, 我们假设指标集合 α_{r+1} 和 $\beta_{N_{r+1}+1}^{r+1}$ 是非空的. 下面我给出 $\mathcal{T}_{\mathcal{K}_*}$ 和 $\mathcal{T}_{\mathcal{K}_*}^2$ 的表达式.

命题 4.1 令 $(X;t)\in\mathcal{K}_*$ 且 X 满足奇异值分解(2.8). 则

$$\mathcal{T}_{\mathcal{K}_*}(X,t)=\begin{cases}\Re^{m\times n}\times\Re, & (X;t)\in\operatorname{int}\mathcal{K}_*,\\ \mathcal{K}_*, & (X;t)=(0;0),\\ \{(H;\eta)\in\Re^{m\times n}\times\Re:\operatorname{Tr}(\overline{\mathrm{H}}_{\alpha\alpha})+\|\overline{H}_{\beta\widehat{\beta}}\|_*\leqslant\eta\}, & \text{其他},\end{cases}$$

其中 $\overline{H}_{\alpha\alpha}=\overline{U}_\alpha^{\mathrm{T}}H\overline{V}_\alpha$, $\overline{H}_{\beta\widehat{\beta}}=\overline{U}_\beta^{\mathrm{T}}H\overline{V}_{\widehat{\beta}}$.

证明 当 $(X;t)\in\operatorname{int}\mathcal{K}_*$ 和 $(X;t)=(0;0)$ 时, 由切锥的定义 (4.3) 可直接得到结论.

下面假设 $(X;t)\in\operatorname{bd}\mathcal{K}_*\setminus\{(0;0)\}$, 即 $\|X\|_*=t\neq 0$, 因此 $\|X\|_*>0$. 由定理 3.1 知, 对任意 $H\in\Re^{m\times n}$ 有

$$
\begin{aligned}
f'(X;H)&=\sum_{i=1}^{m}\sigma_i'(X;H)\\
&=\sum_{k=1}^{r}\sum_{i=1}^{|\alpha_k|}\lambda_i(\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}))+\sum_{i=1}^{|\alpha_{r+1}|}\sigma_{l(i)}'(X;H)\\
&=\sum_{k=1}^{r}\operatorname{Tr}(\mathscr{S}(\overline{H}_{\alpha_k\alpha_k}))+\sum_{i=1}^{|\beta|}\sigma_i(\overline{H}_{\beta\widehat{\beta}})\\
&=\operatorname{Tr}(\overline{H}_{\alpha\alpha})+\|\overline{H}_{\beta\widehat{\beta}}\|_*,
\end{aligned}
\tag{4.5}
$$

由于 f 是 Lipschitz 连续的正常的凸函数, $f_-^{\downarrow}(X;\cdot)=f'(X;\cdot)$, 又因为 $\mathcal{K}_*=\operatorname{epi}f$, 根据 [12, 定理 8.2] 或者 [8, 命题 2.58] 可得到

$$\mathcal{T}_{\mathcal{K}_*}(X,t)=\operatorname{epi}f'(X;\cdot).$$

因此, 由 (4.5) 可知,

$$
\begin{aligned}
\mathcal{T}_{\mathcal{K}_*}(X,t)&=\{(H;\eta)\in\Re^{m\times n}\times\Re: f'(X;H)\leqslant\eta\}\\
&=\{(H;\eta)\in\Re^{m\times n}\times\Re:\operatorname{Tr}(\overline{H}_{\alpha\alpha})+\|\overline{H}_{\beta\widehat{\beta}}\|_*\leqslant\eta\}.
\end{aligned}
$$

命题得证. □

命题 4.2 令 $(X;t)\in\mathcal{K}_*$ 且 X 满足奇异值分解 (2.8), 假设 $(H;\eta)\in\mathcal{T}_{\mathcal{K}_*}(X,t)$. 则有

(1) 若 $(H;\eta)\in\operatorname{int}\mathcal{T}_{\mathcal{K}_*}(X,t)$, 有 $\mathcal{T}^2_{\mathcal{K}_*}((X;t),(H;\eta))=\Re^{m\times n}\times\Re$;

(2) 若 $(X;t)=(0;0)$, 有 $\mathcal{T}^2_{\mathcal{K}_*}((X;t),(H;\eta))=\mathcal{T}_{\mathcal{K}_*}(H,\eta)$;

(3) 若 $(X;t)\in\operatorname{bd}\mathcal{K}_*\setminus\{(0;0)\}$ 且 $(H;\eta)\in\operatorname{bd}T_{\mathcal{K}_*}(X,t)$, 有

$$\mathcal{T}^2_{\mathcal{K}_*}((X;t),(H;\eta))=\{(W;\gamma)\in\Re^{m\times n}\times\Re: v(H,W)\preceq\gamma\},\tag{4.6}$$

其中

$$v(H,W)=\sum_{k=1}^{r}\operatorname{Tr}(V_k(H,W))+\operatorname{Tr}(Q_{\beta\alpha'}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\alpha'})+\|Q_{\beta\beta'}^{\mathrm{T}}V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}'}\|_*,$$

且

$$\alpha'=\bigcup_{j=1}^{N_{r+1}}\beta_j^{r+1},\quad \beta'=\beta^{r+1}_{N_{r+1}+1},\quad \widehat{\beta}'=\beta^{r+1}_{N_{r+1}+1}\cup\beta_0.$$

证明 当 $(H;\eta)\in\operatorname{int}\mathcal{T}_{\mathcal{K}_*}(X,t)$ 及 $(X;t)=(0;0)$ 时, 由二阶切集的定义可直接得到结论.

下面假设 $(X;t)\in\operatorname{bd}\mathcal{K}_*\setminus\{(0;0)\}$ 且 $(H;\eta)\in\operatorname{bd}\mathcal{T}_{\mathcal{K}_*}(X,t)$, 则有 $\|X\|_*=t\neq 0$ 和 $f'(X;H)=\eta\neq 0$ 成立. 利用 [8, 命题 3.41], 由 $\mathcal{K}_*=\operatorname{epi}f$, f 的 Lipschitz 连续性质和二阶连续可微性质,

$$f_-^{\downarrow\downarrow}(X;H,\cdot)=f''(X;H,\cdot)$$

可知,

$$\mathcal{T}^2_{\mathcal{K}_*}((X;t),(H;\eta)) = \operatorname{epi} f''(X;H,\cdot). \tag{4.7}$$

此时, 由定理 3.1 和 3.2 可知, 对任意 $W \in \Re^{m\times n}$ 有

$$\begin{aligned} f''(X;H,W) &= \sum_{i=1}^{m} \sigma_i''(X;H,W) \\ &= \sum_{k=1}^{r}\sum_{i=1}^{|\alpha_k|} \lambda_i(V_k(H,W)) + \sum_{i}^{|\alpha'|} \lambda_i(\mathscr{S}(Q^{\mathrm{T}}_{\beta\beta^{r+1}_{q_b(i)}} V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\beta^{r+1}_{q_b(i)}})) \\ &\quad + \sum_{i=1}^{|\beta'|} \sigma_i(Q^{\mathrm{T}}_{\beta\beta'} V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}'}) \\ &= \sum_{k=1}^{r} \operatorname{Tr}(V_k(H,W)) + \operatorname{Tr}(Q^{\mathrm{T}}_{\beta\alpha'} V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\alpha'}) + \|Q^{\mathrm{T}}_{\beta\beta'} V_{\beta\widehat{\beta}}(H,W)\widehat{Q}_{\widehat{\beta}\widehat{\beta}'}\|_*. \end{aligned}$$

因此, 由 (4.7) 可得 (4.6). 命题得证. □

类似于命题 4.1 和 4.2 的讨论, 我们给出 $\mathcal{T}_{\mathcal{K}_2}$ 和 $\mathcal{T}^2_{\mathcal{K}_2}$ 的表达式.

命题 4.3 令 $(X;t) \in \mathcal{K}_2$ 且 X 满足奇异值分解 (2.8), 则

$$\mathcal{T}_{\mathcal{K}_2}(X,t) = \begin{cases} \Re^{m\times n} \times \Re, & (X;t) \in \operatorname{int} \mathcal{K}_2, \\ \mathcal{K}_2, & (X;t) = (0;0), \\ \{(H;\eta) \in \Re^{m\times n} \times \Re : \mathscr{S}(\overline{U}^{\mathrm{T}}_{\alpha_1} H \overline{V}_{\alpha_1}) \preceq \eta I_{|\alpha_1|}\}, & \text{其他}, \end{cases} \tag{4.8}$$

其中对于任意的实对称矩阵 X 和 Y, $X \preceq Y$ 意味着 $Y - X$ 是半正定的.

命题 4.4 令 $(X;t) \in \mathcal{K}_2$ 且 X 满足奇异值分解 (2.8), 设 $(H;\eta) \in \mathcal{T}_{\mathcal{K}_2}(X;t)$, 则有

(1) 若 $(H;\eta) \in \operatorname{int} \mathcal{T}_{\mathcal{K}_2}(A,s)$, 有 $\mathcal{T}^2_{\mathcal{K}_2}((X;t),(H;\eta)) = \Re^{m\times n} \times \Re$;

(2) 若 $(X;t) = (0;0)$, 有 $\mathcal{T}^2_{\mathcal{K}_2}((X;t),(H;\eta)) = \mathcal{T}_{\mathcal{K}_2}(H,\eta)$;

(3) 若 $(X;t) \in \operatorname{bd} \mathcal{K}_2 \setminus \{(0;0)\}$ 且 $(H;\eta) \in \operatorname{bd} T_{\mathcal{K}_2}(X,t)$, 有

$$\mathcal{T}^2_{\mathcal{K}_2}((X;t),(H;\eta)) = \{(W;\gamma) \in \Re^{m\times n} \times \Re : Q^{\mathrm{T}} V_1(H,W) Q \preceq \gamma I_{|\alpha_1|}\}, \tag{4.9}$$

其中矩阵 Q 的列构成 $V^{\mathrm{T}}_{\alpha_1} H^{\mathrm{T}} U_{\alpha_1} + U^{\mathrm{T}}_{\alpha_1} H V_{\alpha_1}$ 的最大特征值对应的特征向量空间的一组标准直交基.

致谢 衷心感谢审稿人给予的评价和指导意见.

参考文献

1 Chu M T, Funderlic R E, Plemmons R J. Structured low rank approximation. Linear Algebra Appl, 2003, 366: 157–172

2 Doyle J C. Robustness of multiloop linear feedback systems. In: Proceedings of the 17th IEEE Conference on Decision and Control. Washington: IEEE, 1979, 12–18

3 Doyle J C, Wall J E, Stein G. Performance and robustness analysis for structured uncertainty. In: Proceedings 21st IEEE Conference on Decision and Control Orlando. Washington: IEEE, 1982, 629–636

4 Polak E, Wardi Y. Nondiferentiable optimization algorithm for designing control systems having singular value inequalities. Automatica, 1982, 18: 267–283

5 Sandel N R. Robust stability of systems with applications to singular perturbations. Automatica, 1979, 15: 467–470

6 Ding C, Sun D F, Toh K C. An introduction to a class of matrix cone programming. Preprint

7 Shapiro A. First and second order analysis of nonlinear semidefinite programs. Math Program Ser B, 1997, 77: 301–320
8 Bonnans J F, Shapiro A. Perturbation Analysis of Optimization Problems. New York: Springer, 2000
9 Lancaster P. On eigenvalues of matrices dependent on a parameter. Numer Math, 1964, 6: 377–387
10 Torki M. Second-order directional derivatives of all eigenvalues of a symmetric matrix. Nonlinear Anal Ser A Theory Method, 2001, 46: 1133–1150
11 Horn R A, Johnson C R. Matrix Analysis. Cambridge: Cambridge University Press, 1985
12 Rockafellar R T, Wets J B. Variational Analysis. New York: Springer-Verlag, 1998

The second-order directional derivatives of singular values

ZHANG Ning, ZHANG LiWei & XIAO XianTao

Abstract This paper derives the formula for the second-order derivative of any singular value function in a direct way. As applications, the specific expressions of the tangent cones and the second order tangent sets for the epigraphs of spectral norm and nuclear norm are established, which play important roles in the study of the first and the second-order optimality conditions for matrix optimization.

Keywords singular value, second-order directional derivative, tangent cone, second-order tangent set, spectral norm, nuclear norm

MSC(2010) 90C30

doi: 10.1360/012011-668