From Linear Programming to Matrix Programming: A Paradigm Shift in Optimization

Defeng Sun

Department of Mathematics and Risk Management Institute

National University of Singapore

Partially based on joint works with Chao Ding, Kim Chuan Toh, and Bin Wu at NUS

December 16, 2011

Linear Programming

The primal form of standard Linear Programming (LP):

$$
\text{(P)} \qquad \begin{array}{ll} \min & \langle c, x\rangle \\ \text{s.t.} & \mathcal{A}x = b\,, \\ & x \geq 0\,, \end{array}
$$

where $\mathcal{A} : \Re^n \to \Re^m$ is a linear operator, $c \in \Re^n$, and $b \in \Re^m$.
The dual form takes:

$$
\text{(D)} \qquad \begin{array}{ll} \max & \langle b, y\rangle \\ \text{s.t.} & \mathcal{A}^* y - c = z\,, \\ & z \leq 0\,, \end{array}
$$

where $\mathcal{A}^*$ is the adjoint of the linear operator $\mathcal{A}$.

Brief History of LP

- **Simplex Method**
  - G.B. Dantzig (1947)
  - Very efficient
  - Not polynomial algorithm. Klee and Minty (1972) gave an example
  - Average analysis versus worst-case analysis
- **Khachiyan's Ellipsoid Method (1979)**
  - Polynomial algorithm
  - Less efficient
- **Interior-Point Algorithms**
  - Karmarkar (1984)
  - Polynomial algorithm
  - Efficient for some large-scale sparse LPs
- **Efficient Softwares Available**

Let $\mathcal{X}$ be the Cartesian product of several finite dimensional real matrix spaces, symmetric or non-symmetric, with the inner product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$.

Let $f : \mathcal{X} \to (-\infty, \infty]$ be a closed proper convex function.

The Fenchel conjugate of f is defined by

$$f^*(z) := \sup_{x \in \mathcal{X}} \{ \langle z, x \rangle - f(x) \} .$$

For example, if $f(x) = \delta_{\Re^n_+}(x)$, the indicator function over $\Re^n_+$, then $f^*(x) = \delta_{(-\Re^n_+)}(x)$.

We are ready to define matrix programming ...

Matrix Programming

The "standard" Matrix Programming (MP) and its dual take the forms:

$$(\mathrm{P}) \qquad \begin{array}{ll} \min & \langle c, x\rangle + f(x) \\ \text{s.t.} & \mathcal{A}x = b \end{array}$$

and

$$(\mathrm{D}) \qquad \begin{array}{ll} \max & \langle b, y\rangle - f^*(z) \\ \text{s.t.} & \mathcal{A}^*y - c = z\,, \end{array}$$

where $\mathcal{A}^*$ is the adjoint of the linear operator $\mathcal{A} : \mathcal{X} \to \Re^m$, $c \in \mathcal{X}$, $b \in \Re^m$.

Here, f and f^* should be "simple" and "computable". In standard linear programming, $f(x) = \delta_{\Re^n_+}(x)$, the indicator function over $\Re^n_+$ and $f^*(x) = \delta_{(-\Re^n_+)}(x)$.

Matrix Cone Programming

When $f(x) = \delta_{\mathcal{K}}(x)$, the indicator function of a closed convex cone $\mathcal{K}$, we can write (P) and (D) equivalently as the following Matrix Cone Programming (MCP):

$$
(\mathrm{P}) \quad \begin{array}{ll} \min & \langle c, x \rangle \\ \text{s.t.} & \mathcal{A}x = b \,, \\ & x \in \mathcal{K} \,, \end{array} \qquad (\mathrm{D}) \quad \begin{array}{ll} \max & \langle b, y \rangle \\ \text{s.t.} & \mathcal{A}^* y - c = z \,, \\ & z \in \mathcal{K}^\circ \,, \end{array}
$$

where $\mathcal{K}^\circ$ is the polar of $\mathcal{K}$, i.e,

$$
\mathcal{K}^\circ := \{ z \in \mathcal{X} \,|\, \langle z, x \rangle \leq 0 \quad \forall\, x \in \mathcal{K} \} \,.
$$

The dual cone of $\mathcal{K}$ is denoted by $\mathcal{K}^* := -\mathcal{K}^\circ$.

Semidefinite Programming

Semidefinite programming (SDP): Let $\mathcal{S}^n$ and $\mathcal{S}^n_+$ be, respectively, the space of $n \times n$ symmetric matrices and the cone of real positive semidefinite matrices in $\mathcal{S}^n$. When $f(x) = \delta_{\mathcal{S}^n_+}(x)$, the indictor function over $\mathcal{S}^n_+$, we have the familiar SDP:

$$
\text{(P)} \quad \begin{array}{ll} \min & \langle c, x\rangle \\ \text{s.t.} & \mathcal{A}x = b\,, \\ & x \in \mathcal{S}^n_+\,, \end{array} \qquad \text{(D)} \quad \begin{array}{ll} \max & \langle b, y\rangle \\ \text{s.t.} & \mathcal{A}^* y - c = z\,, \\ & z \in -\mathcal{S}^n_+\,. \end{array}
$$

The SDP cone $\mathcal{S}^n_+$ is self-dual, i.e., $\mathcal{S}^n_+ = (\mathcal{S}^n_+)^*$. There are many non-self-dual matrix cones.

Copositive Cone Programming

Copositive cone programming: $f(x) = \delta_{\mathcal{C}}(x)$, the indictor function over the cone of copositive matrices, i.e.,

$$\mathcal{C} := \{x \in \mathcal{S}^n \,|\, a^T x a \geq 0 \text{ for all } a \in \Re^n_+\}\,.$$

The function $f^*(x) = \delta_{(-\mathcal{C}^*)}(x)$, where $\mathcal{C}^*$, the dual cone of $\mathcal{C}$, is the cone of completely positive matrices, i.e.,

$$\mathcal{C}^* = \text{conv}\{aa^T \,|\, a \in \Re^n_+\}\,.$$

Copositive programming is hard. Murty and Kabadi [1] has shown that checking whether a given matrix $x \in \mathcal{C}$ is a co-NP-complete decision problem.

[1] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming* 39 (1987) 117–129.

Desirable Properties of f

More specifically, we require

- The Moreau-Yosida regularization of f

$$\psi_f(x) := \min_{z \in \mathcal{X}} f(z) + \frac{1}{2}\|z - x\|^2$$

has a closed form solution, denoted by $P_f(x)$ [non-expansive operator], or at least admits an effective algorithm.

- One can easily compute the directional derivative of

$$\nabla \psi_f(x) = x - P_f(x).$$

Thus, to be able to characterize the Bouligand-sub-differential of $\nabla \psi_f$.

- The function $\nabla \psi_f$ is (strongly) semismooth.

Moreau's Decomposition

Note that Moreau's decomposition

$$P_f(x) + P_{f^*}(x) = x, \quad \forall\, x \in \mathcal{X}$$

is very useful in checking these properties for f.

If $f(x) = \delta_K(x)$, the indicator function over a closed convex cone K, then

$$P_f(x) = \Pi_K(x) \quad \text{and} \quad P_{f^*}(x) = \Pi_{K^o}(x)\,,$$

where Π_D is the metric projection over a closed convex set D and K^o is the polar of K.

For example, $K = \mathcal{S}^n_+$, the cone of real positive semi-definite symmetric matrices, $\Pi_{\mathcal{S}^n_+}(x)$ needs one eigenvalue decomposition[divide and conquer–$4n^3$ operations].

More Examples

More generally, if f is positively homogenous, then $f^*(x) = \delta_{\partial f(0)}(x)$ and

$$x - P_f(x) = P_{f^*}(x) = \Pi_{\partial f(0)}(x)\,, \quad \forall\, x \in \mathcal{X}\,.$$

Let us first look at one simple example:

$$\min_{y \in \Re^k} \quad \|A_0 - \textstyle\sum_{i=1}^k y_i A_i\|_2\,,$$

where A_i are m by n matrices, $\|\cdot\|_2$ is the spectral (operator) norm of matrices (the largest singular value).

Use $\|\cdot\|_*$ to denote the nuclear norm (the sum of all singular values) and B_*^1 to denote the unit nuclear norm ball.

We can equivalently write it in the form of (D):

$$\begin{array}{ll} \max & \langle 0, y\rangle - \|Z\|_2 \\ \text{s.t.} & \mathcal{A}y - A_0 = Z \end{array}$$

and the corresponding form of (P):

$$\begin{array}{ll} \min & \langle A_0, X\rangle + \delta_{B_*^1}(X) \\ \text{s.t.} & \mathcal{A}^* X = 0\,. \end{array}$$

Slightly more complicated:

$$\min_{y\in\Re^k} \quad \|A_0 - \textstyle\sum_{i=1}^{k} y_i A_i\|_2 + \lambda\|y\|_1\,,$$

$\lambda > 0$ and $\|\cdot\|_1$ is the l_1-norm, in order to get a sparse y. Use $\mathrm{epi}\, f_1$ to denote the epigraph of the l_1-norm function.

Now we may write it as (D)

$$\begin{array}{ll} \max & \langle(-\lambda, 0), (y_0, y)\rangle - \|Z\|_2 - \delta_{\mathrm{epi}\, f_1}(z_0, z) \\ \text{s.t.} & \mathcal{A}y - A_0 = Z\,, \\ & (y_0, y) = (z_0, z). \end{array}$$

The corresponding (P) form is:

$$\begin{array}{ll} \min & \langle A_0, X\rangle + \langle 0, (x_0, x)\rangle + \delta_{B_*^1}(X) + \delta_{(-\mathrm{epi}\, f_\infty)}(x_0, x) \\ \text{s.t.} & (x_0, \mathcal{A}^* X + x) = (-\lambda, 0)\,, \end{array}$$

where $\mathrm{epi}\, f_\infty$ is the epigraph of the l_∞ norm function.

The (P) form can be simplified as:

$$\begin{array}{ll} \min & \langle A_0, X\rangle + \langle 0, x\rangle + \delta_{B_*^1}(X) + \delta_{B_\infty^\lambda}(x) \\ \text{s.t.} & \mathcal{A}^* X + x = 0\,, \end{array}$$

where B_∞^λ is the l_∞ norm ball centered at the origin with radius λ.

A bit more complicated:

$$\min_{y \in \Re^k} \quad \tfrac{1}{2}\|A_0 - \textstyle\sum_{i=1}^k y_i A_i\|_2^2 + \lambda \|y\|_1 \,.$$

Now we may write it as the (D) form:

$$\begin{array}{ll} \max & \langle 0, y\rangle - \tfrac{1}{2}\|Z\|_2^2 - \lambda\|z\|_1 \\ \text{s.t.} & \mathcal{A}y - A_0 = Z \,, \\ & y = z \,. \end{array}$$

The corresponding (P) form is:

$$\begin{array}{ll} \min & \langle A_0, X\rangle + \langle 0, x\rangle + \frac{1}{2}\|X\|_*^2 + \delta_{B_\infty^\lambda}(x) \\ \text{s.t.} & \mathcal{A}^* X + x = 0\,. \end{array}$$

In low rank matrix optimization problems, we need to solve (P) form for the nonsymmetric problem:

$$\begin{array}{ll} \min & \langle C, X\rangle + \sum_{j=1}^{k} \sigma_j(X) \\ \text{s.t.} & \mathcal{A}(X) = b \end{array}$$

and the (P) form for the symmetric problem:

$$\begin{array}{ll} \min & \langle C, X\rangle + \sum_{j=1}^{k} \lambda_j(X) + \delta_{\mathcal{S}^n_+}(X) \\ \text{s.t.} & \mathcal{A}(X) = b\,. \end{array}$$

One non-symmetric matrix cone

Define

$$\mathcal{K}^{\varepsilon}_{m,n} := \{(t, X) \in \Re \times \Re^{m\times n} \,|\, \varepsilon^{-1} t \geq \|X\|_2\}$$

for $\varepsilon > 0$.

We drop m and n if they are clear from the context and omit ε if it is 1.
So $\mathcal{K}$ is the epigraph of the operator norm $\|\cdot\|_2$.

When $n = 1$, $\mathcal{K}$ is the epigraph of the l_2-norm, which is better known as the second order cone, or Lorentz cone, or ice-cream cone.

Why bother?

Note that we can write $t \geq \|X\|_2$ (here, $X \in \Re^{m \times n}$) equivalently as

$$\mathcal{S}^{m+n} \ni \begin{bmatrix} tI_m & X \\ X^T & tI_n \end{bmatrix} \succeq 0\,.$$

However, this changes the dimension from mn to $\frac{1}{2}(m+n)^2$. No one will do it if $m \ll n$ or $n \ll m$. Think about the second order cone case ($n = 1$).

Most importantly, now it is clear that we cannot expect IPMs to solve large scale SDPs. Such transformations must be avoided. The remedy – go back to the problem where it comes from and resort to nonsmooth analysis – semismooth analysis.

Why nonsmooth?

Recall that

$$\nabla \psi_f(x) = x - P_f(x)$$

is only Lipschitz continuous.

[Non-smoothness allows non-singularity!] while Smoothness $\Longrightarrow$ Singularity.

Non-singularity makes the conjugate gradient methods to solve the corresponding linear systems possible.

Where are we?

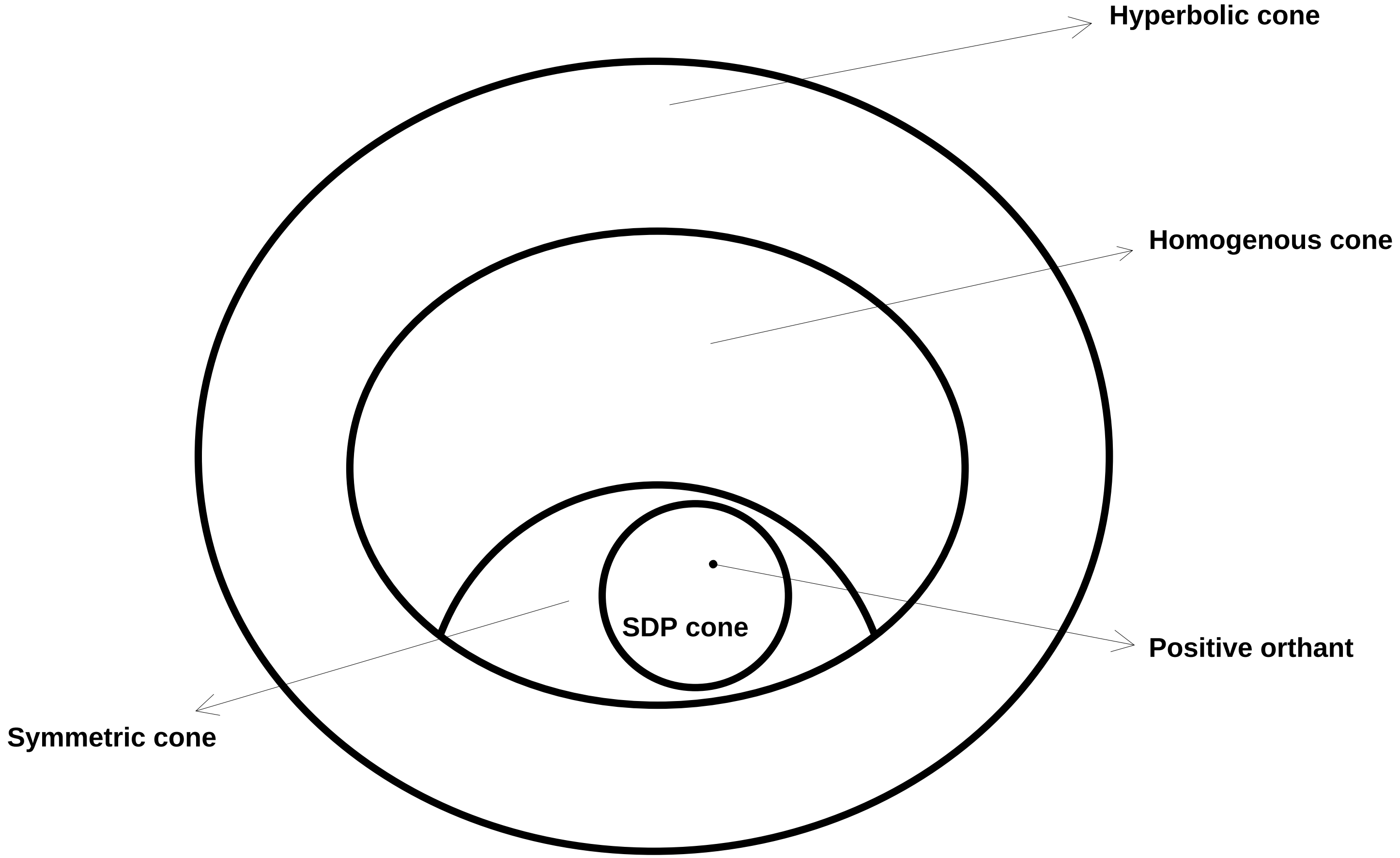

Figure 1: The cones relationship

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

$$X = \overline{U}\,[\Sigma(X)\ \ 0]\,\overline{V}^T = \overline{U}\,[\Sigma(X)\ \ 0]\,\left[\overline{V}_1\ \ \overline{V}_2\right]^T = \overline{U}\Sigma(X)\overline{V}_1^T\,,$$

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 $\overline{V} = \left[\overline{V}_1\ \ \overline{V}_2\right]$.

The set of such matrices (U, V) in the singular value decomposition is denoted by $\mathcal{O}^{m,n}(X)$, i.e.,

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

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

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

For any $(t, x) \in \Re \times \Re^n$, $\Pi_{\mathcal{D}_n^\varepsilon}(t, x)$ is the unique optimal solution to the following simple quadratic convex optimization problem

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

We can solve (1) at a cost of $O(n)$ operations.

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.,

$$\mathcal{M}_n^\varepsilon := \left\{(t, X) \in \Re \times \mathcal{S}^n \,|\, \varepsilon^{-1} t \geq \lambda_{\max}(X)\right\}. \tag{2}$$

Proposition 1. *Let X have the eigenvalue decomposition*

$$X = \overline{P}\mathrm{diag}(\lambda(X))\overline{P}^T, \tag{3}$$

where $\overline{P} \in \mathcal{O}^n$. Then,

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

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

Theorem 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*

$$\Pi_{\mathcal{K}^\varepsilon}(t,X) = \left(\bar{t}, \overline{U}\,[\mathrm{diag}(\bar{y}) \;\; 0]\,\overline{V}^T\right), \tag{4}$$

where

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

and for any positive constant $\varepsilon > 0$, we denote the closed convex cone $\mathcal{C}_m^\varepsilon$ by

$$\mathcal{C}_m^\varepsilon := \left\{(t,x) \in \Re \times \Re^m \,|\, \varepsilon^{-1} t \geq \|x\|_\infty\right\}.$$

Proof. Trivial. Just use von Neumann's inequality
$\|\sigma(X) - \sigma(Y)\| \leq \|X - Y\|$.

Note that for any $(t, x) \in \Re \times \Re^m$, $\Pi_{\mathcal{C}_m^\varepsilon}(t, x)$ is the unique optimal solution to the following convex optimization problem

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

This is simple quadratic programming – it can be solved in $O(m)$ operations. So the cost of this part is negligible.

Define the index sets a, b and c by

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

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

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

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

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

For convenience, write $\sigma_0(X) = +\infty$ and $\sigma_{n+1}(X) = -\infty$. Let $s_0 = 0$ and $s_k = \sum_{i=1}^k \sigma_i(X)$, $k = 1,\ldots,m$. Let $\bar{k}$ be the smallest integer $k \in \{0,1,\ldots,m\}$ such that

$$\sigma_{k+1}(X) \le (s_k + \varepsilon t)/(k + \varepsilon^2) < \sigma_k(X)\,. \tag{5}$$

Denote

$$\theta(t, \sigma(X)) := (s_{\overline{k}} + \varepsilon t)/(\overline{k} + \varepsilon^2)\,. \tag{6}$$

Define three index sets α, β and γ in $\{1, \ldots, n\}$ by

$$\alpha := \{i \,|\, \sigma_i(X) > \theta^\varepsilon(t, \sigma(X))\}, \quad \beta := \{i \,|\, \sigma_i(X) = \theta^\varepsilon(t, \sigma(X))\}$$

and

$$\gamma := \{i \,|\, \sigma_i(X) < \theta^\varepsilon(t, \sigma(X))\}\,.$$

Let $\delta := \sqrt{1 + \bar{k}}$. Define a linear operator $\rho : \Re \times \Re^{m \times n} \to \Re$ as follows

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

Denote

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

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

$$(\Omega_1)_{ij} := \begin{cases} \dfrac{g_i(t,\sigma(X)) - g_j(t,\sigma(X))}{\sigma_i(X) - \sigma_j(X)} & \text{if } \sigma_i(X) \neq \sigma_j(X)\,, \\ 0 & \text{otherwise}\,, \end{cases} \tag{7}$$

$$(\Omega_2)_{ij} := \begin{cases} \dfrac{g_i(t,\sigma(X)) + g_j(t,\sigma(X))}{\sigma_i(X) + \sigma_j(X)} & \text{if } \sigma_i(X) + \sigma_j(X) \neq 0\,, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

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

$$(\Omega_3)_{ij} := \begin{cases} \dfrac{g_i(t,\sigma(X))}{\sigma_i(X)} & \text{if } \sigma_i(X) \neq 0\,, \\ 0 & \text{if } \sigma_i(X) = 0\,, \end{cases} \tag{9}$$

Theorem 2. *The metric projector over the matrix cone* $\mathcal{K}$, $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is directionally differentiable at* (t, X)*. For any given direction* $(\eta, H) \in \Re \times \Re^{m\times n}$*, let* $A := \overline{U}^T H \overline{V}_1$*,* $B := \overline{U}^T H \overline{V}_2$*.*

Then the directional derivative $\Pi'_{\mathcal{K}}((t,X);(\eta,H))$ *can be computed as follows*

(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*

$$\overline{\eta} = \delta^{-1}\psi_0^\delta(\eta, H)\,,$$

$$\overline{H} = \overline{U}\begin{bmatrix} \overline{\eta} I_{|\alpha|} & 0 & (\Omega_1)_{\alpha\gamma}\circ S(A)_{\alpha\gamma} \\ 0 & \Psi^\delta(\eta, H) & S(A)_{\beta\gamma} \\ (\Omega_1)_{\gamma\alpha}\circ S(A)_{\gamma\alpha} & S(A)_{\gamma\beta} & S(A)_{\gamma\gamma} \end{bmatrix}\overline{V}_1^T$$

$$+\overline{U}\begin{bmatrix} (\Omega_2)_{aa}\circ T(A)_{aa} & (\Omega_2)_{ab}\circ T(A)_{ab} \\ (\Omega_2)_{ba}\circ T(A)_{ba} & T(A)_{bb} \end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} (\Omega_3)_{ac'}\circ B_{ac'} \\ B_{bc'} \end{bmatrix}\overline{V}_2^T\,,$$

where $\left(\psi_0^\delta(\eta, H), \Psi^\delta(\eta, H)\right) \in \Re \times \mathcal{S}^{|\beta|}$ is given by

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

In particular, if $t = \|X\|_2 > 0$, we have that $\bar{k} = 0$, $\delta = 1$, $\alpha = \emptyset$, $\rho(\eta, H) = \eta$ and

$$\overline{\eta} = \psi_0^\delta(\eta, H), \quad \overline{H} = \overline{U} \begin{bmatrix} \Psi^\delta(\eta, H) + T(A)_{\beta\beta} & A_{\beta\gamma} \\ A_{\gamma\beta} & A_{\gamma\gamma} \end{bmatrix} \overline{V}_1^T + \overline{U} B \overline{V}_2^T \, ;$$

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

$$\overline{\eta} = \delta^{-1} \psi_0^\delta(\eta, H) \, , \tag{10}$$

$$\overline{H} = \overline{U} \begin{bmatrix} \overline{\eta} I_{|\alpha|} & 0 \\ 0 & \Psi_1^\delta(\eta, H) \end{bmatrix} \overline{V}_1^T + \overline{U} \begin{bmatrix} 0 \\ \Psi_2^\delta(\eta, H) \end{bmatrix} \overline{V}_2^T \, , \tag{11}$$

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

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

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

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

Moreover, $\Pi_{\mathcal{K}}(\cdot,\cdot)$ is 1-order B-differentiable (or B-diff. of degree 2) at (t, X), i.e., for any $(\eta, H) \in \Re \times \Re^{m\times n}$ with $(\eta, H) \to (0,0)$, we have

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

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

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

(ii) $\|X\|_2 > t > -\|X\|_*$ *but* $\sigma_{\bar{k}+1}(X) < \theta(t, \sigma(X))$*;*

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

In this case, for any $(\eta, H) \in \Re \times \Re^{m\times n}$, $\Pi'_{\mathcal{K}}(t, X)(\eta, H) = (\bar{\eta}, \overline{H})$, *where under condition (i),* $(\overline{\eta}, \overline{H}) = (\eta, H)$*; under condition (ii),*

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

and

$$\begin{aligned}
\overline{H} &= \overline{U}\begin{bmatrix} \delta^{-1}\rho(\eta, H)I_{|\alpha|} & (\Omega_1)_{\alpha\gamma} \circ S(A)_{\alpha\gamma} \\ (\Omega_1)_{\gamma\alpha} \circ S(A)_{\gamma\alpha} & S(A)_{\gamma\gamma} \end{bmatrix}\overline{V}_1^T \\
&+ \overline{U}\begin{bmatrix} (\Omega_2)_{aa} \circ T(A)_{aa} & (\Omega_2)_{ab} \circ T(A)_{ab} \\ (\Omega_2)_{ba} \circ T(A)_{ba} & T(A)_{bb} \end{bmatrix}\overline{V}_1^T + \overline{U}\begin{bmatrix} (\Omega_3)_{ac'} \circ B_{ac'} \\ B_{bc'} \end{bmatrix}\overline{V}_2^T
\end{aligned}$$

with $A := \overline{U}^T H \overline{V}_1$, $B := \overline{U}^T H \overline{V}_2^T$; and under condition (iii), $(\overline{\eta}, \overline{H}) = (0, 0)$.

Theorem 4. $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is strongly G-semismooth at any* $(t, X) \in \Re \times \Re^{m\times n}$ *and* $\Pi_{\mathcal{K}}(\cdot,\cdot)$ *is* 1*-order B-diff at any* $(t, X) \in \Re \times \Re^{m\times n}$

Note that a locally Lipschitz function $G : \Re^p \to \Re^q$ is said to be strongly G-semismooth at x if

$$G(x+h) - G(x) - \partial G(x+h)h = O(\|h\|^2)\,.$$

Do we know more now?

Five matrix cones can be included:
1) The epigraph of the Frobenius norm (the second order cone):

$$\{(t, X) \mid t \geq \|X\|_F\}.$$

2) The epigraph of the l_∞ norm

$$\{(t, X) \mid t \geq \|X\|_\infty\}.$$

3) The epigraph of the l_1 norm

$$\{(t, X) \mid t \geq \|X\|_1\}.$$

4) The epigraph of the operator norm

$$\{(t, X) \mid t \geq \|X\|_2\}.$$

5) The epigraph of the nuclear norm

$$\{(t, X) \mid t \geq \|X\|_*\}.$$

More can be done ...

The epigraph of any absolutely symmetric convex functions instead of just the norm functions.

One does not need to consider the epigraphs only, instead one may focus on the Moreau-Yosida regulation of interesting convex functions.

Many more need to be done on second order variational analysis for matrix optimization problems.

The proximal point algorithm (PPA) is currently the targeted numerical approach. Better ideas needed.