Recent Developments in Nonlinear Optimization Theory

Defeng Sun
Department of Mathematics
National University of Singapore
Republic of Singapore

July 9-11, 2006

2 Variational Analysis on Metric Projectors Over Closed Convex Sets

Let Z be a finite-dimensional Hilbert vector space equipped with a scalar product $\langle \cdot, \cdot \rangle$ and its induced norm $\| \cdot \|$ and D be a nonempty closed convex set in Z. For any $z \in Z$, let $\Pi_D(z)$ denote the metric projection of z onto D:

$$\begin{array}{ll} \min & \frac{1}{2}\langle y - z, y - z \rangle \\ \text{s.t.} & y \in D. \end{array} \tag{1}$$

The operator $\Pi_D : Z \to Z$ is called the metric projection operator or metric projector over D.

Proposition 2.1 *Let D be a nonempty closed convex set in Z. Then the point $y \in D$ is an optimal solution to (1) if and only if it satisfies*

$$\langle z - y, d - y \rangle \leq 0 \quad \forall\, d \in D\,. \tag{2}$$

Proof. "$\Rightarrow$" Suppose that $y \in D$ is an optimal solution to (1). Let d be an arbitrary point in D. Then $y_t := (1-t)y + td \in D$ for any $t \in [0,1]$. This, together with the fact that y is an optimal solution, implies that

$$\|z - y_t\|^2 \geq \|z - y\|^2 \quad \forall\, t \in [0,1],$$

which further implies

$$\|(1-t)(z-y) + t(z-d)\|^2 \geq \|z-y\|^2 \;\; \forall\, t \in [0,1].$$

Thus,

$$(t^2 - 2t)\|z - y\|^2 + 2t(1-t)\langle z - y, z - d\rangle + t^2\|z - d\|^2 \geq 0 \;\forall\, t \in [0,1].$$

By taking $t \downarrow 0$ and dividing t on both sides of the above equation, we obtain

$$-2\|z - y\|^2 + 2\langle z - y, z - d\rangle \geq 0\,,$$

which turns into (2).

"$\Leftarrow$" Suppose that $y \in D$ satisfies (2). Assume on the contrary that y does not solve (1). Then we have by the assumption,

$$\langle z - y, \Pi_D(z) - y\rangle \leq 0$$

and by the sufficiency part,

$$\langle z - \Pi_D(z), y - \Pi_D(z)\rangle \leq 0\,.$$

Summing up the above two inequalities leads to

$$\langle \Pi_D(z) - y, \Pi_D(z) - y \rangle \le 0\,.$$

This implies that $y = \Pi_D(z)$. The contradiction shows that y solves (1). □

Note that Proposition 2.1 holds even if Z is infinite-dimensional.

If D is a nonempty closed convex cone, then (2) is equivalent to

$$\langle z - \Pi_D(z), \Pi_D(z) \rangle = 0 \quad \& \quad \langle z - \Pi_D(z), d \rangle \le 0 \quad \forall\, d \in D\,. \quad (3)$$

Proposition 2.2 *Let D be a nonempty closed convex set in Z. Then the metric projector $\Pi_D(\cdot)$ satisfies*

$$\langle y - z, \Pi_D(y) - \Pi_D(z)\rangle \geq \|\Pi_D(y) - \Pi_D(z)\|^2 \quad \forall\, y, z \in Z\,. \quad (4)$$

Note that (4) implies

$$\|\Pi_D(y) - \Pi_D(z)\| \leq \|y - z\| \quad \forall\, y, z \in Z\,.$$

Proof. Let $y, z \in Z$. Then by Proposition 2.1, we have

$$\langle z - \Pi_D(z), \Pi_D(y) - \Pi_D(z)\rangle \leq 0$$

and

$$\langle y - \Pi_D(y), \Pi_D(z) - \Pi_D(y)\rangle \leq 0\,.$$

Summing them up gives the desired inequality (4). □

The metric projector $\Pi_D(\cdot)$ is only globally Lipschitz continuous and is not differentiable everywhere, but we have

Proposition 2.3 *Let D be a nonempty closed convex set in Z. Let*

$$\theta(z) := \frac{1}{2}\|z - \Pi_D(z)\|^2, \quad z \in Z\,.$$

Then θ is continuously differentiable with

$$\nabla\theta(z) = z - \Pi_D(z)\,, \quad z \in Z\,.$$

Proof. For any $z \in Z$, let

$$Q(z) := z - \Pi_D(z).$$

Then we have for $\Delta z \to 0$ that

$$\begin{aligned}
&\theta(z+\Delta z) - \theta(z) \\
&= \frac{1}{2}\langle Q(z+\Delta z) - Q(z), Q(z+\Delta z) + Q(z)\rangle \\
&= \frac{1}{2}\langle \Delta z - [\Pi_D(z+\Delta z) - \Pi_D(z)], Q(z+\Delta z) + Q(z)\rangle \\
&= \langle \Delta z - [\Pi_D(z+\Delta z) - \Pi_D(z)], Q(z)\rangle + O(\|\Delta z\|^2) \\
&= \langle Q(z), \Delta z\rangle - \langle \Pi_D(z+\Delta z) - \Pi_D(z), Q(z)\rangle + O(\|\Delta z\|^2) \\
&= \langle Q(z), \Delta z\rangle - \langle \Pi_D(z+\Delta z) - \Pi_D(z), z - \Pi_D(z)\rangle + O(\|\Delta z\|^2) \\
&\geq \langle Q(z), \Delta z\rangle + O(\|\Delta z\|^2) \quad \text{(by (2))}
\end{aligned}$$

and similarly

$$
\begin{aligned}
&\theta(z+\Delta z)-\theta(z)\\
&=\frac{1}{2}\langle \Delta z-[\Pi_D(z+\Delta z)-\Pi_D(z)], Q(z+\Delta z)+Q(z)\rangle\\
&=\langle \Delta z-[\Pi_D(z+\Delta z)-\Pi_D(z)], Q(z+\Delta z)\rangle+O(\|\Delta z\|^2)\\
&=\langle Q(z+\Delta z),\Delta z\rangle-\langle \Pi_D(z+\Delta z)-\Pi_D(z), Q(z+\Delta z)\rangle+O(\|\Delta z\|^2)\\
&=\langle Q(z),\Delta z\rangle+\langle \Pi_D(z)-\Pi_D(z+\Delta z), Q(z+\Delta z)\rangle+O(\|\Delta z\|^2)\\
&\le \langle Q(z),\Delta z\rangle+O(\|\Delta z\|^2)\quad \text{(by (2))}\,.
\end{aligned}
$$

Thus θ is Fréchet differentiable at z with

$$\nabla\theta(z)=z-\Pi_D(z)\,.$$

The continuity of $\nabla\theta(\cdot)$ follows from the global Lipschitz continuity of $\Pi_D(\cdot)$. □

Recall that the normal cone $\mathcal{N}_D(y)$ at y in the sense of convex analysis is

$$\mathcal{N}_D(y) = \begin{cases} \{d \in Y \,:\, \langle d, z - y \rangle \leq 0 \;\; \forall\, z \in D\} & \text{if } y \in D, \\ \emptyset & \text{if } y \notin D. \end{cases}$$

Proposition 2.4 *Let D be a nonempty closed convex set in Z. Then a point $\mu \in \mathcal{N}_D(y)$ if and only if*

$$y = \Pi_D(y + \mu)\,. \tag{5}$$

Note that $\mu \in \mathcal{N}_D(y)$ already implies that $y \in D$.

Proof. "$\Rightarrow$" Suppose that $\mu \in \mathcal{N}_D(y)$. Then $y \in D$ and

$$\langle \mu, z - y \rangle \leq 0 \quad \forall\, z \in D\,.$$

Thus,

$$\langle (y + \mu) - y, z - y \rangle \leq 0 \quad \forall\, z \in D\,,$$

which, according to Proposition 2.1, implies $y = \Pi_D(y + \mu)$.

"$\Leftarrow$" Suppose that $y = \Pi_D(y + \mu)$. Then $y \in D$. By Proposition 2.1, we have

$$\langle (y + \mu) - y, z - y \rangle \leq 0 \quad \forall\, z \in D\,,$$

i.e.,

$$\langle \mu, z - y \rangle \leq 0 \quad \forall\, z \in D\,.$$

That is, $\mu \in \mathcal{N}_D(y)$. □

Proposition 2.5 *Let D be a nonempty closed convex cone in Z and $D^o := -D^*$ be the polar of D. Then any $z \in Z$ can be uniquely decomposed into*

$$z = \Pi_D(z) + \Pi_{D^o}(z)\,. \tag{6}$$

Proof. Let $u := z - \Pi_D(z)$. By (3), we have

$$\langle u, \Pi_D(z) \rangle = 0 \quad \& \quad \langle u, d \rangle \le 0 \quad \forall d \in D\,.$$

Thus $u \in D^o$, $\langle z - u, u \rangle = 0$, and

$$\langle z - u, w \rangle = \langle z - (z - \Pi_D(z)), w \rangle = \langle \Pi_D(z), w \rangle \le 0 \quad \forall\, w \in D^o.$$

Hence, $u = \Pi_{D^o}(z)$. The uniqueness of the decomposition is obvious. $\square$.

For A and B in $\mathcal{S}^p$, define

$$\langle A, B \rangle := \mathrm{Tr}\left(A^T B\right) = \mathrm{Tr}\left(AB\right) \,,$$

where "Tr" denotes the trace of a square matrix (i.e., the sum of all diagonal elements of the symmetric matrix). Let $A \in \mathcal{S}^p$ have the following spectral decomposition

$$A = P \Lambda P^T,$$

where Λ is the diagonal matrix of eigenvalues of A and P is a corresponding orthogonal matrix of orthonormal eigenvectors.

Let

$$A_+ := P\Lambda_+ P^T.$$

Then, $\langle A - A_+, A_+ \rangle = \langle \Lambda - \Lambda_+, \Lambda_+ \rangle = 0$ and

$$\langle A - A_+, H \rangle = \langle \Lambda - \Lambda_+, P^T H P \rangle \leq 0 \quad \forall\, H \in \mathcal{S}^p_+.$$

Thus, by (3), we obtain that :

$$\Pi_{\mathcal{S}^p_+}(A) = A_+ = P\Lambda_+ P^T\,.$$

Let $\Xi : \mathcal{O} \subseteq Y \to Z$ be a locally Lipschitz continuous function on the open set $\mathcal{O}$, where Y is another finite-dimensional REAL Hilbert space.

Then by the Rademacher theorem, Ξ is almost everywhere (in the Lebesgue sense) Fréchet differential in $\mathcal{O}$. We denote by $\mathcal{O}_\Xi$ the set of points in $\mathcal{O}$ where Ξ is Fréchet differentiable. If $\mathcal{O} \equiv Y$, we use $\mathcal{D}_\Xi$ to represent Y_Ξ. Then Clarke's generalized Jacobian of Ξ at y is:

$$\partial \Xi(y) := \mathrm{conv}\{\partial_B \Xi(y)\},$$

where "conv" denotes the convex hull and

$$\partial_B \Xi(y) := \{V \,:\, V = \lim_{k\to\infty} \mathcal{J}\Xi(y^k)\,,\, y^k \to y\,,\, y^k \in \mathcal{O}_\Xi\}.$$

Proposition 2.6 *Let D be a nonempty closed convex set in Z. For any $y \in Z$ and $V \in \partial \Pi_D(y)$,*

(i) *V is self-adjoint;*

(ii) $\langle d, Vd \rangle \geq 0 \quad \forall\, d \in Z$*; and*

(iii) $\langle Vd, d - Vd \rangle \geq 0 \quad \forall\, d \in Z$.

Proof. (i) Define $\varphi : Z \longrightarrow \Re$ by

$$\varphi(z) := \frac{1}{2}[\langle z, z \rangle - \langle z - \Pi_D(z), z - \Pi_K(z) \rangle], \quad z \in Z\,.$$

Then, by Proposition 2.3, φ is continuously differentiable with

$$\nabla \varphi(z) = z - [z - \Pi_D(z)] = \Pi_D(y)\,, \quad z \in Z\,.$$

It then follows that if $\Pi_D(\cdot)$ is Fréchet differentiable at some z, then $\mathcal{J}\Pi_D(z)$ is self-adjoint. Thus, V, as the limit of $\mathcal{J}\Pi_D(y^k)$ for some $y^k \in \mathcal{D}_{\Pi_D}$ converging to y, is also self-adjoint.

(ii) is a special case of (iii).

(iii) First, we consider $z \in \mathcal{D}_{\Pi_D}$. By Proposition 2.2, for any $d \in Z$ and $t \geq 0$, we have

$$\langle \Pi_D(z+td) - \Pi_D(z), td\rangle \geq ||\Pi_D(z+td) - \Pi_D(z)||^2, \quad \text{for all } t \geq 0.$$

Hence,

$$\langle \mathcal{J}\Pi_D(z)d, d\rangle \geq \langle \mathcal{J}\Pi_D(z)d, \mathcal{J}\Pi_D(z)d\rangle. \tag{7}$$

Next, let $V \in \partial\Pi_D(y)$. Then, by Carathéodory's theorem, there exist a positive integer $\kappa > 0$, $V^i \in \partial_B \Pi_D(y)$, $i = 1, 2, \ldots, \kappa$ such that

$$V = \sum_{i=1}^{\kappa} \lambda_i V^i \,,$$

where $\lambda_i \geq 0$, $i = 1, 2, \ldots, \kappa$, and $\sum_{i=1}^{\kappa} \lambda_i = 1$.

Let $d \in Z$. For each $i = 1, \ldots, \kappa$ and $k = 1, 2, \ldots,$ there exists $y^{i_k} \in \mathcal{D}_{\Pi_D}$ such that

$$||y - y^{i_k}|| \leq 1/k$$

and

$$||\mathcal{J}\Pi_D(y^{i_k}) - V^i|| \leq 1/k$$

By (7), we have

$$\langle \mathcal{J}\Pi_D(y^{i_k})d, d\rangle \geq \langle \mathcal{J}\Pi_D(y^{i_k})d, \mathcal{J}\Pi_D(y^{i_k})d\rangle.$$

Hence,

$$\langle V^i d, d\rangle \geq \langle V^i d, V^i d\rangle,$$

and so,

$$\sum_{i=1}^{\kappa} \lambda_i \langle V^i d, d\rangle \geq \sum_{i=1}^{\kappa} \lambda_i \langle V^i d, V^i d\rangle. \tag{8}$$

Define $\theta(z) := ||z||^2,\ z \in Z$. By the convexity of θ, we have

$$\theta\Big(\sum_{i=1}^{\kappa} \lambda_i V^i d\Big) \leq \sum_{i=1}^{\kappa} \lambda_i \theta(V^i d) = \sum_{i=1}^{\kappa} \lambda_i \langle V^i d, V^i d\rangle = \sum_{i=1}^{\kappa} \lambda_i ||V^i d||^2 .$$

Hence,

$$\sum_{i=1}^{\kappa} \lambda_i ||V^i d||^2 \geq \Big\langle \sum_{i=1}^{\kappa} \lambda_i V^i d, \sum_{i=1}^{\kappa} \lambda_i V^i d \Big\rangle. \tag{9}$$

By using (8) and (9), we obtain for all $d \in Z$ that

$$\langle Vd, d\rangle \geq \langle Vd, Vd\rangle.$$

The proof is completed. □

Recall that if $A \in \mathcal{S}^p$ has the following spectral decomposition

$$A = P \Lambda P^T,$$

where Λ is the diagonal matrix of eigenvalues of A and P is a corresponding orthogonal matrix of orthonormal eigenvectors, then

$$A_+ = \Pi_{\mathcal{S}^p_+}(A) = P \Lambda_+ P^T.$$

Define

$$\alpha := \{i : \lambda_i > 0\}, \ \beta := \{i : \lambda_i = 0\}, \ \gamma := \{i : \lambda_i < 0\}.$$

Write

$$\Lambda = \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} \text{ and } P = [\ P_\alpha \quad P_\beta \quad P_\gamma \ ].$$

Define $U \in \mathcal{S}^p$:

$$U_{ij} := \frac{\max\{\lambda_i, 0\} + \max\{\lambda_j, 0\}}{|\lambda_i| + |\lambda_j|}, \quad i, j = 1, \ldots, p,$$

where $0/0$ is defined to be 1.

The operator $\Pi_{\mathcal{S}^p_+}(\cdot)$ is strongly semismooth at A, i.e., in addition to the directional differentiability of $\Pi_{\mathcal{S}^p_+}(\cdot)$ at A, for any $H \in \mathcal{S}^p$ and $V \in \partial \Pi_{\mathcal{S}^p_+}(A+H)$ we have

$$\Pi_{\mathcal{S}^p_+}(A+H) - \Pi_{\mathcal{S}^p_+}(A) - V(H) = O(\|H\|^2) \,. \qquad (10)$$

The directional derivatve of $\Pi_{\mathcal{S}^p_+}(\cdot)$ at A has a very compact form

$$\Pi'_{\mathcal{S}^p_+}(A;H) = P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma} \circ P_\alpha^T H P_\gamma \\ P_\beta^T H P_\alpha & \Pi_{\mathcal{S}^{|\beta|}_+}(P_\beta^T H P_\beta) & 0 \\ P_\gamma^T H P_\alpha \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T,$$

where $\circ$ denotes the Hadamard product. Note that $\Pi'_{\mathcal{S}^p_+}(A;H)$ does not depend on any particularly chosen P.

The following result needs a long but not very complicated proof.

Proposition 2.7 *Let*

$$\Theta(\cdot) := \Pi'_{\mathcal{S}^p_+}(A; \cdot).$$

It holds that

$$\partial_B \Pi_{\mathcal{S}^p_+}(A) = \partial_B \Theta(0).$$

Proposition 2.8 *Let $\Psi : X \to Y$ be continuously differentiable on an open neighborhood $\widehat{N}$ of $\bar{x}$ and $\Xi : \mathcal{O} \subseteq Y \to Z$ be a locally Lipschitz continuous function on an open set $\mathcal{O}$ containing $\bar{y} := \Psi(\bar{x})$.*

Suppose that Ξ is directionally differentiable at every point in $\mathcal{O}$ and that $\mathcal{J}\Psi(\bar{x}) : X \to Y$ is onto. Then it holds that

$$\partial_B \Phi(\bar{x}) = \partial_B \Xi(\bar{y}) \mathcal{J}\Psi(\bar{x}),$$

where $\Phi : \widehat{N} \to Z$ is defined by

$$\Phi(x) := \Xi(\Psi(x)), \quad x \in \widehat{N}.$$

Proof. By shrinking $\widehat{N}$ if necessary, we may assume that $\Xi(\widehat{N}) \subseteq \mathcal{O}$. Then Ξ is Lipschitz continuous and directionally differentiable on $\mathcal{O}$. By further shrinking $\widehat{N}$ if necessary, we may also assume that for each $x \in \widehat{N}$, $\mathcal{J}\Psi(x)$ is onto.

We shall first show that Φ is F-differentiable at $x \in \widehat{N}$ if and only if Ξ is F-differentiable at $\Psi(x)$, which ensures that

$$\partial_B \Phi(\bar{x}) \subseteq \partial_B \Xi(\bar{y}) \mathcal{J}\Psi(\bar{x}).$$

Certainly, Φ is F-differentiable at $x \in \widehat{N}$ if Ξ is F-differentiable at $\Psi(x)$. Now, suppose that Φ is F-differentiable at $x \in \widehat{N}$. Then, since Ξ is directionally differentiable at $\Psi(x)$, for any $d \in X$ we have

$$\mathcal{J}\Phi(x)d = \Xi'(\Psi(x); \mathcal{J}\Psi(x)d),$$

which implies that for any $s, t \in \Re$ and $u, v \in X$,

$$\Xi'(\Psi(x); s\mathcal{J}\Psi(x)u + t\mathcal{J}\Psi(x)v) = \Xi'(\Psi(x); \mathcal{J}\Psi(x)(su + tv))$$

$$= \mathcal{J}\Phi(x)(su + tv)$$

$$= s\mathcal{J}\Phi(x)u + t\mathcal{J}\Phi(x)v$$

$$= s\Xi'(\Psi(x); \mathcal{J}\Psi(x)u) + t\Xi'(\Psi(x); \mathcal{J}\Psi(x)v).$$

By the surjectivity of $\mathcal{J}\Psi(x)$, we can conclude that $\Xi'(\Psi(x); \cdot)$ is a linear operator and so Ξ is Gâteau differentiable at $\Psi(x)$. Since Ξ is assumed to be locally Lipschitz continuous on $\mathcal{O}$, Ξ is F-differentiable at $\Psi(x)$.

Next, we show that the second half inclusion holds:

$$\partial_B \Phi(\bar{x}) \supseteq \partial_B \Xi(\bar{y}) \mathcal{J}\Psi(\bar{x}).$$

Let $W \in \partial_B \Xi(\bar{y})$. Then there exists a sequence $\{y^k\}$ in $\mathcal{O}$ converging to $\bar{y}$ such that Ξ is F-differentiable at y^k and $W = \lim_{k\to\infty} \mathcal{J}\Xi(y^k)$.

By applying the classical Inverse Function Theorem to

$$\Psi\left(\bar{x} + \mathcal{J}\Psi(\bar{x})^*(y - \bar{y})\right) - \Psi(\bar{x}) = 0,$$

we obtain that there exists a sequence $\{\tilde{y}^k\}$ in $\mathcal{O}$ converging to $\bar{y}$ such that

$$\Psi\left(\bar{x} + \mathcal{J}\Psi(\bar{x})^*(\tilde{y}^k - \bar{y})\right) - \Psi(\bar{x}) = y^k - \Psi(\bar{x})$$

for all k sufficiently large.

Let $\tilde{x}^k := \bar{x} + \mathcal{J}\Psi(\bar{x})^*(\tilde{y}^k - \bar{y})$. Then $y^k = \Psi(\tilde{x}^k)$ and Φ is F-differentiable at $\tilde{x}^k$ with

$$\mathcal{J}\Phi(\tilde{x}^k) = \mathcal{J}\Xi(y^k)\mathcal{J}\Psi(\tilde{x}^k).$$

By using the fact that $\tilde{y}^k \to \bar{y}$ implies $\tilde{x}^k \to \bar{x}$, we know that there exists a $V \in \partial_B\Phi(\bar{x})$ such that

$$W\mathcal{J}\Psi(\bar{x}) = \lim_{k\to\infty}\mathcal{J}\Xi(y^k)\lim_{k\to\infty}\mathcal{J}\Psi(\tilde{x}^k) = \lim_{k\to\infty}\mathcal{J}\Phi(\tilde{x}^k) = V \in \partial_B\Phi(\bar{x}).$$

This completes the proof. □

Proposition 2.9 *For any* $V \in \partial_B \Pi_{\mathcal{S}^p_+}(A)$ (*respectively,* $\partial \Pi_{\mathcal{S}^p_+}(A)$), *there exists a* $W \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ (*respectively,* $\partial \Pi_{\mathcal{S}^{|\beta|}_+}(0)$) *such that*

$$V(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & W(\widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}^T_{\alpha\gamma} \circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T \quad \forall\, H \in \mathcal{S}^p, \tag{11}$$

where $\widetilde{H} := P^T H P$.

Conversely, for any $W \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$ (*respectively,* $\partial \Pi_{\mathcal{S}^{|\beta|}_+}(0)$) , *there exists a* $V \in \partial_B \Pi_{\mathcal{S}^p_+}(A)$ (*respectively,* $\partial \Pi_{\mathcal{S}^p_+}(A)$) *such that (11) holds.*

Proof. We only need to prove that (11) holds for $V \in \partial_B \Pi_{\mathcal{S}^p_+}(A)$ and $W \in \partial_B \Pi_{\mathcal{S}^{|\beta|}_+}(0)$.

Let $\Theta(\cdot) := \Pi'_{\mathcal{S}^p_+}(A;\cdot)$. Define $\Psi : \mathcal{S}^p \to \mathcal{S}^p$ by $\Psi(H) := P^T H P$, $H \in \mathcal{S}^p$ and $\Xi : \mathcal{S}^p \to \mathcal{S}^p$ by

$$\Xi(B) := P \begin{bmatrix} B_{\alpha\alpha} & B_{\alpha\beta} & U_{\alpha\gamma} \circ B_{\alpha\gamma} \\ B_{\alpha\beta}^T & \Pi_{\mathcal{S}^{|\beta|}_+}(B_{\beta\beta}) & 0 \\ B_{\alpha\gamma}^T \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T, \quad B \in \mathcal{S}^p.$$

Then we have

$$\Theta(H) = \Xi(\Psi(H)), \quad H \in \mathcal{S}^p.$$

Since $\Pi_{\mathcal{S}_+^{|\beta|}}$ is directionally differentiable everywhere and $\mathcal{J}\Psi(H) : \mathcal{S}^p \to \mathcal{S}^p$ is onto, we know from Proposition 2.8 that

$$\partial_B \Theta(0) = \partial_B \Xi(0) \mathcal{J}\Psi(0).$$

This, together with Proposition 2.7, completes the proof. □

Next, we consider an application of the variational analysis of the metric projector to a financial engineering problem: Given a symmetric matrix $G \in \mathcal{S}^n$, its nearest correlation matrix is the optimal solution to

$$\begin{array}{ll} \min & \frac{1}{2}\| G - X \|^2 \\ \text{s.t.} & X_{ii} = 1, \quad i = 1, \ldots, n\,, \\ & X \in \mathcal{S}^n_+\,. \end{array} \tag{12}$$

Define: $\mathcal{A} : \mathcal{S}^n \rightarrow \Re^n$ by

$$\mathcal{A}(X) = (X_{11}, X_{22}, \ldots, X_{nn})^T.$$

The adjoint of $\mathcal{A}$ is given by

$$\mathcal{A}^*(y) = \mathrm{diag}(y_1, y_2, \ldots, y_n).$$

Then, by using the Karush-Kuhn-Tucker (KKT) theory, we may solve the correlation matrix problem by solving the equation:

$$F(y) := \mathcal{A}\,(G + \mathcal{A}^* y)_+ - e = 0, \quad y \in \Re^n \,,$$

where $e \in \Re^n$ is the vector of all ones.

Let y^* be a root of $F(y) = 0$. Then we can recover the optimal solution to the correlation matrix problem by letting

$$X^* = (G + \mathcal{A}^* y^*)_+ \,.$$

Indeed, the dual problem is

$$\min \ \theta(y)$$

where

$$\theta(y) = \frac{1}{2}\|\Pi_{\mathcal{S}^n_+}(G + \mathcal{A}^* y)\|^2 - \langle e, y\rangle - \frac{1}{2}||G\|^2, \quad y \in \Re^n .$$

Then we have

$$F(y) = \nabla\theta(y) = \mathcal{A}\,\Pi_{\mathcal{S}^n_+}(G + \mathcal{A}^* y) - e = 0, \quad y \in \Re^n.$$

In numerical computations, we use the following globalized Newton's method for solving the dual problem. Recall that for any $y \in \Re^n$, $\nabla\theta(y) = F(y) - e$.

Algorithm 2.1 (Newton's Method)

Step 0. Given $y^0 \in \Re^n$, $\eta \in (0,1)$, $\rho, \sigma \in (0, 1/2)$. $k := 0$.

Step 1. Select an element $V_k \in \partial F(y^k)$ and apply the conjugate gradient (CG) method of Hestenes and Stiefel to find an approximate solution d^k to

$$\nabla\theta(y^k) + V_k d = 0 \tag{13}$$

such that

$$\|\nabla\theta(y^k) + V_k d^k\| \leq \eta_k \|\nabla\theta(y^k)\| \tag{14}$$

where $\eta_k := \min\{\eta, \|\nabla\theta(y^k)\|\}$.

Step 2. (continued)

If (14) is not achievable or if the condition

$$\nabla\theta(y^k)^T d^k \leq -\eta_k \|d^k\|^2 \tag{15}$$

is not satisfied, let $d^k := -B_k^{-1}\nabla\theta(y^k)$, where B_k is any symmetric positive definite matrix in $\mathcal{S}^n$.

Let m_k be the smallest nonnegative integer m such that

$$\theta(y^k + \rho^m d^k) - \theta(y^k) \leq \sigma\rho^m \nabla\theta(y^k)^T d^k.$$

Set $t_k = \rho^{m_k}$ and $y^{k+1} = y^k + t_k d^k$.

Step 3. Replace k by $k+1$ and go to Step 1.

Theorem 2.1 *Suppose that in Algorithm 2.1 both* $\{\|B_k\|\}$ *and* $\{\|B_k^{-1}\|\}$ *are uniformly bounded. Then the iteration sequence* $\{y^k\}$ *generated by Algorithm 2.1 converges to the unique solution* y^* *of* $F(y) = 0$ *quadratically.*

For details on the above Newton's method for computing the nearest correlation matrix problem, see

- H.-D. Qi and D. Sun. A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications* (2006).

Source code in MatLab is available at
`http://www.math.nus.edu.sg/ matsundf`

The material on the basic properties of metric projectors is quite standard. For the properties on the Jacobian of metric projectors, see the following papers:

- D. Sun and J. Sun. Semismooth matrix valued functions. *Mathematics of Operations Research* 27 (2002) 150–169.

- J.S. Pang, D. Sun, and J. Sun. Semismooth homeomorphisms and strong stability of semidefinite and Lorentz complementarity problems. *Mathematics of Operations Research* 28 (2003) 39–63.

- F. MENG, D. SUN, AND G. ZHAO. Semismoothness of solutions to generalized equations and the Moreau-Yosida regularization. *Mathematical Programming* 104 (2005) 561–581.

- D. SUN. The strong second order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research* 31 (2006).