Asymmetric Proximal Point Algorithms with Moving Proximal Centers

Deren Han (handeren@njnu.edu.cn)

School of Mathematical Sciences, Nanjing Normal University

Nanjing, 210023, China.

September 2, 2014

Outline

1. Introduction
2. APPA with Moving Proximal Centers
3. Worst-case Rate of Convergence
4. Linear Convergence
5. Two Concrete Algorithms
   - The Saddle Point Problem
   - Multiblock Decomposible Convex Optimization
6. Conclusion

VI

$$(x - x^*)^T F(x^*) \geq 0, \ \forall x \in \Omega, \tag{1}$$

- Ω: a nonempty, closed and convex set in $\mathcal{R}^N$;
- F: a continuous and monotone mapping defined on $\mathcal{R}^N$.

VI

$$(x - x^*)^T F(x^*) \geq 0, \ \forall x \in \Omega, \tag{1}$$

- Ω: a nonempty, closed and convex set in $\mathcal{R}^N$;
- F: a continuous and monotone mapping defined on $\mathcal{R}^N$.

Examples: Equations, Complementarity Problems, Constrained Optimization Problems, Saddle Point Problems, etc.

Proximal Point Algorithms

- Classical PPA:

$$(x - x^{k+1})^T \left(F(x^{k+1}) + \frac{1}{c_k}(x^{k+1} - x^k) \right) \geq 0, \ \forall x \in \Omega, \qquad (2)$$

$\frac{1}{c_k}$ is the proximal parameter and x^k is the proximal center.

Proximal Point Algorithms

- Classical PPA:

$$(x - x^{k+1})^T \left(F(x^{k+1}) + \frac{1}{c_k}(x^{k+1} - x^k) \right) \geq 0, \ \forall x \in \Omega, \qquad (2)$$

$\frac{1}{c_k}$ is the proximal parameter and x^k is the proximal center.

- A General Version:

$$(x - x^{k+1})^T \left(F(x^{k+1}) + M_k(x^{k+1} - x^k) \right) \geq 0, \ \forall x \in \Omega, \qquad (3)$$

where the metric proximal parameter $M_k \in \mathcal{R}^{n \times n}$ is positive definite and symmetric.

Proximal Point Algorithms

- Classical PPA:

$$(x - x^{k+1})^T \left(F(x^{k+1}) + \frac{1}{c_k}(x^{k+1} - x^k) \right) \geq 0, \ \forall x \in \Omega, \qquad (2)$$

$\frac{1}{c_k}$ is the proximal parameter and x^k is the proximal center.

- A General Version:

$$(x - x^{k+1})^T \left(F(x^{k+1}) + M_k(x^{k+1} - x^k) \right) \geq 0, \ \forall x \in \Omega, \qquad (3)$$

where the metric proximal parameter $M_k \in \mathcal{R}^{n \times n}$ is positive definite and symmetric.

$M_k := 1/c_k M$,

$$(x - x^{k+1})^T \left(c_k F(x^{k+1}) + M(x^{k+1} - x^k) \right) \geq 0, \ \forall x \in \Omega. \qquad (4)$$

Role of M in Algorithms:

Make the subproblems easier:

- Preconditioner: T. Pock and A. Chambolle [1].
- Decomposable of the subproblems.

[1] T. Pock and A. Chambolle. *Diagonal preconditioning for first order primal-dual algorithms in convex optimization*, IEEE Inter. Con. Comput. Vis., 2011, pp. 1762-1769.

A simple example

The saddle point problem:

$$\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \Phi(u, v) := f(u) + v^T A u - g(v). \qquad (5)$$

A simple example

The saddle point problem:

$$\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \Phi(u, v) := f(u) + v^T A u - g(v). \qquad (5)$$

PDHG scheme:

$$\begin{cases} \hat{u}^{k+1} & = & u^k - \tau A^T v^k, \\ u^{k+1} & = & \arg\min_{u \in \mathcal{U}} f(u) + \frac{1}{2\tau}\|u - \hat{u}^{k+1}\|^2, \\ \bar{u}^{k+1} & = & u^{k+1} + \theta(u^{k+1} - u^k), \\ \hat{v}^{k+1} & = & v^k + \sigma A \bar{u}^{k+1}, \\ v^{k+1} & = & \arg\min_{v \in \mathcal{V}} g(v) + \frac{1}{2\sigma}\|v - \hat{v}^{k+1}\|^2 \end{cases} \qquad (6)$$

A simple example

The saddle point problem:

$$\min_{u \in \mathcal{U}} \max_{v \in \mathcal{V}} \Phi(u,v) := f(u) + v^T A u - g(v). \tag{5}$$

PDHG scheme:

$$\begin{cases} \hat{u}^{k+1} &= u^k - \tau A^T v^k, \\ u^{k+1} &= \arg\min_{u \in \mathcal{U}} f(u) + \frac{1}{2\tau}\|u - \hat{u}^{k+1}\|^2, \\ \bar{u}^{k+1} &= u^{k+1} + \theta(u^{k+1} - u^k), \\ \hat{v}^{k+1} &= v^k + \sigma A \bar{u}^{k+1}, \\ v^{k+1} &= \arg\min_{v \in \mathcal{V}} g(v) + \frac{1}{2\sigma}\|v - \hat{v}^{k+1}\|^2 \end{cases} \tag{6}$$

PPA point of view

$$M := \begin{pmatrix} \frac{1}{\tau} I_m & -A^T \\ -\theta A & \frac{1}{\sigma} I_n \end{pmatrix}. \tag{7}$$

Difficulty in Analysis:

- Can not bound the progress of consecutive iterations using M-norm;

Difficulty in Analysis:

- Can not bound the progress of consecutive iterations using M-norm;
- Remedies: Correction steps;

Difficulty in Analysis:

- Can not bound the progress of consecutive iterations using M-norm;
- Remedies: Correction steps;
- **Our Motivation:** Schemes without corrections.

Definitions

Let $F(\cdot)$ be a mapping from $\mathcal{R}^N$ into $\mathcal{R}^N$. Then, $F(\cdot)$ is said to be

- Monotone: if

$$(F(x) - F(y))^T (x - y) \geq 0 \qquad \forall x, y \in \mathcal{R}^N;$$

- Strongly monotone with modulus μ if

$$(F(x) - F(y))^T (x - y) \geq \mu \|x - y\|^2 \qquad \forall x, y \in \mathcal{R}^N.$$

- Lipschitz continuous: if

$$\|F(x) - F(y)\| \leq L \|x - y\| \qquad \forall x, y \in \mathcal{R}^N;$$

- Co-coercive with modulus $\sigma > 0$ if

$$(F(x) - F(y))^T (x - y) \geq \sigma \|F(x) - F(y)\|^2 \qquad \forall x, y \in \mathcal{R}^N. \quad (8)$$

Reformulation

$M_s \equiv \frac{1}{2}(M + M^T)$: The symmetric part of M. Define

- $C := M_s^{-1/2}(M - M_s)M_s^{-1/2}$;
- $K \equiv M_s^{1/2}\Omega = \{M_s^{1/2}x : \; x \in \Omega\}$;
- $K^* \equiv M_s^{1/2}\Omega^* = \{M_s^{1/2}x^* : \; x^* \in \Omega^*\}$;
- $\tilde{F}_M(y) \equiv M_s^{-1/2}F(M_s^{-1/2}y), \; \forall y \in K, ;$
- For any $x \in \Omega$, we define $y := M_s^{1/2}x$. Thus,

$$y' \equiv M_s^{1/2}x', \; y^{k+1} \equiv M_s^{1/2}x^{k+1}, \text{ and } y^k \equiv M_s^{1/2}x^k. \tag{9}$$

Then, a vector x^* is a solution of the variational inequality (1) if and only if $y^* := M_s^{1/2}x^*$ is a solution of the variational inequality of finding $y \in K$ such that

$$(y' - y^*)^T\tilde{F}_M(y^*) \geq 0, \quad \forall y' \in K. \tag{10}$$

The solution set of (10) is thus given by K^*.

A Useful Lemma

Lemma 1

Let the mapping $G: \mathcal{R}^n \to \mathcal{R}^n$ *be co-coercive on a nonempty, closed, convex subset* W *in* $\mathcal{R}^n$ *with modulus* $\sigma > 1/2$*, Then, for any* $x, y, z \in W$*, we have*

$$(x-y)^T(G(z)-G(y)) \geq -\frac{\nu}{2}\|x-z\|^2, \tag{11}$$

where ν *is an arbitrary number satisfying*

$$0 < \frac{1}{2\sigma} < \nu < 1. \tag{12}$$

Co-Coerciveness

Lemma 2

[Proposition 12.5.3](Facchinei and Pang 2003) The mapping $\alpha\tilde{F}_M - C$ is co-coercive over K with modulus greater than $1/2$ if either one of the following two conditions holds:

1. *F is Lipschitz continuous on Ω with modulus L and a $\tau \in (0,1)$ exists such that, for all $y_1,\ y_2 \in K$,*

$$\|\alpha\tilde{F}_M(y_2) - \alpha\tilde{F}_M(y_1) - M_s^{-1/2}MM_s^{-1/2}(y_2 - y_1)\| \le \tau\|y_2 - y_1\|;$$

2. *$F(x) = Qx + q$, for some positive semidefinite matrix $Q \equiv D + E$ with D positive definite and E symmetric, and $0 < \|I_n + H\| < 2$,*

The Exact Version

Algorithm 1: APPA-MPC — The Exact Version

Step 0. Given a positive definite matrix $M \in \mathbb{R}^{n\times n}$, $\epsilon > 0$, and $x^0 \in \Omega$. Choose α such that $\alpha\tilde{F}_M - C$ is co-coercive over K with the modulus $\sigma > 1/2$. Set $k := 0$.

Step 1 [APPA Step]. Find the next iterate x^{k+1} by solving the subproblem

$$(x' - x^{k+1})^T[F(x^{k+1}) + M(x^{k+1} - (x^k - \alpha M^{-1}F(x^k)))] \geq 0,\ \forall x' \in \Omega. \tag{3.1}$$

Step 2 [Convergence Verification]. If

$$\|x^{k+1} - x^k\| \leq \epsilon, \tag{3.2}$$

then stop; Otherwise, set $k := k + 1$ and go to Step 1.

The Exact Version

Algorithm 1: APPA-MPC — The Exact Version

Step 0. Given a positive definite matrix $M \in \mathbb{R}^{n\times n}$, $\epsilon > 0$, and $x^0 \in \Omega$. Choose α such that $\alpha\tilde{F}_M - C$ is co-coercive over K with the modulus $\sigma > 1/2$. Set $k := 0$.

Step 1 [APPA Step]. Find the next iterate x^{k+1} by solving the subproblem

$$(x' - x^{k+1})^T[F(x^{k+1}) + M(x^{k+1} - (x^k - \alpha M^{-1}F(x^k)))] \geq 0, \ \forall x' \in \Omega. \tag{3.1}$$

Step 2 [Convergence Verification]. If

$$\|x^{k+1} - x^k\| \leq \epsilon, \tag{3.2}$$

then stop; Otherwise, set $k := k+1$ and go to Step 1.

Remark: The exact version of APPA-MPC (“APPA-MPC-E" in short): The proximal center in x^k in classical PPA is shifted to $x^k - \alpha M^{-1}F(x^k)$ in (3.1).

Convergence

Lemma 3

The subproblem (3.1) of the APPA-MPC-E can be rewritten as

$$(y' - y^{k+1})^T[\tilde{F}_M(y^{k+1}) + (I + C)(y^{k+1} - y^k) + \alpha \tilde{F}_M(y^k)] \geq 0, \ \forall y' \in K. \quad (13)$$

Convergence

Lemma 3

The subproblem (3.1) of the APPA-MPC-E can be rewritten as

$$(y' - y^{k+1})^T[\tilde{F}_M(y^{k+1}) + (I + C)(y^{k+1} - y^k) + \alpha\tilde{F}_M(y^k)] \geq 0, \ \forall y' \in K. \quad (13)$$

Lemma 4

Let y^* be an arbitrary solution point in K^*.. Then, we have

$$\|y^{k+1} - y^*\|^2 \leq \|y^k - y^*\|^2 - (1 - \nu)\|y^k - y^{k+1}\|^2, \quad (14)$$

where ν is an arbitrary number satisfying (12).

The Inexact Version

Algorithm 2: APPA-MPC — An Inexact Version

Step 0. Given a positive definite matrix $M \in \mathbb{R}^{n\times n}$, $\epsilon > 0$ and $x^0 \in \Omega$. Choose α such that $\alpha\tilde{F}_M - C$ is co-coercive over K with the modulus $\sigma > 1/2$. Choose a sequence of $\{\epsilon_k\}$ with $\epsilon_k > 0$ for all k and $\sum_{k=1}^{\infty} \epsilon_k < +\infty$. Set $k := 0$.

Step 1 [PPA Step]. Solve the subproblem (3.1) iteratively and find x^{k+1} such that

$$\max\{\|x^{k+1} - \bar{x}^{k+1}\|, \|F(x^{k+1}) - F(\bar{x}^{k+1})\|\} \le \epsilon_k, \tag{4.3}$$

where

$$\bar{x}^{k+1} := P_\Omega[x^{k+1} - (F(x^{k+1}) + M(x^{k+1} - x^k) + \alpha F(x^k))]. \tag{4.4}$$

Step 2 [Convergence Verification]. If

$$\max\{\|x^{k+1} - x^k\|, \|x^{k+1} - \bar{x}^{k+1}\|\} \le \epsilon, \tag{4.5}$$

then stop; Otherwise, set $k := k + 1$ and go to Step 1.

Convergence

Lemma 5

Let $\{x^k\}$ *be the sequence generated by APPA-MPC-I. Then, there exist a positive scalar* Γ *such that the following inequality holds for any* $y' \in K$*:*

$$(y'-\bar{y}^{k+1})^T(\tilde{F}_M(\bar{y}^{k+1})+(I+C)(\bar{y}^{k+1}-y^k)+\alpha\tilde{F}_M(y^k)) \geq -\epsilon_k\|y'-\bar{y}^{k+1}\|^2-\Gamma\epsilon_k. \quad (15)$$

Convergence

Lemma 5

Let $\{x^k\}$ be the sequence generated by APPA-MPC-I. Then, there exist a positive scalar Γ such that the following inequality holds for any $y' \in K$:

$$(y' - \bar{y}^{k+1})^T (\tilde{F}_M(\bar{y}^{k+1}) + (I + C)(\bar{y}^{k+1} - y^k) + \alpha \tilde{F}_M(y^k)) \geq -\epsilon_k \|y' - \bar{y}^{k+1}\|^2 - \Gamma \epsilon_k. \quad (15)$$

Theorem 6

The sequence $\{x^k\}$ generated by the APPA-MPC-I globally converges to a solution point of the variational inequality (1).

APPA-MPC-E

Theorem 7

Let $\{x^k\}$ *be the sequence generated by the APPA-MPC-E. For an integer* $t > 0$*, let*

$$y_t := \frac{1}{t+1} \sum_{k=0}^{t} y^{k+1}, \tag{16}$$

then we have $y_t \in K$ *and*

$$(y_t - y')^T \tilde{F}_M(y') \le \frac{1}{2(1+\alpha)(t+1)} \|y' - y^0\|^2, \quad \forall y' \in K. \tag{17}$$

APPA-MPC-E

Theorem 7

Let $\{x^k\}$ *be the sequence generated by the APPA-MPC-E. For an integer* $t > 0$, *let*

$$y_t := \frac{1}{t+1} \sum_{k=0}^{t} y^{k+1}, \tag{16}$$

then we have $y_t \in K$ *and*

$$(y_t - y')^T \tilde{F}_M(y') \le \frac{1}{2(1+\alpha)(t+1)} \|y' - y^0\|^2, \quad \forall y' \in K. \tag{17}$$

Lemma 8 (Theorem 2.3.5, Facchinei and Pang (2003))

The solution set K^* *of the variational inequality is convex and it can be characterized as*

$$K^* = \bigcap_{y' \in K} \{y^* \in K \mid (y' - y^*)^T \tilde{F}_M(y') \ge 0.\} \tag{18}$$

APPA-MPC-I

Theorem 9

Let $\{x^k\}$ be the sequence generated by the APPA-MPC-I. For an integer $t > 0$, let

$$y_t := \frac{1}{t+1} \sum_{k=0}^{t} \bar{y}^{k+1}, \tag{19}$$

then we have $y_t \in K$ and

$$(y_t - y')^T \tilde{F}_M(y') \le \frac{1}{2(1+\alpha)(t+1)} \left\{ \|y' - y^0\|^2 + N(y') \sum_{k=0}^{t} \epsilon_k \right\}, \quad \forall y' \in K.$$

Theorem 10

Suppose that F is strongly monotone with the modulus $\mu_f > 0$ and Lipschitz continuous with the constant L_f. Let $\{x^k\}$ be the sequence generated by the proposed APPA-MPC-E. Then it converges to a solution point in Ω^* linearly.

Theorem 10

Suppose that F is strongly monotone with the modulus $\mu_f > 0$ and Lipschitz continuous with the constant L_f. Let $\{x^k\}$ be the sequence generated by the proposed APPA-MPC-E. Then it converges to a solution point in Ω^* linearly.

Theorem 11

Suppose that F is strongly monotone with the modulus $\mu_f > 0$ and Lipschitz continuous with the constant L_f. Moreover, assume that

$$\epsilon_k \leq \frac{3(1-\nu)}{8\Gamma + 2}\|y^k - \bar{y}^{k+1}\|^2, \quad \forall k \geq 0.$$

Let $\{x^k\}$ be the sequence generated by the proposed APPA-MPC-I. Then it converges to a solution point in Ω^* linearly.

目录

1 Introduction

2 APPA with Moving Proximal Centers

3 Worst-case Rate of Convergence

4 Linear Convergence

5 Two Concrete Algorithms

- The Saddle Point Problem
- Multiblock Decomposible Convex Optimization

6 Conclusion

The saddle-point problem

Recall that the problem is

$$\min_{u\in\mathcal{U}} \max_{v\in\mathcal{V}} \Phi(u,v) := f(u) + v^T Au - g(v), \tag{20}$$

where $f:\ \mathcal{U}\to\mathcal{R}$, $g:\ \mathcal{V}\to\mathcal{R}$ are convex but not necessarily smooth functions; $\mathcal{U}\subseteq\mathcal{R}^n$ and $\mathcal{V}\subseteq\mathcal{R}^m$ are two nonempty, closed and convex sets; and $A\in\mathcal{R}^{m\times n}$.

The algorithm

Algorithm 3: A Primal-Dual Hybrid Gradient Algorithm for the Saddle Point Problem (1.7)

Input: Choose $u^0 \in \mathbb{R}^n$, $v^0 \in \mathbb{R}^m$, $s_u^0 \in \partial f(u^0) + N_{\mathcal{U}}(u^0)$, $s_v^0 \in \partial g(v^0) + N_{\mathcal{V}}(v^0)$, $\theta \in \mathbb{R}$, and $\tau, \sigma > 0$ such that the matrix M defined by (1.9) is positive definite. Choose α such that $\alpha \tilde{F}_M - C$ is co-coercive over K with the modulus $\sigma > 1/2$.

while *Not converged* **do**

$\quad \breve{u}^{k+1} = u^k - \tau A^T v^k - \tau\alpha(s_u^k + A^T v^k)$

$\quad u^{k+1} = \arg\min_{u \in \mathcal{U}} f(u) + \frac{1}{2\tau}\|u - \breve{u}^{k+1}\|^2$

$\quad s_u^{k+1} = \frac{1}{\tau}(\breve{u}^{k+1} - u^{k+1})$

$\quad \bar{u}^{k+1} = u^{k+1} + \theta(u^{k+1} - u^k)$

$\quad \breve{v}^{k+1} = v^k + \sigma A\bar{u}^{k+1} - \sigma\alpha(s_v^k - Au^k)$

$\quad v^{k+1} = \arg\min_{v \in \mathcal{V}} g(v) + \frac{1}{2\sigma}\|v - \breve{v}^{k+1}\|^2$

$\quad s_v^{k+1} = \frac{1}{\sigma}(\breve{v}^{k+1} - v^{k+1})$

end

目录

1 Introduction

2 APPA with Moving Proximal Centers

3 Worst-case Rate of Convergence

4 Linear Convergence

5 Two Concrete Algorithms
- The Saddle Point Problem
- Multiblock Decomposible Convex Optimization

6 Conclusion

The Problems

$$\min \left\{ \sum_{i=1}^{m} \theta_i(x_i) \;\middle|\; \sum_{i=1}^{m} A_i x_i = b,\ x_i \in \mathcal{X}_i;\ i = 1, \cdots, m \right\}, \tag{21}$$

where $\theta_i : \mathcal{R}^{n_i} \to \mathcal{R}$ are closed proper convex functions (not necessarily smooth); $A_i \in \mathcal{R}^{l \times n_i}$; $\mathcal{X}_i \subset \mathcal{R}^{n_i}$ are closed and convex nonempty sets; $b \in \mathcal{R}^l$; and $\sum_{i=1}^m n_i = n$.

The Decomposition Algorithm

$$M = \begin{pmatrix} \kappa I & -\beta A_1^T A_2 & \cdots & -\beta A_1^T A_m & 0 \\ 0 & \kappa I & \cdots & -\beta A_2^T A_m & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \kappa I & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\beta} I \end{pmatrix}. \tag{22}$$

The algorithm

Algorithm 4: A Splitting Algorithm for (7.6)

Input: Choose $x_i^0 \in \mathbb{R}^{n_i}$, $s_i^0 \in \partial\theta_i(x_i^0)$, $i = 1, \cdots, m$, and $\beta > 0$. Choose $\kappa > 0$ such that the matrix M defined by (7.12) is positive definite. Choose α such that $\alpha\tilde{F}_M - C$ is co-coercive over K with the modulus $\sigma > 1/2$.

while *Not converged* **do**

  for $i = 1, \cdots, m$ **do**

    $\breve{x}_i^{k+1} := x_i^k - \frac{1}{\kappa}\alpha(s_i^k - A_i^T\lambda^k)$

    $\phi_i^k := \sum_{j=1}^{i-1} A_j x_j^{k+1} + \sum_{j=i+1}^{m} A_j x_j^k - b - \frac{1}{\beta}\lambda^k$

    $x^{k+1} = \arg\min_{x_i \in \mathcal{X}_i} \theta_i(x_i) + \frac{\beta}{2}\|A_i x_i + \phi_i^k\|^2 + \frac{\kappa}{2}\|x_i - \breve{x}_i^{k+1}\|^2$

    $\lambda^{k+1} = \lambda^k - \beta\left(\sum_{i=1}^m A_i x_i^{k+1} - b\right)$

    $s_i^{k+1} = \kappa(\breve{x}_i^{k+1} - x_i^{k+1}) - \beta A_i^T\left(A_i x_i^{k+1} + \phi_i^k\right)$

  end

end

- We have proposed an APPA-MPC(Exact& Inexact);

- We have proposed an APPA-MPC(Exact& Inexact);
- Proved the global convergence ;

- We have proposed an APPA-MPC(Exact& Inexact);
- Proved the global convergence ;
- Worst-case convergence rate & Asymptotical linear rate of convergence.

- We have proposed an APPA-MPC(Exact& Inexact);
- Proved the global convergence ;
- Worst-case convergence rate & Asymptotical linear rate of convergence.
- Two concrete algorithms;

- We have proposed an APPA-MPC(Exact& Inexact);
- Proved the global convergence ;
- Worst-case convergence rate & Asymptotical linear rate of convergence.
- Two concrete algorithms;
- Numerical performances?

- We have proposed an APPA-MPC(Exact& Inexact);
- Proved the global convergence ;
- Worst-case convergence rate & Asymptotical linear rate of convergence.
- Two concrete algorithms;
- Numerical performances?

Thank you !