A Barrier-Based Smoothing Proximal Point Algorithm for Nonlinear Complementarity Problems over Closed Convex Cones

IMS Workshop on Optimization, Singapore

21st November 2012

Chek Beng CHUA
Nanyang Technological University
cbchua@ntu.edu.sg

Pre-print @ http://www.ntu.edu.sg/home/cbchua

Nonlinear Complementarity Problem

$\mathrm{NCP}_K(F)$:

Given continuous nonlinear $F : \mathbb{E} \to \mathbb{E}$, and closed convex cone $K \subset \mathbb{E}$, find $x \in \mathbb{E}$ such that

$$x \in K, \quad F(x) \in K^\sharp, \quad \text{and} \quad \langle x, F(x) \rangle = 0$$

$(\mathbb{E}, \langle \cdot, \cdot \rangle)$: finite dimensional Euclidean space
$K^\sharp$: dual cone of K

Normal Map Equation (NME):

$$F(\Pi_K(z)) + z - \Pi_K(z) = 0$$

$\Pi_K : z \in \mathbb{E} \mapsto \arg\min_{x \in K}\{\frac{1}{2}\|z - x\|^2 = \frac{1}{2}\langle z - x, z - x \rangle\}$ is the *Euclidean projector*

$$z \text{ solves NME} \quad \Leftrightarrow \quad x = \Pi_K(z) \text{ solves } \mathrm{NCP}_K(F)$$

Nonlinear Complementarity Problem

Optimization over convex cones:

$$\max\{-h(y) : b - A^T y \in K^\sharp\}$$

has sufficient optimality conditions

$$-\nabla h(y) + Ax = 0, \quad b - A^T y \in K^\sharp, \quad x \in K, \quad \langle x, b - A^T y \rangle = 0$$

or equivalently,

$$K \times \mathbb{R}^m \ni \begin{bmatrix} x \\ y \end{bmatrix} \perp \begin{bmatrix} b - A^T y \\ -\nabla h(y) + Ax \end{bmatrix} \in K^\sharp \times \{0\} = (K \times \mathbb{R}^m)^\sharp$$

NME:

$$\begin{bmatrix} b - A^T y \\ -\nabla h(y) + A\Pi_K(z) \end{bmatrix} + \begin{bmatrix} z - \Pi_K(z) \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

PPA:

$$\boxed{F(\Pi_K(z)) + z - \Pi_K(z) + c_k^{-1}(z - z_k) = 0}$$

PPA/NP:

$$\boxed{F(\Pi_K(z)) + z - \Pi_K(z) + c_k^{-1}(\Pi_K(z) - \Pi_K(z_k)) = 0}$$

SPPA:

$$\boxed{F(p(z,\mu)) + z - p(z,\mu) + c_k^{-1}(p(z,\mu) - p(z_k,\mu_k)) = 0, \quad \mu = \frac{\gamma_k^{-1}}{1+\gamma_k^{-1}}\mu_k}$$

SPPA/CCO:

$$\boxed{\begin{bmatrix} b - A^T y + z - p(z,\mu) \\ -\nabla h(y) + Ap(z,\mu) \end{bmatrix} + \begin{bmatrix} c_k^{-1}(p(z,\mu) - p(z_k,\mu_k)) \\ d_k^{-1}(y - y_k) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \; \mu = \frac{\gamma_k^{-1}}{1+\gamma_k^{-1}}\mu_k}$$

SCE:

$$\boxed{-\nabla h(y) + \sqrt{c_k}Ap(\sqrt{c_k}A^T y - b_k, \mu) + d_k^{-1}(y - y_k) = 0}$$

Proximal Point Algorithm

Proximal Point Algorithm (PPA):

Finds a zero of set-valued $T : \mathbb{H} \rightrightarrows \mathbb{H}$ by iteratively finding approximate zeros x_{k+1} of

$$T_k : x \mapsto T(x) + c_k^{-1}(x - x_k)$$

$\mathbb{H}$ = Hilbert space
$\{c_k\}$ = sequence of positive real numbers

Moreau-Yosida regularization (with $T = \partial f$):

$$y = \arg\min_x \{f(x) + \tfrac{1}{2c_k}\|y - x_k\|^2\} \quad \Leftrightarrow \quad T_k(y) = 0$$

Theorem (Convergence of PPA, Rockafellar, 1976). *If T is maximal monotone, $\{c_k^{-1}\}$ is bounded with $\{c_k\|T_k(x_{k+1})\|\}$ summable and $T^{-1}(0) \neq \emptyset$, then*

$$\lim_{k\to\infty} x_k \in T^{-1}(0)$$

Maximal Monotonicity

Definition (Maximal Monotone). $T : \mathbb{H} \rightrightarrows \mathbb{H}$ is *monotone* if

$$\forall (z,w),(z',w') \in \mathcal{G}(T),\ \ \langle z - z', w - w' \rangle_{\mathbb{H}} \geq 0$$

$\mathcal{G}(T)$: graph of T; i.e., $\{(z,w) \in \mathbb{H}^2 : w \in T(z)\}$
It is *maximal* if $\mathcal{G}(T) \nsubseteq \mathcal{G}(T')$ $\forall T'$ monotone

Proximal mapping: For monotone $T : \mathbb{H} \rightrightarrows \mathbb{H}$

$$\forall (z,w),(z',w') \in \mathcal{G}(cT + I),\ \ \langle z - z', w - w' \rangle_{\mathbb{H}} \geq \|z - z'\|_{\mathbb{H}}^2$$

$\therefore$ the *proximal mapping* $(cT + I)^{-1}$ is nonexpansive, whence single-valued

The PPA approximately finds $(c_k T + I)^{-1}(x_k)$ at each iteration

Theorem (Minty's criterion, Minty, 1962). $\forall c > 0$, $\forall$ *monotone* $T : \mathbb{H} \rightrightarrows \mathbb{H}$

$$T \text{ is maximal monotone} \quad \Leftrightarrow \quad \operatorname{dom}((cT + I)^{-1}) = \mathbb{H}$$

Monotonicity of Normal Map

F maximal monotone $\quad \Rightarrow \quad x \mapsto F(x) + \Pi_K^{-1}(x) - x$ maximal monotone

Π_K is maximal monotone but $z \mapsto F(\Pi_K(z)) + z - \Pi_K(z)$ may not be, even when F is

Conic optimization:

$$\begin{bmatrix} x \\ y \end{bmatrix} \in K \times \mathbb{R}^m \mapsto \begin{bmatrix} b - A^T y \\ -\nabla h(y) + Ax \end{bmatrix}$$ is maximal monotone when h is convex

Nonlinear Proximal Term

PPA with nonlinear proximal term: Consider instead

For some $R : \mathbb{H} \to \mathbb{H}$, iteratively find approximate zeros z_{k+1} of

$$T_k : z \mapsto T(z) + c_k^{-1}(R(z) - R(z_k))$$

Corollary (Convergence of PPA with nonlinear proximal term)**.** *If TR^{-1} is maximal monotone, $\{c_k^{-1}\}$ is bounded with $\{c_k \|T_k(z_{k+1})\|\}$ summable and $T^{-1}(0) \neq \emptyset$, then*

$$\lim_{k\to\infty} R(z_k) \in R(T^{-1}(0))$$

Corollary. *If F is maximal monotone, $\{c_k^{-1}\}$ is bounded with*

$$\{c_k \|F(\Pi_K(z_{k+1})) + z_{k+1} - \Pi_K(z_{k+1}) + c_k^{-1}(\Pi_K(z_{k+1}) - \Pi_K(z_k))\|\}$$

summable and $\mathrm{NCP}_K(F)$ has a solution, then $\Pi_K(z_k)$ converges to a solution.

Smoothing Approximation

Definition (Smoothing Approximation of Euclidean Projector)**.**
A *smoothing approximation of the Euclidean projector* is a C^1 map $p : \mathbb{E} \times \mathbb{R}_{++} \to \mathbb{E}$ satisfying

$$p(\cdot, \mu) \underset{\mu \to 0}{\longrightarrow} \Pi_K.$$

It is said to be *uniform* if the convergence is uniform.

Examples:

- $K = \mathbb{R}_+$: $\quad p(z, \mu) = \dfrac{z + \sqrt{z^2 + 4\mu^2}}{2} \underset{\mu \to 0}{\longrightarrow} \Pi_{\mathbb{R}_+}(z) = \max\{0, z\} =: z_+$
- $K = \mathbb{S}_+$: $\quad p(Z, \mu) = \dfrac{Z + \sqrt{Z^2 + 4\mu^2 I}}{2} \underset{\mu \to 0}{\longrightarrow} \Pi_{\mathbb{S}_+}(Z) = \sum_{i=1}^{n} (\lambda_i)_+ q_i q_i^T$

Smoothing Approximation

Non-interior continuation methods:

$$x \geq 0, s \geq 0, xs = 0 \quad \Leftrightarrow \quad x = (x-s)_+$$

$$\updownarrow$$

$$x \geq 0, s \geq 0, xs = \mu^2 \quad \Leftrightarrow \quad x = \frac{(x-s) + \sqrt{(x-s)^2 + 4\mu^2}}{2}$$

Log-determinant optimization problem: Wang, Sun and Toh (2010) observed that "the term $-\mu \log \det X$ acts as a smoothing term."

Smoothing Approximation

Barrier for K: A convex C^2 function $f : \mathrm{int}(K) \to \mathbb{R}$ with positive definite Hessians $\nabla^2 f(x)$ such that

$$f(x_k) \to \infty \quad \forall \mathrm{int}(K) \ni x_k \to \mathrm{bd}(K)$$

$\partial f : x \mapsto \begin{cases} \{\nabla f(x)\} & \text{if } x \in \mathrm{int}(K), \\ \emptyset & \text{if } x \notin \mathrm{int}(K) \end{cases}$ is a maximal monotone map

Minty's criterion
$\Rightarrow \quad \forall \mu > 0$, $(I + \mu \partial f(\cdot/\mu))^{-1}$ is a bijection from $\mathbb{E}$ to $\mathrm{int}(K)$

Theorem (Barrier-based smoothing approximation).
$p : (z, \mu) \mapsto (I + \mu \partial f(\cdot/\mu))^{-1}(z)$ *is a smoothing approximation of* Π_K *if*

$$\limsup_{\mu \to 0^+, y \to x} \mu \partial f(y/\mu) \subseteq N_K(x) \quad \forall x \in \mathbb{E}$$

Corollary. f *is logarithmically homogeneous*
$\Rightarrow \quad p$ *is a smoothing approximation of* Π_K

Auxiliary Equation

Smoothed Normal Map Equation:

$$F(p(z,\mu_+)) + z - p(z,\mu_+) = 0 \quad \text{where } p(\cdot,0) := \Pi_K$$

Auxiliary equation:

$$\begin{bmatrix} F(p(z,\mu_+)) + z - p(z,\mu_+) \\ \mu_+ \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

PPA/NP for auxiliary equation:

Iteratively find approximate zeros (z_k, μ_k) of

$$T_k : (z,\mu) \mapsto \begin{bmatrix} F(p(z,\mu_+)) + z - p(z,\mu_+) \\ \mu_+ \end{bmatrix} + \begin{bmatrix} c_k^{-1}(p(z,\mu_+) - p(z_k,(\mu_k)_+)) \\ \gamma_k^{-1}(\mu_+ - (\mu_k)_+) \end{bmatrix}$$

We used $R : (z,\mu) \in \mathbb{E} \times \mathbb{R} \mapsto (p(z,\mu_+),\mu_+)$

Smoothing PPA

SPPA for NME:

Starting with $(z_0, \mu_0) \in \mathbb{E} \times \mathbb{R}_{++}$, iteratively set $\mu_{k+1} = \frac{\gamma_k^{-1}}{1+\gamma_k^{-1}} \mu_k$ and find approximate solution z_{k+1} to

$$F(p(z, \mu_{k+1})) + z - p(z, \mu_{k+1}) + c_k^{-1}(p(z, \mu_{k+1}) - p(z_k, \mu_k)) = 0,$$

Theorem. *If $C_k : \mathbb{H} \to \mathbb{H}$ are bijective linear maps, $C_k T R^{-1}$ are maximal monotone, $\{C_k^{-1}\}$ is bounded with $\{\|C_k(T(z_{k+1}) + C_k^{-1}(R(z_{k+1}) - R(z_k)))\|\}$ summable and $T^{-1}(0) \neq \emptyset$, then*

$$\lim_{k \to \infty} R(z_k) \in R(T^{-1}(0))$$

Smoothing PPA

Lemma. *For* $C : (z, \mu) \in \mathbb{E} \times \mathbb{R} \mapsto (cz, \gamma\mu)$, $R : (z, \mu) \mapsto (p(z, \mu_+), \mu_+)$ *with the barrier* f *satisfying*

$$c\gamma^{-1} \sup_{x \in \mathrm{int}(K)} \left\langle \nabla f(x) - \nabla^2 f(x) x, (\nabla^2 f(x))^{-1} \nabla f(x) - x \right\rangle \leq 4\omega$$

for some $\omega > 0$ *and* F *monotone, then* $C^{-1} T R^{-1}$ *is maximal monotone under the inner product* $((x, \mu), (x', \mu')) \mapsto \langle x, x' \rangle + \omega \mu \mu'$

Lemma. *The above* $\sup$ *is* ϑ *if* f *is* ϑ*-logarithmically homogeneous*

Theorem (Convergence of SPPA for NME). *If* F *is monotone,* f *is logarithmically homogeneous,* $\{c_k^{-1}\}$, $\{\gamma_k^{-1}\}$ *and* $\{c_k \gamma_k^{-1}\}$ *are bounded with*

$$\{c_k \| F(p(z, \mu_{k+1})) + z - p(z, \mu_{k+1}) + c_k^{-1} (p(z, \mu_{k+1}) - p(z_k, \mu_k)) \| \}$$

summable and $\mathrm{NCP}_K(F)$ *has a solution, then* $p(z_k, \mu_k)$ *converges to a solution.*

Convex Conic Optimization

In each iteration, we need to approximate the solution to

$$\begin{bmatrix} b - A^T y + z - p(z, \mu_{k+1}) \\ -\nabla h(y) + Ap(z, \mu_{k+1}) \end{bmatrix} + \begin{bmatrix} c_k^{-1}(p(z, \mu_{k+1}) - p(z_k, \mu_k)) \\ d_k^{-1}(y - y_k) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Change of variable: $w = \sqrt{c_k}(z - p(z, \mu_{k+1})) + p(z, \mu_{k+1})/\sqrt{c_k}$

Logarithmic homogeneity of f

$\Rightarrow \quad p(w, \mu_{k+1}) = p(z, \mu_{k+1})/\sqrt{c_k} \quad$ and $\quad w - p(w, \mu_{k+1}) = \sqrt{c_k}(z - p(z, \mu_{k+1}))$

$$\Rightarrow \quad z - p(z, \mu_{k+1}) + c_k^{-1} p(z, \mu_{k+1}) = (w - p(w, \mu_{k+1}))/\sqrt{c_k} + p(w, \mu_{k+1})/\sqrt{c_k} = w/\sqrt{c_k}$$

Proximal mapping equation then becomes

$$\begin{bmatrix} b - A^T y + w/\sqrt{c_k} - c_k^{-1} p(z_k, \mu_k) \\ -\nabla h(y) + \sqrt{c_k} A p(w, \mu_{k+1}) + d_k^{-1}(z - z_k) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Change of Variable

Denote $b_k = \sqrt{c_k}(b - c_k^{-1} p(z_k, \mu_k))$ in

$$\begin{bmatrix} b - A^T y + w/\sqrt{c_k} - c_k^{-1} p(z_k, \mu_k) \\ -\nabla h(x) + \sqrt{c_k} A p(w, \mu_{k+1}) + d_k^{-1}(y - y_k) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

gives

$$\begin{bmatrix} w \\ -\nabla h(x) + \sqrt{c_k} A p(w, \mu_{k+1}) + d_k^{-1}(y - y_k) \end{bmatrix} = \begin{bmatrix} \sqrt{c_k} A^T y - b_k \\ 0 \end{bmatrix}$$

Schur complement equation:

$$-\nabla h(x) + \sqrt{c_k} A p(\sqrt{c_k} A^T y - b_k, \mu_{k+1}) + d_k^{-1}(y - y_k) = 0$$

and set $w = \sqrt{c_k} A^T y - b_k$

Recovering z: $\quad z = (w - p(w, \mu_{k+1}))/\sqrt{c_k} + \sqrt{c_k} p(w, \mu_{k+1})$

Solving Schur Complement Equation

Recall: $p(\cdot, \mu) = (I + \mu\partial f(\cdot/\mu))^{-1}$

If g is the conjugate of $\frac{1}{2}\|\cdot\|^2 + \mu^2 f(\cdot/\mu)$, then $\partial g = (I + \mu\partial f(\cdot/\mu))^{-1}$

$\therefore$ $p(\cdot, \mu)$ is the derivative of the convex conjugate g

Schur complement equation: Solving

$$-\nabla h(x) + \sqrt{c_k} A p(\sqrt{c_k} A^T y - b_k, \mu_{k+1}) + d_k^{-1}(y - y_k) = 0$$

is equivalent to minimizing the convex function

$$-h(x) + g(\sqrt{c_k} A^T y - b_k, \mu_{k+1}) + \frac{1}{2d_k}\|y - y_k\|^2$$

Preliminary Numerical Experiments: SDP

n-logarithmically homogeneous barrier: $f : X \in \mathbb{S}^n_{++} \mapsto -\log\det X$

Smoothing approximation: $p(Z,\mu) = (I + \mu\partial f(\cdot/\mu))^{-1}(Z) = \dfrac{Z + \sqrt{Z^2 + 4\mu^2 I}}{2}$

To solve the Schur complement equation, we use inexact Newton's method with strong-Wolfe line search + PCG

Jacobian of Schur complement equation at y:

$$\Delta_y \mapsto c_k A p_W(W, \mu_{k+1}) A^T \Delta_y + d_k^{-1} \Delta_y$$

with $W = \sqrt{c_k} A^T y - b_k$ and

$$p_W(W,\mu) : \Delta_W \mapsto \sum_{i,j=1}^{n} [\lambda_i, \lambda_j]_{p(\cdot,\mu)} (q_i^T \Delta_W q_j) q_i q_j^T \qquad \left(W = \sum_{i=1}^{n} \lambda_i q_i q_i^T \right)$$

$$[\lambda_i, \lambda_j]_{p(\cdot,\mu)} = \frac{1}{2}\left(1 + \frac{\lambda_i + \lambda_i}{\sqrt{\lambda_i^2 + 4\mu^2} + \sqrt{\lambda_i^2 + 4\mu^2}}\right) \in (0,1) \qquad \text{when } \mu > 0$$

Preliminary Numerical Experiments: SDP

Accelerating computation of $p_W(W,\mu)$:

$$\sum_{i,j=1}^{n} [\lambda_i,\lambda_j]_{p(\cdot,\mu)}(q_i^T \Delta_W q_j) q_i q_j^T = \Delta_W - \sum_{i,j=1}^{n} (1-[\lambda_i,\lambda_j]_{p(\cdot,\mu)})(q_i^T \Delta_W q_j) q_i q_j^T$$

Either

1. drop (i,j)'th term from sum when $[\lambda_i,\lambda_j]_{p(\cdot,\mu)} \approx 0$

   or

2. drop (i,j)'th term from sum when $[\lambda_i,\lambda_j]_{p(\cdot,\mu)} \approx 1$

Same technique used in Newton-CG Augmented Lagrangian Algorithm (SDPNAL, Sun, Toh and Zhao, 2010)

SDPNAL can be viewed as the limiting version of SPPA as $\mu \to 0$
$\therefore$ we compare with SDPNAL

Preliminary Numerical Experiments: SDPLIB Graph partitioning

Stop when primal & dual relative infeasibility $\leq 1e-6$

Name of problem	$n\|m$		Iter M|Sub	Relative infeas. Primal	Dual	Relative gap	Time (m:s)
equalG11	801|801	SPPA	20|69	7.4e-07	7.5e-07	5.6e-06	1:49
		SDPNAL	29|103	2.1e-07	6.5e-07	-1.0e-04	1:39
equalG51	1001|1001	SPPA	19|64	6.7e-07	4.5e-07	2.9e-06	2:40
		SDPNAL	26|149	9.5e-07	7.5e-07	-2.1e-05	3:54

		SPPA			SDPNAL		
Name of problem	$n\|m$	PCG /Sub	Rank /PCG	LS /Sub	PCG /Sub	Rank /PCG	LS /Sub
equalG11	801|801	16.0	133.9	1.1	19.9	63.4	1.0
equalG51	1001|1001	11.6	277.8	1.0	20.1	82.2	1.0

M = #Main iterations   Sub = #Sub-iterations = #Newton steps
LS = #Line searches $(n - \text{Rank})^2 = \#[\lambda_i, \lambda_j]_{p(\cdot,\mu)}$ dropped

Preliminary Numerical Experiments: SDPLIB Max-cut

Name of problem	$n\|m$		Iter M|Sub	Relative infeas. Primal	Dual	Relative gap	Time (h:m:s)
maxG11	800|800	SPPA	19|63	8.8e-07	9.6e-07	2.6e-06	2:39
		SDPNAL	30|105	2.5e-07	8.0e-07	-3.9e-06	1:35
maxG51	1000|1000	SPPA	15|58	1.9e-08	1.8e-07	1.2e-07	3:06
		SDPNAL	23|70	3.1e-07	8.7e-07	-2.6e-06	1:11
maxG32	2000|2000	SPPA	20|66	1.7e-08	6.7e-07	2.2e-06	25:21
		SDPNAL	31|109	2.2e-07	5.3e-07	-3.8e-06	14:57
maxG55	5000|5000	SPPA	15|57	6.2e-07	9.0e-07	1.6e-06	4:19:36
		SDPNAL	24|73	7.6e-07	7.0e-07	-3.2e-06	1:08:6
maxG60	7000|7000	SPPA	16|62	1.5e-08	2.1e-07	3.0e-07	8:24:34
		SDPNAL	25|83	1.4e-07	4.7e-07	-2.3e-06	3:27:21

		SPPA			SDPNAL		
Name of problem	$n\|m$	PCG /Sub	Rank /PCG	LS /Sub	PCG /Sub	Rank /PCG	LS /Sub
maxG11	800|800	39.5	86.9	1.1	25.8	16.3	1.0
maxG51	1000|1000	15.5	402.0	1.2	8.1	45.5	1.0
maxG32	2000|2000	44.8	125.0	1.1	26.9	28.8	1.0
maxG55	5000|5000	20.6	2582.7	1.2	8.1	133.7	1.0
maxG60	7000|7000	19.6	1784.9	1.4	10.6	172.5	1.0

Preliminary Numerical Experiments: SDPLIB
Lovász theta function

Name of problem	$n\|m$		Iter M|Sub	Relative infeas. Primal	Dual	Relative gap	Time (h:m:s)
thetaG11	801|2401	SPPA	25|138	3.7e-08	3.9e-14	1.5e-08	5:59
		SDPNAL	31|159	2.6e-07	2.0e-12	1.1e-07	2:38
thetaG51	1001|6910	SPPA	24|400	5.9e-04	9.6e-06	1.4e-04	3:36:35
		SDPNAL	37|491	3.4e-03	7.8e-07	1.9e-03	35:16

		SPPA			SDPNAL		
Name of problem	$n\|m$	PCG /Sub	Rank /PCG	LS /Sub	PCG /Sub	Rank /PCG	LS /Sub
thetaG11	801|2401	40.4	26.2	1.7	23.8	14.8	1.4
thetaG51	1001|6910	360.2	92.1	1.1	66.6	110.3	1.5

THANK YOU

Pre-print @ http://www.ntu.edu.sg/home/cbchua