# A semi-smooth Newton method for solving general projection equations*

Nicolas F. Armijo † **Yunier Bello-Cruz**‡ Gabriel Haeser §

Midwest Optimization Meeting, Waterloo 2024

*Supported by NSF Grant # DMS - 2307328
†Institute of Applied Mathematics, IME, USP, Brazil.
‡Department of Mathematical Sciences, NIU, USA.
§Institute of Applied Mathematics, IME, USP, Brazil.

# Schedule

# Projection Equations

## Piecewise Linear Equation (Bellocruz et al., 2016)

$$x^+ + Tx = b.$$

with $x^+$ being the positive part of $x$ and $T$ a matrix.

## Second-Order Cone (Bellocruz et al., 2017)

$$\Pi_{\mathbb{L}^n}(x) + Tx = b.$$

$\mathbb{L}^n$ is the second-order cone.

## General Cone

$$\Pi_{\mathcal{K}}(x) + Tx = b.$$

$\mathcal{K} \subset \mathbb{X}$ is a convex and closed cone, $\Pi_{\mathcal{K}}(x)$ is the projection of $x$ onto $\mathcal{K}$ and $T : \mathbb{X} \to \mathbb{X}$ is a linear mapping.

# Quadratic Programming Application

## Piecewise Linear Case

$$\begin{pmatrix} \min & \frac{1}{2}x^T Qx + q^T x \\ \text{s.t.} & x \in \mathbb{R}^n_+ \end{pmatrix}$$

## Second-Order Cone Case

$$\begin{pmatrix} \min & \frac{1}{2}x^T Qx + q^T x \\ \text{s.t.} & x \in \mathbb{L}^n \end{pmatrix}$$

## General Cone Case

$$\begin{pmatrix} \min & \frac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle \\ \text{s.t.} & x \in \mathcal{K} \end{pmatrix}.$$

where $Q\colon \mathbb{X} \to \mathbb{X}$ is a linear mapping and $\langle \cdot, \cdot \rangle$ is the inner product in $\mathbb{X}$.

# Optimality and Complementarity Conditions

## Conic Quadratic Programming

$$\begin{aligned} Qx + q - \mu &= 0, \\ \langle \mu, x \rangle &= 0. \end{aligned}$$

where $\mu \in \mathcal{K}^*$ is a Lagrange multiplier.

## Equivalent Form

$$\langle Qx + q, x \rangle = 0, \quad x \in \mathcal{K}, \quad Qx + q \in \mathcal{K}^*.$$

# Corresponding Equations

## Piecewise Linear Case

$$(Q - Id)x^+ + x = -q$$

## Case with Second-Order Cone

$$(Q - Id)\Pi_{\mathbb{L}^n} + x = -q$$

## Case with General Cone

$$(Q - Id)\Pi_{\mathcal{K}} + x = -q$$

# Conic Quadratic Programming

## KKT

$$\langle Qx + q, x\rangle = 0, \quad x \in \mathcal{K}, \quad Qx + q \in \mathcal{K}^*.$$

## Projection Equation

$$(Q - Id)\Pi_{\mathcal{K}} + x = -q$$

## Theorem (Bellocruz et al., 2017) - Solutions of the Equation $\rightarrow$ KKT

If $x$ is a solution of the Projection Equation, then $\overline{x} = \Pi_{\mathcal{K}}(x)$ is a solution of the KKT system.

## Theorem - KKT $\rightarrow$ Solutions of the Projection Equation

If $\overline{x}$ is a KKT solution, then $x = \overline{x} - (Q\overline{x} + q)$ is a solution of the Projection Equation.

# Properties of the Projection

We can try to solve the equation using Newton's method, but $\Pi_{\mathcal{K}}$ is not differentiable at some points.

**Theorem**

The projection operator $\Pi_{\mathcal{K}}(\cdot)$ is differentiable almost everywhere. The Jacobian $P'_{\mathcal{K}}(x)$ (when it exists) and the generalized Jacobian $V(x) \in \partial_C \Pi_{\mathcal{K}}(x)$, for all $x \in \mathbb{X}$, are self-adjoint and positive definite operators. Furthermore, the following properties are satisfied:

**(i)** $\|V(x)\| \leq 1$, $\forall V(x) \in \partial_C \Pi_{\mathcal{K}}(x)$ with $x \in \mathbb{X}$.

**(ii)** $V(x)x = \Pi_{\mathcal{K}}(x), \forall V(x) \in \partial_C \Pi_{\mathcal{K}}(x)$ with $x \in \mathbb{X}$.

# Conic Quadratic Programming

## Solving the Equation

We can solve using Newton's method or a variant of Newton's method. For $F(x) = 0$ we use

$$F(x^k) + V(x^k)(x^{k+1} - x^k) = 0,$$

where $V(x) \in \partial_C F(x)$ is Clarke's subdifferential.

## Semismooth Newton for Quadratic Programming

For $F(x) = (Q - Id)\Pi_{\mathcal{K}}(x) + x + q$ it results in

$$((Q - Id)V(x^k) + Id)x^{k+1} = -q,$$

where $V(x) \in \partial_C \Pi_{\mathcal{K}}(x)$ is Clarke's subdifferential.

# Some Properties

**Proposition**

If $\|Q - Id\| < 1$, then $(Q - Id)\Pi_{\mathcal{K}}(x) + x = -q$ has a unique solution for all $q \in \mathbb{X}$.

**Proposition**

If $Q$ is nonsingular and $\|Q^{-1} - Id\| < 1$, then $(Q - Id)\Pi_{\mathcal{K}}(x) + x = -q$ has a unique solution for all $q \in \mathbb{X}$.

**Proposition**

If $V(x^{k+1}) = V(x^k)$, then $x^{k+1}$ is a solution of the equation.

# Sufficient Conditions for Convergence

## Theorem

Let $q \in \mathbb{X}$ and $Q\colon \mathbb{X} \to \mathbb{X}$ be a linear operator. Suppose that $Q - Id$ has an inverse and $\|Q - Id\| < 1$. Then, the equation has a unique solution $\overline{x}$, and for any initial point $x^0$, the sequence generated by the semismooth Newton method $\{x^k\}$ is well-defined. Additionally, if $\|Q - Id\| < \frac{1}{2}$ then the method converges $Q$-linearly to $\overline{x}$ satisfying

$$\|x^{k+1} - \overline{x}\| \leq \frac{\|Q - Id\|}{1 - \|Q - Id\|}\|x^k - \overline{x}\|, \; k \in \mathbb{N}.$$

# Sufficient Conditions for Convergence

**Theorem**

Let $q \in \mathbb{X}$ and $Q\colon \mathbb{X} \to \mathbb{X}$ be a positive definite operator. Then for any $x^0$, the sequence generated by the semismooth Newton method $\{x^k\}$ is well-defined. Additionally, if $Q - Id$ is nonsingular, then the equation has a unique solution $\overline{x}$, and if $\|Q - Id\| < 1$ the sequence converges $Q$-linearly to $\overline{x}$ and satisfies

$$\|x^{k+1} - \overline{x}\| \leq \|Q - Id\| \|x^k - \overline{x}\|, \; k \in \mathbb{N}.$$

# Nearest Correlation Matrix Problem

## Definition

$$\begin{pmatrix} \min & \frac{1}{2}\|X - G\|^2 \\ \text{s.t.} & \operatorname{diag}(X) = e \\ & X \in \mathbb{S}^n_+ \end{pmatrix},$$

where $e$ is the vector of 1s and $\operatorname{diag}(X)$ returns the diagonal vector of $X$.

## Linear Constraint

The linear constraint $\operatorname{diag}(X) = e$ does not fit directly. We have to generalize it!!

# Quadratic Cone Problem with Linear Constraints

## Definition

$$\begin{pmatrix} \min & \frac{1}{2}\langle x, Qx\rangle + \langle q, x\rangle \\ \text{s.t.} & \mathcal{A}x = b \\ & x \in \mathcal{K} \end{pmatrix},$$

where $\mathcal{A}\colon \mathbb{X} \to \mathbb{Y}$ is a linear mapping and $b \in \mathbb{Y}$.

## Optimality and Complementarity Conditions

$$\begin{aligned} Qx + q + \mathcal{A}^*\lambda - \mu &= 0 \\ \mathcal{A}x - b &= 0 \\ \langle \mu, x\rangle &= 0, \end{aligned}$$

where $\mu \in \mathcal{K}^*$ and $\lambda \in \mathbb{Y}$.

# Quadratic Cone Problem with Linear Constraints

## Equivalent Formulation

$$\left\langle \begin{pmatrix} Qx + \mathcal{A}^*\lambda + q \\ \mathcal{A}x - b \end{pmatrix}, \begin{pmatrix} x \\ \lambda \end{pmatrix} \right\rangle = 0, \quad \begin{pmatrix} Qx + \mathcal{A}^*\lambda + q \\ \mathcal{A}x - b \end{pmatrix} \in K^*.$$

with $(x, \lambda) \in K := \mathcal{K} \times \mathbb{Y}$ and $K^* := \mathcal{K}^* \times \{0\}$.

## Equation for Problem with Linear Constraints

$$\begin{pmatrix} (Q - Id)\Pi_{\mathcal{K}}(x) + \mathcal{A}^*\lambda + x \\ \mathcal{A}\Pi_{\mathcal{K}}(x) \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix},$$

# Quadratic Cone Problem with Linear Constraints

## System of Equations for the Problem

$$\left(\begin{pmatrix} Q & \mathcal{A}^* \\ \mathcal{A} & 0 \end{pmatrix} - Id\right) \Pi_K(x, \lambda) + \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix},$$

The same results can be applied but now for the matrix (linear mapping) $\begin{pmatrix} Q & \mathcal{A}^* \\ \mathcal{A} & 0 \end{pmatrix}$.

# Quadratic Cone Problem with Linear Constraints

## Semismooth Newton for the Quadratic Problem with Linear Constraints

$$\begin{pmatrix} (Q - Id)V(x^k)x^{k+1} + x^{k+1} + \mathcal{A}^*\lambda^{k+1} \\ \mathcal{A}V(x^k)x^{k+1} \end{pmatrix} = \begin{pmatrix} -q \\ b \end{pmatrix},$$

# Nearest Correlation Matrix

## Returning to the Problem

$$\begin{pmatrix} \min & \frac{1}{2}\|X - G\|^2 \\ \text{s.t.} & \mathcal{A}(X) = b \\ & X \in \mathbb{S}^n_+ \end{pmatrix}.$$

## Iteration

$$\begin{pmatrix} X^{k+1} + \mathcal{A}^*(\Lambda^{k+1}) \\ \mathcal{A}V(X^k)X^{k+1} \end{pmatrix} = \begin{pmatrix} G \\ b \end{pmatrix}.$$

# Nearest Correlation Matrix

## Problem

$$\begin{pmatrix} \min & \frac{1}{2}\|X - G\|^2 \\ \text{s.t.} & \text{diag}(X) = e \\ & X \in \mathbb{S}^n_+ \end{pmatrix}.$$

## Equations and Iteration

$$\begin{pmatrix} X + \text{Diag}(\lambda) \\ \text{diag}(\Pi_{\mathbb{S}^n_+}(X)) \end{pmatrix} = \begin{pmatrix} G \\ e \end{pmatrix}.$$

The semismooth Newton method for the Nearest Correlation Matrix results in

$$\begin{pmatrix} X^{k+1} + \text{Diag}(\lambda^{k+1}) \\ \text{diag}(V(X^k)X^{k+1}) \end{pmatrix} = \begin{pmatrix} G \\ e \end{pmatrix}.$$

# Nearest Correlation Matrix

## Observation

The off-diagonal elements of $X^{k+1}$ must be equal to the off-diagonal elements of $G$. We define $D^{k+1} = \text{Diag}(\text{diag}(X^{k+1}))$ and $\hat{G} = G - \text{Diag}(\text{diag}(G))$, obtaining

$$X^{k+1} = D^{k+1} + \hat{G},$$
$$\lambda^{k+1} = \text{diag}(G) - \text{diag}(D^{k+1}).$$

## Final Iteration

After some calculations substituting into $\text{diag}(V(X^k)X^{k+1}) = e$ we get

$$\text{diag}(D^{k+1}) = (\text{Diag}(\text{diag}(V(X^k)))^{-1}[e - \text{diag}(V(X^k)\hat{G})].$$

# Nearest Correlation Matrix

## Choice of $V(X)$

A choice for $V(X)$ is the subdifferential by Malick, 2006.
We use

$$V(X) = UDU^T,$$

where $X = U\Lambda U^T$, $D_{ii} = 1$ if $\Lambda_{ii} > 0$ and $D_{ii} = 0$ if $\Lambda_{ii} \leq 0$.

## Proposition

Let $X \in \mathbb{S}^n$. If $\mathrm{diag}(X) > 0$, then $\mathrm{diag}(V(X)) > 0$.

# Numerical Experiments

## Proposed Method

Qi and Sun in 2006 applied semismooth Newton to

$$\tilde{F}(y) := \mathcal{A}\Pi_{\mathbb{S}^n_+}(G + \mathcal{A}^* y) - e,$$

in particular for

$$\tilde{F}(y) := \mathrm{diag}(\Pi_{\mathbb{S}^n_+}(G + \mathrm{Diag}(y))) - e.$$

# Numerical Experiments

## Qi and Sun, 2006

- They use the Clarke subdifferential found by Malick in 2006.
- They use Conjugate Gradients to solve the linear systems.

## Higham, 2010

- Uses minres preconditioning the matrix $G$ before iterating, $D^{-\frac{1}{2}}GD^{-\frac{1}{2}}$.

# Numerical Experiments

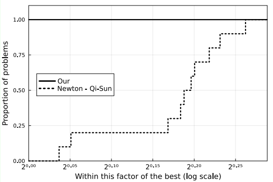
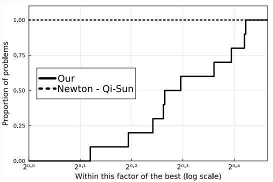
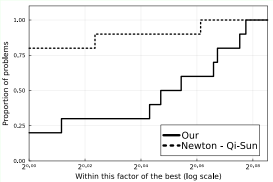


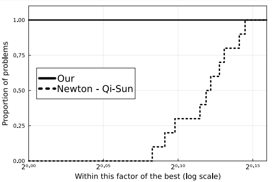

Figura: Performance profile




(a) $n = 5000$




(b) $n = 8000$




(c) $n = 10000$

Figura: Performance profiles

# Work in Progress

## Nonlinear Cone Problem

The problem and method can be generalized to the NCP problem.

$$\begin{pmatrix} \min & f(x) \\ \text{s.t.} & g(x) \in \mathcal{K} \end{pmatrix}$$

with $f : \mathbb{X} \to \mathbb{R}$, $g : \mathbb{X} \to \mathbb{Y}$ and $\mathcal{K} \subset \mathbb{Y}$ being a convex closed cone.

# Work in Progress

## KKT System

$$\begin{aligned} \nabla f(x) - (Dg(x))^*\lambda &= 0 \\ \langle \lambda, g(x) \rangle &= 0 \\ g(x) &\in \mathcal{K}. \end{aligned}$$

## Equivalent Form

$$\left\langle \begin{pmatrix} \nabla f(x) - (Dg(x))^*\lambda \\ g(x) \end{pmatrix}, \begin{pmatrix} x \\ \lambda \end{pmatrix} \right\rangle = 0, \begin{pmatrix} \nabla f(x) - (Dg(x))^*\lambda \\ g(x) \end{pmatrix} \in K^*$$

$(\overline{x}, \overline{\lambda}) \in \mathrm{K}$, where $\mathrm{K} := \mathbb{X} \times \mathcal{K}^*$ and therefore $\mathrm{K}^* = \{0\} \times \mathcal{K}$.

# Work in Progress

## Equations for the NCP

The equation for the new case is

$$\begin{aligned} \nabla f(x) - (Dg(x))^* \Pi_{\mathcal{K}^*}(\lambda) &= 0 \\ g(x) - \Pi_{\mathcal{K}^*}(\lambda) + \lambda &= 0. \end{aligned}$$

We can return to the original cone using the Moreau decomposition $\lambda = \Pi_{\mathcal{K}}(\lambda) - \Pi_{\mathcal{K}^*}(-\lambda)$.

## Theorem

- If $(x, \lambda)$ solves the system of equations, then $(x, \Pi_{\mathcal{K}^*}(\lambda))$ solves KKT.
- If $(x, \sigma)$ is a solution of the KKT system, then $(x, \lambda)$ is a solution of the system of equations $\lambda := \sigma - g(y)$.

# Work in Progress

## Particular Case $g \equiv Id$

$$\begin{aligned} \nabla f(x) - \Pi_{\mathcal{K}^*}(\lambda) &= 0 \\ x - \Pi_{\mathcal{K}^*}(\lambda) + \lambda &= 0. \end{aligned}$$

Resulting in

$$\nabla f(\Pi_{\mathcal{K}}(y)) - \Pi_{\mathcal{K}}(y) + y = 0.$$

## Proposition

Let $f : \mathbb{X} \to \mathbb{R}$ such that $\nabla f$ is Lipschitz continuous. If $\|\mathrm{Id} - \nabla^2 f(z)\| < 1, \forall z$ then the equation has a unique solution.

# Work in Progress

## Semismooth Newton for NCP

$$\begin{pmatrix} \nabla^2 f(x^k) - (D^2 g(x^k))^* \Pi_{\mathcal{K}^*}(\lambda^k) & -(Dg(x^k))^* V_{\mathcal{K}^*}(\lambda^k) \\ Dg(x^k) & Id - V_{\mathcal{K}^*}(\lambda^k) \end{pmatrix} \begin{pmatrix} x^{k+1} - x^k \\ \lambda^{k+1} - \lambda^k \end{pmatrix} + \begin{pmatrix} \nabla f(x^k) - (Dg(x^k))^* \Pi_{\mathcal{K}^*}(\lambda^k) \\ g(x^k) - \Pi_{\mathcal{K}^*}(\lambda^k) + \lambda^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

# Work in Progress

## Mixed Projection Cone Problem

$$\begin{pmatrix} \min & \langle C, X\rangle + \lambda \text{Rank}(X) \\ \text{s.t.} & AX = B \\ & \text{Rank}(X) \leq \sigma \\ & x \in \mathcal{K}. \end{pmatrix}$$

$A \in \mathbb{R}^{l\times n}$, $B \in \mathbb{R}^{l\times m}$, $X \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{n\times m}$, $\sigma \geq 0$ and $\mathcal{K}$ is a cone.

## Some Applications

- Low-Rank Matrix Completion.
- Minimum Dimension Euclidean Distance Embedding.
- Quadratically Constrained Quadratic Optimization (QCQP) Relaxation.

# Work in Progress

## Proposition (Bertsimas, 2022)

For any $X \in \mathbb{R}^{n\times m}$, $\text{Rank}(X) \leq \sigma \iff \exists Y \in \mathcal{Y}_n : \text{Tr}(Y) \leq \sigma$, $X = YX$. Where $\mathcal{Y}_n := \{P \in \mathbb{S}^n : P^2 = P\}$ is the set of orthogonal projection matrices of size $n \times n$.

## Resulting Problem

$$\begin{pmatrix} \min & \langle C, X\rangle + \lambda \text{Tr}(Y) \\ \text{s.t.} & AX = B \\ & X = YX \\ & Y^2 = Y \\ & \text{Tr}(Y) \leq \sigma \\ & Y \in \mathbb{S}^n \\ & X \in \mathcal{K}. \end{pmatrix}$$

# Conclusions and Future Ideas

- We managed to generalize the results for cone programming.
- We applied the method to the Nearest Correlation Matrix problem obtaining some results, but we can improve!!
- Consider another choice for $V(x)$.
- Consider the Newton matrix and the step size of the method, especially for the problems of interest.