An Introduction to a Class of Matrix Cone Programming

Defeng Sun

Department of Mathematics and Risk Management Institute
National University of Singapore

This is a joint work with Chao Ding and Kim Chuan Toh at NUS

July 25, 2010

Let $X \in \Re^{m\times n}$ admit the following singular value decomposition:

$$X = \overline{U}\,[\Sigma(X) \;\; 0]\,\overline{V}^T = \overline{U}\,[\Sigma(X) \;\; 0]\,\begin{bmatrix}\overline{V}_1 & \overline{V}_2\end{bmatrix}^T = \overline{U}\Sigma(X)\overline{V}_1^T\,, \qquad (1)$$

where $\overline{U} \in \mathcal{O}^m$, $\overline{V} \in \mathcal{O}^n$ and $\overline{V}_1 \in \Re^{n\times m}$, $\overline{V}_2 \in \Re^{n\times(n-m)}$ and $\overline{V} = \begin{bmatrix}\overline{V}_1 & \overline{V}_2\end{bmatrix}$. The set of such matrices (U, V) in the singular value decomposition (1) is denoted by $\mathcal{O}^{m,n}(X)$, i.e.,

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

Define the index sets a, b and c by

$$a := \{i \,|\, \sigma_i(X) > 0\}, \quad b := \{i \,|\, \sigma_i(X) = 0\} \quad \text{and} \quad c := \{m+1, \ldots, n\}\,. \tag{2}$$

Let $\overline{\mu}_1 > \overline{\mu}_2 > \ldots > \overline{\mu}_r > 0$ be the nonzero distinct singular values of X. Then, let

$$a_k := \{i \,|\, \sigma_i(X) = \overline{\mu}_k\}, \quad k = 1, \ldots, r\,. \tag{3}$$

For any positive constant $\varepsilon > 0$, denote the closed convex cone $\mathcal{D}_n^\varepsilon$ by

$$\mathcal{D}_n^\varepsilon := \{(t, x) \in \Re \times \Re^n \,|\, \varepsilon^{-1} t \geq \; x_i, \; i = 1, \ldots, n\}\,. \tag{4}$$

Let $\Pi_{\mathcal{D}_n^\varepsilon}(\cdot)$ be the metric projector over $\mathcal{D}_n^\varepsilon$ under the Euclidean inner product in $\Re^n$. That is, for any $(t, x) \in \Re \times \Re^n$, $\Pi_{\mathcal{D}_n^\varepsilon}(t, x)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\tau - t)^2 + \|y - x\|^2\big) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq y_i, \; i = 1, \ldots, n\,. \end{array} \tag{5}$$

The we have the following useful result about $\Pi_{\mathcal{D}_n^\varepsilon}(\cdot, \cdot)$.

For any $x \in \Re^n$, let $x^{\downarrow}$ be the vector of components of x being arranged in the non-increasing order $x_1^{\downarrow} \geq \ldots \geq x_n^{\downarrow}$. Let $\mathrm{sgn}(x)$ be the sign vector of x, i.e., $(\mathrm{sgn})_i(x) = 1$ if $x_i \geq 0$ and -1 otherwise. We use "$\circ$" to denote the Hadamard product operation either for two vectors or two matrices of the same dimensions.

Proposition 1. *Assume that $\varepsilon > 0$ and $(t, x) \in \Re \times \Re^n$ are given. Let π be a permutation of $\{1, \ldots, n\}$ such that $x^{\downarrow} = x_{\pi}$, i.e., $x_i^{\downarrow} = x_{\pi(i)}$, $i = 1, \ldots, n$ and π^{-1} the inverse of π. For convenience, write $x_0^{\downarrow} = +\infty$ and $x_{n+1}^{\downarrow} = -\infty$. Let $\bar{\kappa}$ be the smallest integer $k \in \{0, 1, \ldots, n\}$ such that*

$$x_{k+1}^{\downarrow} \leq \Big(\sum_{j=1}^{k} x_j^{\downarrow} + \varepsilon t \Big) / (k + \varepsilon^2) < x_k^{\downarrow} \, . \tag{6}$$

Define $\bar{y} \in \Re^n$ and $\bar{\tau} \in \Re_+$, respectively, by

$$\bar{y}_i := \begin{cases} \Big(\sum_{j=1}^{\bar{\kappa}} x_j^{\downarrow} + \varepsilon t\Big)/(\bar{\kappa} + \varepsilon^2) & \text{if } 1 \le i \le \bar{\kappa}\,, \\ x_i^{\downarrow} & \text{otherwise} \end{cases}$$

and

$$\bar{\tau} := \varepsilon \bar{y}_1 = \varepsilon \Big(\sum_{j=1}^{\bar{k}} x_j^{\downarrow} + \varepsilon t\Big)/(\bar{k} + \varepsilon^2)\,.$$

The metric projection $\Pi_{\mathcal{D}_n^\varepsilon}(t,x)$ is computed by $\Pi_{\mathcal{D}_n^\varepsilon}(t,x) = (\bar{\tau}, \bar{y}_{\pi^{-1}})$.

For any positive constant $\varepsilon > 0$, denote the closed convex cone $\mathcal{C}_n^\varepsilon$ by

$$\mathcal{C}_n^\varepsilon := \{(t, x) \in \Re \times \Re^n \,|\, \varepsilon^{-1} t \geq \|x\|_\infty\} . \tag{7}$$

Let $\Pi_{\mathcal{C}_n^\varepsilon}(\cdot, \cdot)$ be the metric projector over $\mathcal{C}_n^\varepsilon$ under the Euclidean inner product in $\Re^n$. That is, for any $(t, x) \in \Re \times \Re^n$, $\Pi_{\mathcal{C}_n^\varepsilon}(t, x)$ is the unique optimal solution to the following convex optimization problem

$$\begin{array}{ll} \min & \frac{1}{2}\big((\tau - t)^2 + \|y - x\|^2\big) \\ \text{s.t.} & \varepsilon^{-1}\tau \geq \|y\|_\infty . \end{array} \tag{8}$$

In the following discussions, we frequently drop n from $\mathcal{C}_n^\varepsilon$ when its size can be found from the context.

Assume that $\varepsilon > 0$ and $(t,x) \in \Re \times \Re^n$ are given. Let π be a permutation of $\{1,\ldots,n\}$ such that $|x|^{\downarrow} = |x|_{\pi}$, i.e., $|x|^{\downarrow}_i = |x|_{\pi(i)}$, $i = 1,\ldots,n$ and π^{-1} be the inverse of π. Let $|x|^{\downarrow}_0 = +\infty$ and $|x|^{\downarrow}_{n+1} = 0$. Let $s_0 = 0$ and $s_k = \sum_{i=1}^{k} |x|^{\downarrow}_i$, $k = 1,\ldots,n+1$. Let $\overline{k}$ be the smallest integer $k \in \{0,1,\ldots,n\}$ such that

$$|x|^{\downarrow}_{k+1} \le (s_k + \varepsilon t)/(k+\varepsilon^2) < |x|^{\downarrow}_k \tag{9}$$

or $\overline{k} = n+1$ if such an integer does not exist. Denote

$$\theta^{\varepsilon}(t,x) := (s_{\overline{k}} + \varepsilon t)/(\overline{k} + \varepsilon^2)\,. \tag{10}$$

Let α, β and γ be the index sets of $|x|^{\downarrow}$ as

$$\alpha := \{i \,|\, |x|^{\downarrow}_i > \theta^{\varepsilon}(t,x)\}, \quad \beta := \{i \,|\, |x|^{\downarrow}_i = \theta^{\varepsilon}(t,x)\} \tag{11}$$

and

$$\gamma := \{i \,|\, |x|^{\downarrow}_i < \theta^{\varepsilon}(t,x)\}\,. \tag{12}$$

Define $\bar{y} \in \Re^n$ and $\bar{\tau} \in \Re_+$, respectively, by

$$\bar{y}_i := \begin{cases} \max\{\theta^{\varepsilon}(t,x), 0\} & \text{if } i \in \alpha\,, \\ |x|^{\downarrow}_i & \text{otherwise} \end{cases}$$

and

$$\bar{\tau} := \varepsilon \max\{\theta^{\varepsilon}(t,x), 0\}\,.$$

Proposition 2. *Assume that* $\varepsilon > 0$ *and* $(t, x) \in \Re \times \Re^n$ *are given.*

(i) The metric projection $\Pi_{\mathcal{C}^\varepsilon}(t, x)$ *of* (t, x) *onto* $\mathcal{C}^\varepsilon$ *can be computed as follows*

$$\Pi_{\mathcal{C}^\varepsilon}(t, x) = \left(\bar{\tau}, \operatorname{sgn}(x) \circ \bar{y}_{\pi^{-1}}\right). \tag{13}$$

(ii) The mapping $\Pi_{\mathcal{C}^\varepsilon}(\cdot,\cdot)$ is piecewise linear. Denote $\delta := \sqrt{\varepsilon^2 + \overline{k}}$. For any $(\eta, h) \in \Re \times \Re^n$, let

$$\eta' := \begin{cases} \delta^{-1}(\varepsilon\eta + \sum_{i \in \pi^{-1}(\alpha)} \mathrm{sgn}(x_i) h_i) & \text{if } t \geq -\varepsilon^{-1}\|x\|_1 \,, \\ 0 & \text{otherwise} \,. \end{cases}$$

Let $h' = \mathrm{sgn}(x) \circ h$. Then, the directional derivative of $\Pi_{\mathcal{C}^\varepsilon}(\cdot,\cdot)$ at (t, x) along the direction (η, h) is given by

$$(\bar{\eta}, \bar{h}) := \Pi'_{\mathcal{C}^\varepsilon}((t, x); (\eta, h)) \,, \tag{14}$$

with

$$\left(\delta\varepsilon^{-1}\bar{\eta}, (\operatorname{sgn}(x)\circ\bar{h})_{\pi^{-1}(\beta)}\right) = \begin{cases} \Pi_{\mathcal{D}^{\delta}_{|\beta|}}(\eta', h'_{\pi^{-1}(\beta)}) & \text{if } t > -\varepsilon^{-1}\|x\|_1\,, \\ \Pi_{\mathcal{C}^{\delta}_{|\beta|}}(\eta', h'_{\pi^{-1}(\beta)}) & \text{otherwise}\,, \end{cases}$$

$$\bar{h}_i = \bar{\eta}, \quad i \in \pi^{-1}(\alpha) \quad \text{and} \quad \bar{h}_i = h_i, \quad i \in \pi^{-1}(\gamma)\,.$$

Note that if $\beta = \emptyset$, let $\mathcal{D}^{\delta}_{|\beta|} := \Re$ and $\mathcal{C}^{\delta}_{|\beta|} := \Re_+$.

(iii) The mapping $\Pi_{\mathcal{C}^{\varepsilon}}(\cdot,\cdot)$ is differentiable at (t, x) if and only if $t > \varepsilon||x||_\infty$, or $\varepsilon\|x\|_\infty > t > -\varepsilon^{-1}\|x\|_1$ but $|x|^{\downarrow}_{\bar{k}+1} < (s_k + \varepsilon t)/(\bar{k} + \varepsilon^2)$ or $t < -\varepsilon^{-1}\|x\|_1$.

For any positive constant $\varepsilon > 0$, define the matrix cone $\mathcal{M}_n^{\varepsilon}$ in $\mathcal{S}^n$ as the epigraph of the convex function $\varepsilon\lambda_{\max}(\cdot)$, i.e.,

$$\mathcal{M}_n^{\varepsilon} := \{(t, X) \in \Re \times \mathcal{S}^n \,|\, \varepsilon^{-1} t \geq \lambda_{\max}(X)\} . \tag{15}$$

Proposition 3. *Assume that $(t, X) \in \Re \times \mathcal{S}^n$ is given. Let X have the eigenvalue decomposition*

$$X = \overline{P}\mathrm{diag}(\lambda(X))\overline{P}^T , \tag{16}$$

where $\overline{P} \in \mathcal{O}^n$. Let $\Pi_{\mathcal{M}_n^{\varepsilon}}(\cdot,\cdot)$ be the metric projector over $\mathcal{M}_n^{\varepsilon}$ under Frobenius norm in $\mathcal{S}^n$. Then,

$$\Pi_{\mathcal{M}_n^{\varepsilon}}(t, X) = (\bar{t}, \overline{P}\mathrm{diag}(\bar{y})\overline{P}^T) \quad \forall\, (t, X) \in \Re \times \mathcal{S}^n , \tag{17}$$

where $(\bar{t}, \bar{y}) = \Pi_{\mathcal{D}_n^{\varepsilon}}(t, \lambda(X)) \in \Re \times \Re^n$.

Theorem 1. *Assume that* $(t, X) \in \Re \times \Re^{m\times n}$ *is given. Let* X *have the singular value decomposition (1). Let* $\Pi_{\mathcal{K}^\varepsilon}(\cdot,\cdot)$ *be the metric projector over* $\mathcal{K}^\varepsilon$ *under Frobenius norm in* $\Re^{m\times n}$*. For any* $(t, X) \in \Re \times \Re^{m\times n}$*, we have*

$$\Pi_{\mathcal{K}^\varepsilon}(t, X) = \left(\bar{t}, \overline{U}\,[\mathrm{diag}(\bar{y}) \quad 0]\,\overline{V}^T\right), \tag{18}$$

where

$$(\bar{t}, \bar{y}) = \Pi_{\mathcal{C}^\varepsilon}(t, \sigma(X)) \in \Re \times \Re^m .$$

Define $S : \Re^{m\times m} \rightarrow \mathcal{S}^m$ and $T : \Re^{m\times m} \rightarrow \Re^{m\times m}$ as follows

$$S(Z) := \frac{1}{2}(Z + Z^T) \quad \text{and} \quad T(Z) := \frac{1}{2}(Z - Z^T)\,.$$

For any given $(t, X) \in \Re \times \Re^{m\times n}$, Let X have the singular value decomposition (1). For convenience, write $\sigma_0(X) = +\infty$ and $\sigma_{n+1}(X) = 0$. Let $s_0 = 0$ and $s_k = \sum_{i=1}^{k} \sigma_i(X)$, $k = 1, \ldots, n+1$. Let $\overline{k}$ be the smallest integer $k \in \{0, 1, \ldots, n\}$ such that

$$\sigma_{k+1}(X) \leq (s_k + \varepsilon t)/(k + \varepsilon^2) < \sigma_k(X) \tag{19}$$

or $\overline{k} = n + 1$ if such an integer does not exist.

Denote by

$$\theta(t,\sigma(X)) := (s_{\overline{k}} + \varepsilon t)/(\overline{k} + \varepsilon^2)\,. \tag{20}$$

Let α, β and γ be the index sets of $\sigma(X)$, which are defined by (11) and (12). Let $\delta := \sqrt{1+\overline{k}}$. Define a linear operator $\rho : \Re \times \Re^{m\times n} \to \Re$ as follows

$$\rho(\eta, H) := \begin{cases} \delta^{-1}(\mathrm{Tr}(S(\overline{U}_\alpha^T H(\overline{V}_1)_\alpha)) + \eta) & \text{if } t \geq -\|X\|_*\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{21}$$

Denote by

$$\Big(g_0(t,\sigma(X)), g(t,\sigma(X))\Big) := \Pi_{\mathcal{C}}(t,\sigma(X))\,.$$

Define $\Omega_1 \in \Re^{m\times m}$, $\Omega_2 \in \Re^{m\times m}$ and $\Omega_3 \in \Re^{m\times(n-m)}$ (depend on X) as follows, for any $i, j \in \{1, \ldots, m\}$,

$$(\Omega_1)_{ij} := \begin{cases} \dfrac{g_i(t,\sigma(X)) - g_j(t,\sigma(X))}{\sigma_i(X) - \sigma_j(X)} & \text{if } \sigma_i(X) \neq \sigma_j(X)\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{22}$$

$$(\Omega_2)_{ij} := \begin{cases} \dfrac{g_i(t,\sigma(X)) + g_j(t,\sigma(X))}{\sigma_i(X) + \sigma_j(X)} & \text{if } \sigma_i(X) + \sigma_j(X) \neq 0\,, \\ 0 & \text{otherwise} \end{cases} \tag{23}$$

and for any $i \in \{1, \ldots, m\}$ and $j \in \{1, \ldots, n-m\}$

$$(\Omega_3)_{ij} := \begin{cases} \dfrac{g_i(t,\sigma(X))}{\sigma_i(X)} & \text{if } \sigma_i(X) \neq 0\,, \\ 0 & \text{if } \sigma_i(X) = 0\,, \end{cases} \tag{24}$$

Theorem 2. *Assume that* $(t, X) \in \Re \times \Re^{m\times n}$ *is given. Let* X *have the singular value decomposition (1). Then, the metric projector over the matrix cone* $\mathcal{K}$, $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is directionally differentiable at* (t, X). *For any given direction* $(\eta, H) \in \Re \times \Re^{m\times n}$, *the directional derivative* $\Pi'_{\mathcal{K}}((t, X); (\eta, H))$ *is given by*

(i) if $t > \|X\|_2$, *then* $\Pi'_{\mathcal{K}}((t, X); (\eta, H)) = (\eta, H)$.

(ii) if $\|X\|_2 \geq t > -\|X\|_*$, *then* $\Pi'_{\mathcal{K}}((t, X); (\eta, H)) = (\overline{\eta}, \overline{H})$ *with*

$$
\begin{aligned}
\overline{\eta} &= \delta^{-1}\psi_0^\delta(\eta, H)\,, \\
\overline{H} &= \overline{U}\begin{bmatrix} \delta^{-1}\psi_0^\delta(\eta,H)I_{\alpha\alpha} & 0 & (\Omega_1)_{\alpha\gamma}\circ S(A_{\alpha\gamma}) \\ 0 & \Psi^\delta(\eta,H) & S(A_{\beta\gamma}) \\ (\Omega_1)_{\gamma\alpha}\circ S(A_{\gamma\alpha}) & S(A_{\gamma\beta}) & S(A_{\gamma\gamma}) \end{bmatrix}\overline{V}_1^T \\
&\quad +\overline{U}\begin{bmatrix} (\Omega_2)_{aa}\circ T(A_{aa}) & (\Omega_2)_{ab}\circ T(A_{ab}) \\ (\Omega_2)_{ba}\circ T(A_{ba}) & T(A_{bb}) \end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} (\Omega_3)_{ac}\circ B_{ac} \\ B_{bc} \end{bmatrix}\overline{V}_2^T
\end{aligned}
$$

where $A := \overline{U}^T H \overline{V}_1$, $B := \overline{U}^T H \overline{V}_2$ and $\Big(\psi_0^\delta(\eta,H), \Psi^\delta(\eta,H)\Big) \in \Re \times \Re^{|\beta|\times|\beta|}$ is given by

$$
\Big(\psi_0^\delta(\eta,H), \Psi^\delta(\eta,H)\Big) := \Pi_{\mathcal{M}^\delta_{|\beta|}}(\rho(\eta,H), S(\overline{U}_\beta^T H(\overline{V}_1)_\beta))\,. \tag{25}
$$

(iii) if $t = -\|X\|_*$, then $\Pi'_{\mathcal{K}}((t,X);(\eta,H)) = (\overline{\eta},\overline{H})$ with

$$\begin{aligned}\overline{\eta} &= \delta^{-1}\psi_0^\delta(\eta,H)\,,\\ \overline{H} &= \overline{U}\begin{bmatrix}\overline{\eta}I_{\alpha\alpha} & 0\\ 0 & \Psi_1^\delta(\eta,H)\end{bmatrix}\overline{V}_1^T\\ &\quad +\overline{U}[\Omega_2\circ T(A)]\overline{V}_1^T + \overline{U}\begin{bmatrix}(\Omega_3)_{ac}\circ B_{ac}\\ \Psi_2^\delta(\eta,H)\end{bmatrix}\overline{V}_2^T\end{aligned}$$

where $\psi_0^\delta(\eta,H)\in\Re$, $\Psi_1^\delta(\eta,H)\in\Re^{|\beta|\times|\beta|}$ and $\Psi_2^\delta(\eta,H)\in\Re^{|\beta|\times(n-m)}$ are given by

$$\left(\psi_0^\delta(\eta, H), \left[\Psi_1^\delta(\eta, H) \;\; \Psi_2^\delta(\eta, H)\right]\right)$$
$$:= \;\; \Pi_{\mathcal{K}^\delta_{|\beta|,(n-|a|)}}\left(\rho(\eta, H), \left[\overline{U}_\beta^T H V_\beta \;\; \overline{U}_\beta^T H \overline{V}_2\right]\right).$$

(iv) if $t < -\|X\|_*$, then

$$\Pi'_{\mathcal{K}}((t, X); (\eta, H)) = (0, 0)\,.$$

Moreover, $\Pi_{\mathcal{K}}(\cdot, \cdot)$ is strongly B-differentiable at (t, X), i.e., for any $(\eta, H) \in \Re \times \Re^{m\times n}$ and $(\eta, H) \to (0, 0)$, we have

$$\Pi_{\mathcal{K}}(t + \eta, X + H) - \Pi_{\mathcal{K}}(t, X) - \Pi'_{\mathcal{K}}((t, X); (\eta, H)) = O(\|(\eta, H)\|^2)\,. \quad (26)$$

Theorem 3. *The mapping* $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is F-differentiable at* $(t, X) \in \Re \times \Re^{m\times n}$ *if and only if* (t, X) *satisfies one of the following conditions:*

(i) $t > \|X\|_2$*;*

(ii) $\|X\|_2 > t > -\|X\|_*$ *but* $\sigma_{\bar{k}}(X) < \theta(t, \sigma(X))$*, where* $\bar{k}$ *and* $\theta(t, \sigma(X))$ *are given by (9) and (10), respectively;*

(iii) $t < -\|X\|_*$*.*

Denote $\delta := \sqrt{1+\bar{k}}$*. Let* $\rho : \Re \times \Re^{m\times n} \to \Re$ *be the linear operator defined by (21). Then for any* $(\eta, H) \in \Re \times \Re^{m\times n}$*,* $\Pi'_{\mathcal{K}}(t, X)(\eta, H) = (\bar{\eta}, \overline{H})$ *with*

$$\overline{\eta} = \delta^{-1}\rho(\eta, H)$$

and

$$\begin{aligned}\overline{H} &= \overline{U}\begin{bmatrix} \delta^{-1}\rho(\eta, H)I_{\alpha\alpha} & (\Omega_1)_{\alpha\gamma}\circ S(A_{\alpha\gamma}) \\ (\Omega_1)_{\gamma\alpha}\circ S(A_{\gamma\alpha}) & S(A_{\gamma\gamma}) \end{bmatrix}\overline{V}_1^T \\ &+\overline{U}\begin{bmatrix} (\Omega_2)_{aa}\circ T(A_{aa}) & (\Omega_2)_{ab}\circ T(A_{ab}) \\ (\Omega_2)_{ba}\circ T(A_{ba}) & T(A_{bb}) \end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} (\Omega_3)_{ac}\circ B_{ac} \\ B_{bc} \end{bmatrix}\overline{V}_2^T\end{aligned}$$

where $A := \overline{U}^T H \overline{V}_1$ and $B := \overline{U}^T H \overline{V}_2^T$.

Theorem 4. $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is strongly G-semismooth at any* $(t, X) \in \Re \times \Re^{m\times n}$.