The Role of Metric Projectors in Nonlinear Conic Optimization

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

August 1, 2007

Let us consider the matrix correlation problem

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

where $G \in \mathcal{S}^n$ is given, but may not be positive semidefinite.

This is a special problem in stress testing in finance.

- The matrix correlation problem can be cast into Semidefinite Programming.

- IMPs can solve the SDP with a small n (e.g., $n = 80$).

- In practice, n can be large (say, $n = 2,000$).

The Lagrangian dual of the matrix correlation problem is

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

where $\Pi_{\mathcal{S}^n_+}(\cdot)$ denotes the metric projection operator onto $\mathcal{S}^n_+$ and $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$:

$$\mathcal{A}^*(y) = \mathrm{Diag}(y) \quad \text{with} \quad \mathcal{A}(X) = \mathrm{diag}(X) \,.$$

The dual is an unconstrained convex problem with a continuously differentiable objective function

$$\nabla\theta(y) = \mathcal{A}\Pi_{\mathcal{S}^n_+}(G + \mathcal{A}^* y) - e\,.$$

Note that $\theta \notin \mathcal{C}^2$ due to the nonsmoothness of $\Pi_{\mathcal{S}^n_+}$. However, this is NOT a problem for us.

The key is to understand $\Pi_{\mathcal{S}^n_+}$.

Let Y be a finite-dimensional real vector Hilbert space. For any $y \in Y$, let $\Pi_K(y)$ denote the metric projection of y onto the closed convex set $K \subseteq Y$:

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

The operator $\Pi_K : Y \to Y$ is called the metric projection operator or metric projector over K.

The point $s \in K$ is an optimal solution to the metric projection problem if and only if it satisfies

$$\langle y - s, d - s \rangle \leq 0 \quad \forall\, d \in K\,.$$

Note that the above property holds even if Y is infinite dimensional.

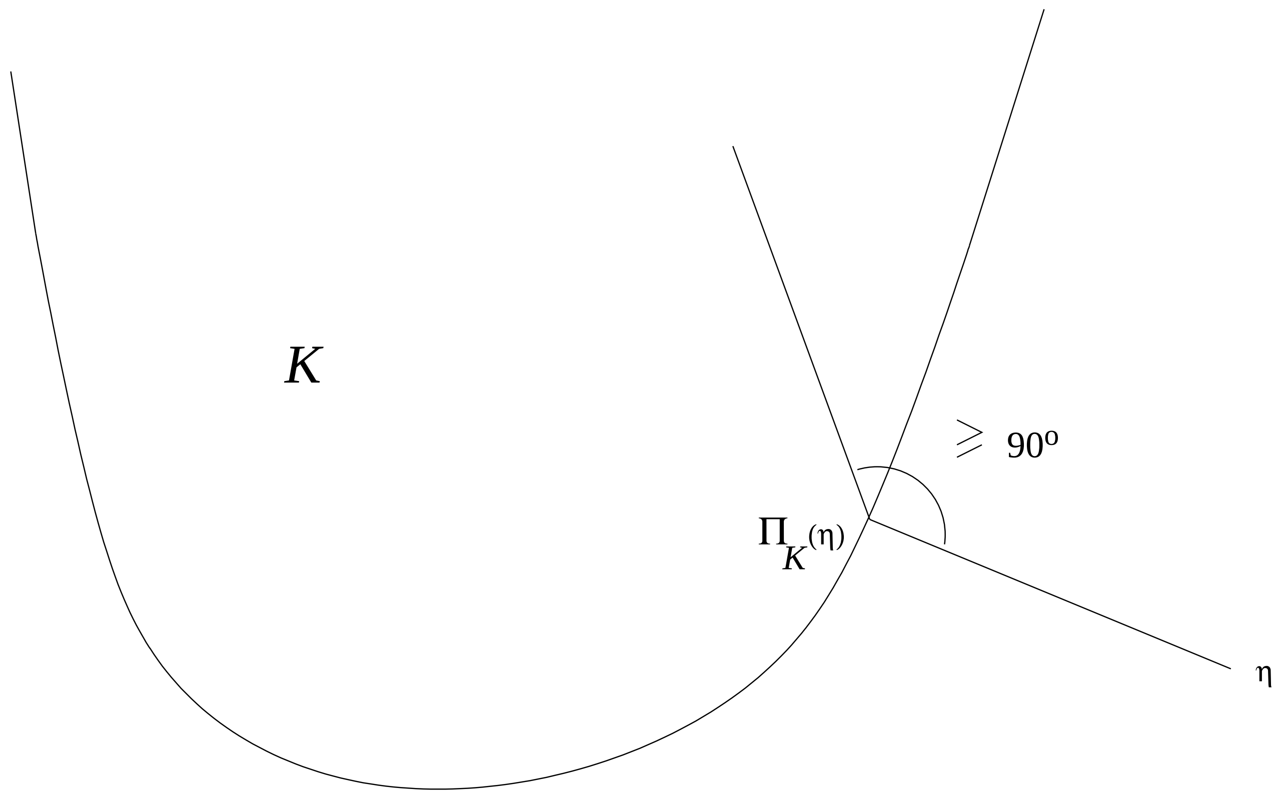

Figure 0.1: Metric projection onto closed convex sets

There are many classical properties for the metric projector. For example, the metric projector $\Pi_K(\cdot)$ satisfies

$$\langle y-z, \Pi_K(y)-\Pi_K(z)\rangle \geq \|\Pi_K(y)-\Pi_K(z)\|^2 \quad \forall\, y, z \in Y\,. \tag{1}$$

Note that (1) implies

$$\|\Pi_K(y) - \Pi_K(z)\| \leq \|y - z\| \quad \forall\, y, z \in Y\,.$$

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

Proposition. Let K be a nonempty closed convex set in Y. Let

$$\theta(y) := \frac{1}{2}\|y - \Pi_K(y)\|^2, \quad y \in Y\,.$$

Then θ is continuously differentiable with

$$\nabla\theta(y) = y - \Pi_K(y)\,, \quad y \in Y\,.$$

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

$$y = \Pi_K(y) + \Pi_{K^o}(y)\,,$$

where K^* is the dual cone of K defined by

$$K^* := \{y \in Y \mid \langle y, d\rangle \geq 0 \ \ \forall\, d \in K\}.$$

Let $\Xi : \mathcal{O} \subseteq Y \to Z$ be a locally Lipschitz continuous function on an open set $\mathcal{O}$. [Z is another finite-dimensional real Hilbert space.]

Then, Rademacher's Theorem says that Ξ is almost everywhere Fréchet differentiable in $\mathcal{O}$.

We denote by $\mathcal{O}_{\Xi}$ the set of points in $\mathcal{O}$ where Ξ is F-differentiable.

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

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

For the metric projector $\Pi_K(\cdot)$, we have

Proposition. For any $y \in Y$ and $V \in \partial \Pi_K(y)$, (a) V is self-adjoint; (b) $\langle d, V d \rangle \geq 0 \quad \forall \, d \in Y$; and (c)

$$V \succeq V^2 \, .$$

For applications, we need more specific characterizations than the above.

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

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

where "Tr" denotes the trace of a square 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

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

Then we know:

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

Note that computing A_+ is equivalent to computing the full eigen-decomposition of A, which in turn needs $9n^3$ flops.

For a typical Pentium IV type desktop PC, it needs about 10 seconds for $n = 1,000$ and less than 90 seconds for $n = 2,000$.

Define

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

Write

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

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.

$\Pi_{\mathcal{S}^p_+}$ is directionally differentiable with $\Pi'_{\mathcal{S}^p_+}(A; H)$ being given by

$$P \begin{bmatrix} P_\alpha^T H P_\alpha & P_\alpha^T H P_\beta & U_{\alpha\gamma} \circ P_\alpha^T H P_\gamma \\ P_\beta^T H P_\alpha & \Pi_{\mathcal{S}^{|\beta|}_+}(P_\beta^T H P_\beta) & 0 \\ P_\gamma^T H P_\alpha \circ U_{\alpha\gamma}^T & 0 & 0 \end{bmatrix} P^T.$$

When $|\beta| = 0$, $\Pi_{\mathcal{S}^n_+}(\cdot)$ is continuously differentiable around A and the above formula reduces to the classical result of Löwner[a].

By using the above formula for $\Pi'_{\mathcal{S}^p_+}(A; H)$, one can easily compute tangent cone of $\mathcal{S}^p_+$ at $A_+ = \Pi_{\mathcal{S}^p_+}(A)$ is

$$\mathcal{T}_{\mathcal{S}^p_+}(A_+) = \left\{B \in \mathcal{S}^p \,:\, P_{\bar\alpha}^T B P_{\bar\alpha} \succeq 0\right\}.$$

where $\bar\alpha := \{1, \ldots, p\} \backslash \alpha$ and $P_{\bar\alpha} := [P_\beta \; P_\gamma]$.

---

[a] K. Löwner. *Über monotone matrixfunktionen.* Mathematische Zeitschrift 38 (1934) 177–216.

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

$$V(H) = P \begin{bmatrix} \widetilde{H}_{\alpha\alpha} & \widetilde{H}_{\alpha\beta} & U_{\alpha\gamma} \circ \widetilde{H}_{\alpha\gamma} \\ \widetilde{H}^T_{\alpha\beta} & W(\widetilde{H}_{\beta\beta}) & 0 \\ \widetilde{H}^T_{\alpha\gamma} \circ U^T_{\alpha\gamma} & 0 & 0 \end{bmatrix} P^T \quad (2)$$

for all $H \in \mathcal{S}^p$, where $\widetilde{H} := P^T H P$. The important fact is that P can be any one used in the eigen-factorization.

One crucial non-classical result is the strong *semismoothness* of $\Pi_{\mathcal{S}^p_+}(\cdot)$ at any $X \in \mathcal{S}^p$:

(i) $\Pi_{\mathcal{S}^p_+}(\cdot)$ is directionally differentiable at X; and

(ii) for any $\mathcal{S}^p \ni Y \to X$ and $V \in \partial \Pi_{\mathcal{S}^p_+}(Y)$,

$$\Pi_{\mathcal{S}^p_+}(Y) - \Pi_{\mathcal{S}^p_+}(X) - V(Y - X) = O(\|Y - X\|)^2.$$

Sensitivity Analysis

Consider the optimization problem

(OP)

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & G(x) \in K, \\ & x \in X\,, \end{aligned}$$

where $f : X \to \Re$ and $G : X \to Y$ are $\mathcal{C}^2$, X, Y finite-dimensional real Hilbert spaces and K is a closed convex set in Y.

The Lagrangian function $L : X \times Y \to \Re$ is defined by

$$L(x, \mu) := f(x) + \langle \mu, G(x) \rangle , \quad (x, \mu) \in X \times Y.$$

Let $\bar{x}$ be a feasible solution to (OP). Robinson's constraint qualification (CQ) is as follows:

$$0 \in \mathrm{int}\{G(\bar{x}) + \mathcal{J}_x G(\bar{x}) X - K\},$$

$$(\text{or} \quad \mathcal{J}_x G(\bar{x}) X + \mathcal{T}_K(G(\bar{x})) = Y),$$

If $\bar{x}$ is a locally optimal solution to (OP) and Robinson's CQ holds at $\bar{x}$, then there exists a Lagrangian multiplier $\bar{\mu} \in Y$, together with $\bar{x}$, satisfying the KKT condition:

$$\mathcal{J}_x L(\bar{x}, \bar{\mu}) = 0 \quad \text{and} \quad \bar{\mu} \in \mathcal{N}_K(G(\bar{x})).$$

Let $\mathcal{M}(\bar{x})$ denote the set of Lagrangian multipliers.

In this talk, we use the nonlinear semidefinite programming as an example to demonstrate the importance of the metric projector in sensitivity analysis

$(NLSDP)$

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & h(x) = 0, \\ & g(x) \in \mathcal{S}^p_+, \\ & x \in X, \end{array}$$

Definition. For any given $B \in \mathcal{S}^p$, define the linear-quadratic function $\Upsilon_B : S^p \times \mathcal{S}^p \to \Re$ by

$$\Upsilon_B(\Gamma, A) := 2\left\langle \Gamma, AB^\dagger A \right\rangle, \quad (\Gamma, A) \in \mathcal{S}^p \times \mathcal{S}^p,$$

where $B^\dagger$ is the Moore-Penrose pseudo-inverse of B.

Proposition. Suppose that $B \in \mathcal{S}^p_+$ and $\Gamma \in \mathcal{N}_{\mathcal{S}^p_+}(B)$. Then for any $V \in \partial \Pi_{\mathcal{S}^p_+}(B+\Gamma)$ and $\Delta B, \Delta\Gamma \in \mathcal{S}^p$ such that

$$\Delta B = V(\Delta B + \Delta \Gamma),$$

it holds that

$$\langle \Delta B, \Delta \Gamma \rangle \geq -\Upsilon_B(\Gamma, \Delta B).$$

Let $\bar{x}$ is a stationary point of $(NLSDP)$. Let $(\bar{\zeta}, \overline{\Gamma}) \in \mathcal{M}(\bar{x})$ such that

$$\mathcal{J}_x L(\bar{x}, \bar{\zeta}, \overline{\Gamma}) = 0, \quad -h(\bar{x}) = 0, \quad \text{and} \quad \overline{\Gamma} \in \mathcal{N}_{\mathcal{S}^p_+}(g(\bar{x})).$$

Let $A := g(\bar{x}) + \bar{\Gamma}$ and

$$g(\bar{x}) = P \begin{bmatrix} \Lambda_\alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} P^T, \text{ and } \quad \overline{\Gamma} = P \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \Lambda_\gamma \end{bmatrix} P^T.$$

The outer approximation set to $\mathrm{aff}(C(\bar{x}))$ with respect to $(\bar{\zeta}, \overline{\Gamma})$ is defined by

$$\begin{aligned}\mathrm{app}(\bar{\zeta}, \overline{\Gamma}) \;=\; \Big\{ d \,:\, \mathcal{J}_x h(\bar{x}) d = 0,\; & P_\beta^T(\mathcal{J}_x g(\bar{x}) d) P_\gamma = 0, \\ & P_\gamma^T(\mathcal{J}_x g(\bar{x}) d) P_\gamma = 0 \Big\}.\end{aligned}$$

Definition. Let $\bar{x}$ be a stationary point of $(NLSDP)$. We say that the strong second order sufficient condition (SSOSC) holds at $\bar{x}$ if

$$\sup_{(\zeta,\Gamma)\in\mathcal{M}(\bar{x})} \left\{ \left\langle d, \mathcal{J}^2_{xx}L(\bar{x},\zeta,\Gamma)d \right\rangle - \Upsilon_{g(\bar{x})}(\Gamma, \mathcal{J}_x g(\bar{x})d) \right\} > 0$$

for all $d \in \widehat{C}(\bar{x})\backslash\{0\}$, where for any $(\zeta,\Gamma) \in \mathcal{M}(\bar{x})$, $(\zeta,\Gamma) \in \Re^m \times \mathcal{S}^p$ and

$$\widehat{C}(\bar{x}) := \bigcap_{(\zeta,\Gamma)\in\mathcal{M}(\bar{x})} \mathrm{app}(\zeta,\Gamma).$$

Next, we define a nondegeneracy condition for $(NLSDP)$, which is an analogue of the LICQ for (NLP). The concept of nondegeneracy originally appeared in Robinson [1984] for (OP).

Definition. We say that a feasible point $\bar{x}$ to (OP) is constraint nondegenerate if

$$\mathcal{J}_x G(\bar{x}) X + \mathrm{lin}(\mathcal{T}_K(\bar{y})) = Y,$$

where $\bar{y} := G(\bar{x})$.

Write down the KKT condition as

$$F(x,\zeta,\Gamma) := \begin{bmatrix} \mathcal{J}_x L(x,\zeta,\Gamma) \\ -h(x) \\ -g(x) + \Pi_{\mathcal{S}^p_+}(g(x)+\Gamma) \end{bmatrix} = 0,$$

which is equivalent to the following generalized equation:

$$0 \in \phi(z) + \mathcal{N}_D(z),$$

where ϕ is $\mathcal{C}^1$ and D is a closed convex set in Z.

Definition. [Robinson'80] Let $\bar{z}$ be a solution of the generalized equation. We say that $\bar{z}$ is a strongly regular solution if there exist neighborhoods $\mathcal{B}$ of the origin $0 \in Z$ and $\mathcal{V}$ of $\bar{z}$ such that for every $\delta \in \mathcal{B}$, the following linearized generalized equation

$$\delta \in \phi(\bar{z}) + \mathcal{J}_z\phi(\bar{z})(z - \bar{z}) + \mathcal{N}_D(z)$$

has a unique solution in $\mathcal{V}$, denoted by $z_{\mathcal{V}}(\delta)$, and the mapping $z_{\mathcal{V}} : \mathcal{B} \to \mathcal{V}$ is Lipschitz continuous.

Let U be a Banach space and $f : X \times U \to \Re$ and $G : X \times U \to Y$.

We say that $(f(x,u), G(x,u))$, with $u \in U$, is a *$\mathcal{C}^2$-smooth parameterization* of (OP) if $f(\cdot,\cdot)$ and $g(\cdot,\cdot)$ are $\mathcal{C}^2$ and there exists a $\bar{u} \in U$ such that $f(\cdot, \bar{u}) = f(\cdot)$ and $G(\cdot, \bar{u}) = G(\cdot)$. The corresponding parameterized problem takes the form:

(OP_u)

$$
\begin{array}{rl}
\min & f(x,u) \\
\text{s.t.} & G(x,u) \in K, \\
 & x \in X\,.
\end{array}
$$

We say that a parameterization is *canonical* if $U := X \times Y$, $\bar{u} = (0, 0) \in X \times Y$, and

$$(f(x, u), G(x, u)) := (f(x) - \langle u_1, x \rangle, G(x) + u_2), \quad x \in X.$$

Definition. [Bonnans and Shapiro'00] Let $\bar{x}$ be a stationary point of (OP). We say that the <u>uniform second order (quadratic) growth condition</u> holds at $\bar{x}$ with respect to a $\mathcal{C}^2$-smooth parameterization $(f(x,u), G(x,u))$ if there exist $c > 0$ and neighborhoods $\mathcal{V}_X$ of $\bar{x}$ and $\mathcal{V}_U$ of $\bar{u}$ such that for any $u \in \mathcal{V}_U$ and any stationary point $x(u) \in \mathcal{V}_X$ of (OP_u), the following holds:

$$f(x,u) \geq f(x(u),u) + c\|x - x(u)\|^2 \quad \forall\, x \in \mathcal{V}_X$$

such that $G(x,u) \in K$.

We say that the uniform second order growth condition holds at $\bar{x}$ if the above inequality holds for every $\mathcal{C}^2$-smooth parameterization of (OP).

Definition. [Kojima'80] and [Bonnans and Shapiro'00]

Let $\bar{x}$ be a stationary point of (OP). We say that $\bar{x}$ is strongly stable with respect to a $\mathcal{C}^2$-smooth parameterization $(f(x,u), G(x,u))$ if there exist neighborhoods $\mathcal{V}_X$ of $\bar{x}$ and $\mathcal{V}_U$ of $\bar{u}$ such that for any $u \in \mathcal{V}_U$, (OP_u) has a unique stationary point $x(u) \in \mathcal{V}_X$ and $x(\cdot)$ is continuous on $\mathcal{V}_U$.

If this holds for any $\mathcal{C}^2$-smooth parameterization, we say that $\bar{x}$ is strongly stable.

Let

$$\Phi(\delta) := F'(\bar{x}, \bar{\zeta}, \overline{\Gamma}; \delta).$$

Theorem. Let $\bar{x}$ be a locally optimal solution to $(NLSDP)$. Suppose that Robinson's CQ holds at $\bar{x}$ so that $\bar{x}$ is necessarily a stationary point of $(NLSDP)$. Let $(\bar{\zeta}, \overline{\Gamma}) \in \Re^m \times \mathcal{S}^p$ be such that $(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is a KKT point of $(NLSDP)$.

Then the following TEN (there are more!) statements are equivalent:

(a) The SSOSC holds at $\bar{x}$ and $\bar{x}$ is constraint nondegenerate.

(b) Any element in $\partial F(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is nonsingular.

(c) The KKT point $(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is strongly regular.

(d) The uniform second order growth condition holds at $\bar{x}$ and $\bar{x}$ is constraint nondegenerate.

(e) The point $\bar{x}$ is strongly stable and $\bar{x}$ is constraint nondegenerate.

(continued)

(f) F is a locally Lipschitz homeomorphism near $(\bar{x}, \bar{\zeta}, \overline{\Gamma})$.

(g) For every $V \in \partial_B F(\bar{x}, \bar{\zeta}, \overline{\Gamma})$, $\operatorname{sgn} \det V = \operatorname{ind}(F, (\bar{x}, \bar{\zeta}, \overline{\Gamma})) = \pm 1$.

(h) Φ is a globally Lipschitz homeomorphism.

(i) For every $V \in \partial_B \Phi(0)$, $\operatorname{sgn} \det V = \operatorname{ind}(\Phi, 0) = \pm 1$.

(j) Any element in $\partial \Phi(0)$ is nonsingular.

Moreover, if f, g, and h are linear functions (linear SDP problems), each of the above statements is equivalent to

(k) Any element in $\partial_B F(\bar{x}, \bar{\zeta}, \overline{\Gamma})$ is nonsingular.

(l) The point $\bar{x}$ is constraint nondegenerate to the primal and $(\bar{\zeta}, \overline{\Gamma})$ is constraint nondegenerate to the dual [Primal-Dual Non-degeneracies].

(m) More ...

Convergence Analysis

Recall that for given $c > 0$, the augmented Lagrangian function for the equality constrained optimization problem

$$\begin{array}{ll} \min_{x \in X} & f(x) \\ \text{s.t.} & h(x) = 0 \,. \end{array}$$

takes the following form

$$L_c(x, y) = f(x) + \langle y, h(x) \rangle + c||h(x)||^2/2 \,.$$

The general optimization problem OP can be written as

$$\begin{array}{ll} \min\limits_{x\in X, s\in Y} & f(x) \\ \text{s.t.} & h(x) = 0\,, \\ & s - g(x) = 0 \\ & s \in K\,. \end{array}$$

So, it is natural to define the augmented Lagrangian function for (OP) as

$$
\begin{aligned}
L_c(x,\zeta,\xi) \; := \; \min_{s\in K} \; \Big\{ & f(x) + \langle \zeta, h(x)\rangle + c||h(x)||^2/2 \\
& + \langle \xi, s-g(x)\rangle + c||s-g(x)||^2/2 \Big\}\,.
\end{aligned}
$$

Suppose that K is a closd convex cone. Then the augmented Lagrangian function reduces to

$$
\begin{aligned}
& L_c(x,\zeta,\xi) \\
\overset{K \text{ is cone}}{=} \; & f(x) + \langle \zeta, h(x) \rangle + c||h(x)||^2/2 \\
& + \frac{1}{2c}(\|\Pi_{K^*}(\xi - cg(x))\|^2 - \|\xi\|^2)\,.
\end{aligned}
$$

The augmented Lagrangian function $L_c(\cdot, \zeta, \xi)$ is continuously differentiable with

$$\begin{aligned}\nabla_x L_c(x, \zeta, \xi) &= \nabla f(x) + \nabla h(x)(\zeta + ch(x)) \\ &\quad - \nabla g(x) \Pi_{K^*}(\xi - cg(x))\,.\end{aligned}$$

Let $c_0 > 0$ be given. Let $(\zeta^0, \xi^0) \in \Re^m \times K^*$ be the initial estimated Lagrange multiplier. At the kth iteration, the augmented Lagrangian method determines x^k by minimizing $L_{c_k}(x, \zeta^k, \xi^k)$, computes (ζ^{k+1}, ξ^{k+1}) by

$$\begin{cases} \zeta^{k+1} := \zeta^k + c_k h(x^k)\,, \\ \xi^{k+1} := \Pi_{K^*}(\xi^k - c_k g(x^k))\,, \end{cases}$$

and updates c_{k+1} by

$$c_{k+1} := c_k \quad \text{or} \quad c_{k+1} := \kappa c_k$$

according to certain rules, where $\kappa > 1$ is a preselected positive number.

Let us again consider $(NLSDP)$ with $K = \mathcal{S}^p_+$. Let $\bar{x}$ be a locally optimal solution to $(NLSDP)$. Suppose that Robinson's CQ holds at $\bar{x}$ so that $\bar{x}$ is necessarily a stationary point of $(NLSDP)$. Suppose that the strong second order sufficient condition and the constraint non-degeneracy hold.

Let $\vartheta_c : \Re^m \times Y \mapsto \Re$ be defined as

$$\vartheta_c(\zeta, \xi) := \min_{x \in \mathbb{B}_\varepsilon(\overline{x})} L_c(x, \zeta, \xi), \quad (\zeta, \xi) \in \Re^m \times Y\,. \tag{3}$$

Since for each fixed $x \in X$, $L_c(x, \cdot)$ is a concave function, $\vartheta_c(\cdot)$ is also a concave function as it is the minimum function of a family of concave functions.

We can choose $\varepsilon > 0$ and $\delta_0 > 0$ such that for any $y \in \mathbb{B}_{\delta_0}(\overline{y})$, $x_c(y)$ is the unique minimizer of $L_c(\cdot, y)$ over $\mathbb{B}_\varepsilon(\overline{x})$ and

$$\vartheta_c(y) = L_c(x_c(y), y)\,, \quad y \in \mathbb{B}_{\delta_0}(\overline{y})\,,$$

where $y := (\zeta, \xi)$.

For any $y \in \mathbb{B}_{\delta_0}(\overline{y})$ with $y = (\zeta, \xi) \in \Re^m \times Y$, let

$$\begin{pmatrix} \zeta_c(y) \\ \xi_c(y) \end{pmatrix} := \begin{pmatrix} \zeta + ch(x_c(y)) \\ \Pi_K(\xi - cg(x_c(y))) \end{pmatrix} . \qquad (4)$$

Then we have

$$\nabla_x L_0(x_c(y), \zeta_c(y), \xi_c(y)) = \nabla_x L_c(x_c(y), y) = 0\,,\; y \in \mathbb{B}_{\delta_0}(\overline{y}). \qquad (5)$$

Let $c \geq c_0$. Then the concave function $\vartheta_c(\cdot)$ defined by (3) is continuously differentiable on $\mathbb{B}_{\delta_0}(\overline{y})$ with

$$\nabla \vartheta_c(y) = \begin{pmatrix} h(x_c(y)) \\ -c^{-1}\xi + c^{-1}\Pi_K(\xi - cg(x_c(y))) \end{pmatrix}$$

for $y = (\zeta, \xi) \in \mathbb{B}_{\delta_0}(\overline{y})$. Moreover, $\nabla \vartheta_c(\cdot)$ is semismooth at any point in $\mathbb{B}_{\delta_0}(\overline{y})$. It is strongly semismooth at any point in $\mathbb{B}_{\delta_0}(\overline{y})$ if $\nabla^2 f, \nabla^2 g$, and $\nabla^2 h$ are locally Lipschitz continuous as $\Pi_{\mathcal{S}^p_+}(\cdot)$ is strongly semismooth everywhere.

The augmented Lagrangian method can locally be regarded as the gradient ascent method applied to the dual problem

$$\max \ \vartheta_c(\zeta,\xi) \quad \text{s.t. } (\zeta,\xi) \in \Re^m \times Y$$

with a constant step-length c, i.e., for all k sufficiently large

$$\begin{pmatrix} \zeta^{k+1} \\ \xi^{k+1} \end{pmatrix} = \begin{pmatrix} \zeta^{k} \\ \xi^{k} \end{pmatrix} + c\nabla\vartheta_c(\zeta^k,\xi^k)\,.$$

Here, we show that locally the augmented Lagrangian method can also be treated as an approximate generalized Newton method applied to the following nonsmooth equation

$$\nabla \vartheta_c(\zeta, \xi) = 0$$

with $-c^{-1}\mathcal{I}$ as a good estimate to elements in $\partial \nabla \vartheta_c(\zeta^k, \xi^k)$ for all (ζ^k, ξ^k) sufficiently close to $(\overline{\zeta}, \overline{\xi})$ and c sufficiently large as every element in $\partial \nabla \vartheta_c(\overline{\zeta}, \overline{\xi})$ is in the form of

$$-c^{-1}\mathcal{I} + O(c^{-2}),$$

where $\mathcal{I}$ is the identity operator in $\Re^m \times Y$.

Since $\nabla\vartheta_c(\cdot,\cdot)$ is semismooth at $(\overline{\zeta},\overline{\xi})$, the fast local convergence of the augmented Lagrangian method comes no surprise for those who are familiar with

"the theory on the superlinear convergence of the generalized Newton method for semismooth equations."

Computational Results

Let us come back to the matrix correlation problem

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

and its dual

$$\max_{y \in \Re^n} -\theta(y) := -\frac{1}{2}\| \Pi_{\mathcal{S}_+^n}(G + \mathcal{A}^* y) \|^2 + e^T y\,,$$

Below are some numerical results by applying a generalized Newton method to the dual problem:

n	cputime	It.	Func.	Tol.
500	34.3 s	8	9	3.7×10^{-9}
1,000	4 m 55 s	9	10	3.1×10^{-9}
1,500	14 m 04 s	9	10	4.5×10^{-7}
2,000	33 m 52 s	9	10	2.6×10^{-6}

References

[1] J.F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*, Springer (New York, 2000).

[2] Zi Xian Chan and D.F. Sun. Constraint nondegeneracy, strong regularity and nonsingularity in semidefinite programming. January 2007. Submitted to *SIAM Journal on Optimization*.

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

[4] Houduo Qi and Defeng Sun. A quadratically convergent Newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Applications* 28 (2006) 360–385.

[5] D.F. Sun. The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. *Mathematics of Operations Research* 31 (2006) 761–776.

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

[7] D.F. Sun and J. Sun. Löwner's operator and spectral functions in Euclidean Jordan algebras. December 2004. Submitted to *Mathematics of Operations Research*.

[8] Defeng Sun, Jie Sun, and Liwei Zhang. The rate of convergence of the augmented Lagrangian method for nonlinear semidefinite programming. *Mathematical Programming, Series A*. Published online: 10 May 2007.