Recent Developments of Alternating Direction Method of Multipliers with Multi-Block Variables

Shiqian Ma

Department of Systems Engineering and Engineering Management

The Chinese University of Hong Kong

2014 Workshop on Optimization for Modern Computation
BICMR, Beijing, China
September 2, 2014

Outline

- ADMM for $N = 2$
- Existing work on ADMM for $N \geq 3$
- Convergence Rates of ADMM for $N \geq 3$
- BSUM-M

Alternating Direction Method of Multipliers (ADMM)

- Convex optimization

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) + \ldots + f_N(x_N) \\ \text{s.t.} & A_1 x_1 + A_2 x_2 + \ldots + A_N x_N = b \\ & x_j \in \mathcal{X}_j, \quad j = 1, 2, \ldots, N. \end{array}$$

- f_j: closed convex function
- $\mathcal{X}_j$: closed convex set
- Augmented Lagrangian function

$$\mathcal{L}_\gamma(x_1, \ldots, x_N; \lambda) := \sum_{j=1}^N f_j(x_j) - \langle \lambda, \sum_{j=1}^N A_j x_j - b \rangle + \frac{\gamma}{2} \| \sum_{j=1}^N A_j x_j - b \|_2^2$$

Multi-Block ADMM

- Augmented Lagrangian function

$$\mathcal{L}_\gamma(x_1,\ldots,x_N;\lambda) := \sum_{j=1}^N f_j(x_j) - \langle \lambda, \sum_{j=1}^N A_j x_j - b\rangle + \frac{\gamma}{2}\|\sum_{j=1}^N A_j x_j - b\|_2^2$$

- Multi-Block ADMM

$$\begin{cases} x_1^{k+1} & := & \mathrm{argmin}_{x_1\in\mathcal{X}_1}\ \mathcal{L}_\gamma(x_1, x_2^k,\ldots,x_N^k;\lambda^k) \\ x_2^{k+1} & := & \mathrm{argmin}_{x_2\in\mathcal{X}_2}\ \mathcal{L}_\gamma(x_1^{k+1}, x_2, x_3^k,\ldots,x_N^k;\lambda^k) \\ & \vdots & \\ x_N^{k+1} & := & \mathrm{argmin}_{x_N\in\mathcal{X}_N}\ \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1},\ldots,x_{N-1}^{k+1}, x_N;\lambda^k) \\ \lambda^{k+1} & := & \lambda^k - \gamma\left(\sum_{j=1}^N A_j x_j^{k+1} - b\right). \end{cases}$$

- Update the primal variables in a Gauss-Seidel manner.

ADMM for $N = 2$

- ADMM for $N = 2$

$$\begin{cases} x_1^{k+1} & := & \text{argmin}_{x_1 \in \mathcal{X}_1} \; \mathcal{L}_\gamma(x_1, x_2^k; \lambda^k) \\ x_2^{k+1} & := & \text{argmin}_{x_2 \in \mathcal{X}_2} \; \mathcal{L}_\gamma(x_1^{k+1}, x_2; \lambda^k) \\ \lambda^{k+1} & := & \lambda^k - \gamma \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b \right). \end{cases}$$

- Long history goes back to variational methods for PDEs in 1950s; Relate to Douglas-Rachford and Peaceman-Rachford Operator Splitting Methods for finding zero of monotone operators.

$$\text{Find } x, \; \text{s.t.}, 0 \in A(x) + B(x).$$

- Revisited recently for sparse optimization
  - [Wang-Yang-Yin-Zhang-2008]
  - [Goldstein-Osher-2009]
  - [Boyd-etal-2011]

Global Convergence of ADMM for $N = 2$

- ADMM for $N = 2$

$$\begin{cases} x_1^{k+1} & := & \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \ \mathcal{L}_\gamma(x_1, x_2^k; \lambda^k) \\ x_2^{k+1} & := & \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2; \lambda^k) \\ \lambda^{k+1} & := & \lambda^k - \gamma \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b \right). \end{cases}$$

- Global convergence for any $\gamma > 0$. (Fortin-Glowinski-1983; Gabay-1983; Glowinski-Le Tallec-1989; Eckstein-Bertsekas-1992)
- ADMM for $N = 2$ with fixed dual step size

$$\begin{cases} x_1^{k+1} & := & \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \ \mathcal{L}_\gamma(x_1, x_2^k; \lambda^k) \\ x_2^{k+1} & := & \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \ \mathcal{L}_\gamma(x_1^{k+1}, x_2; \lambda^k) \\ \lambda^{k+1} & := & \lambda^k - \alpha\gamma \left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b \right). \end{cases}$$

- $\alpha > 0$ is a fixed dual step size
- Global convergence for any $\gamma > 0$ and $\alpha \in (0, \frac{1+\sqrt{5}}{2})$.

Sublinear Convergence of ADMM for $N = 2$

- Ergodic $O(1/k)$ convergence (He-Yuan-2012)
- Non-Ergodic $O(1/k)$ convergence (He-Yuan-2012)
- Ergodic $O(1/k)$ convergence (Monteiro-Svaiter-2013)

Linear Convergence Rate of ADMM for $N = 2$

- Douglas-Rachford splitting method converges linearly if B is coercive and Lipschitz (Lions-Mercier-1979)
- Linear convergence for solving linear programs (Eckstein-Bertsekas-1990)
- Linear convergence for quadratic programs (Han-Yuan-2013; Boley-2013)

Generalized ADMM

- Generalized ADMM for $N = 2$ (Deng-Yin-2012)

$$\begin{cases} x_1^{k+1} & := & \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \mathcal{L}_\gamma(x_1, x_2^k; \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_P^2 \\ x_2^{k+1} & := & \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \mathcal{L}_\gamma(x_1^{k+1}, x_2; \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_Q^2 \\ \lambda^{k+1} & := & \lambda^k - \alpha\gamma\left(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b\right). \end{cases}$$

- One sufficient condition for guaranteeing global linear convergence: $P = Q = 0$, $\alpha = 1$, f_2 strongly convex, ∇f_2 Lipschtiz continuous, A_2 full row rank.

ADMM for $N \geq 3$: a counter example

- A negative result (Chen-He-Ye-Yuan-2013): Direct extension of multi-block ADMM is not necessarily convergent
- A counter example:

$$A_1x_1+A_2x_2+A_3x_3 = 0, \text{ where } A = (A_1, A_2, A_3) = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix}$$

The update of multi-block ADMM with $\gamma = 1$ is

$$\begin{pmatrix} 3 & 0 & 0 & 0 & 0 & 0 \\ 4 & 6 & 0 & 0 & 0 & 0 \\ 5 & 7 & 9 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 2 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} \end{pmatrix} = \begin{pmatrix} 0 & -4 & -5 & 1 & 1 & 1 \\ 0 & 0 & -7 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 2 & 2 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \lambda^k \end{pmatrix}$$

ADMM for $N \geq 3$: a counter example

- Equivalently,

$$\begin{pmatrix} x_2^{k+1} \\ x_3^{k+1} \\ \lambda^{k+1} \end{pmatrix} = M \begin{pmatrix} x_2^{k} \\ x_3^{k} \\ \lambda^{k} \end{pmatrix}, \quad \text{where}$$

$$M = \frac{1}{162} \begin{pmatrix} 144 & -9 & -9 & -9 & 18 \\ 8 & 157 & -5 & 13 & -8 \\ 64 & 122 & 122 & -58 & -64 \\ 56 & -35 & -35 & 91 & 56 \\ -88 & -26 & -26 & -62 & 88 \end{pmatrix}.$$

- Note that $\rho(M) > 1$.

Theorem (Chen-He-Ye-Yuan-2013)

There existing an example where the direct extension of ADMM of three blocks with a real number initial point is not necessarily convergent for any choice of $\gamma > 0$.

ADMM for $N \geq 3$: Strong convexity?

$$\begin{array}{ll} \min & 0.05x_1^2 + 0.05x_2^2 + 0.05x_3^2 \\ \text{s.t.} & \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0. \end{array}$$

- For $\gamma = 1$, $\rho(M) = 1.0087 > 1$
- Able to find a proper initial point such that ADMM diverges
- Even for strongly convex programming, the extended ADMM is not necessarily convergent for a certain $\gamma > 0$.

ADMM for $N \geq 3$: Strong convexity works!

Global convergence

Theorem (Han-Yuan-2012)

If $f_i, i = 1, \ldots, N$ are strongly convex with parameter σ_i's, and

$$0 < \gamma < \min_{i=1,\ldots,N} \left\{ \frac{2\sigma_i}{3(N-1)\lambda_{\max}(A_i^\top A_i)} \right\},$$

then multi-block ADMM globally converges.

Convergence Rate?

ADMM for $N \geq 3$: weaker condition and convergence rate

$$u := (x_1, \ldots, x_N),\ \bar{x}_i^t = \frac{1}{t+1}\sum_{k=0}^{t} x_i^{k+1}, 1 \leq i \leq N,\ \bar{\lambda}^t = \frac{1}{t+1}\sum_{k=0}^{t} \lambda^{k+1}.$$

Theorem (Lin-Ma-Zhang-2014a)

If $f_2, \ldots, f_N$ are strongly convex, f_1 is convex, and

$$\gamma \leq \min_{2 \leq i \leq N-1} \left\{ \frac{2\sigma_i}{(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}, \frac{2\sigma_N}{(N-2)(N+1)\lambda_{\max}(A_N^\top A_N)} \right\},$$

then $|f(\bar{u}^t) - f(u^*)| = O(1/t)$, and $\left\| \sum_{i=1}^{N} A_i \bar{x}_i^t - b \right\| = O(1/t)$.

- Weaker condition
- Ergodic $O(1/t)$ convergence rate in terms of objective value and primal feasibility

ADMM for $N \geq 3$: non-ergodic convergence rate

- Optimality measure: if

$$\begin{cases} A_2 x_2^{k+1} - A_2 x_2^k = 0, \\ A_3 x_3^{k+1} - A_3 x_3^k = 0, \\ A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b = 0, \end{cases}$$

  then $(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \lambda^{k+1})$ is optimal.

- Define

$$\begin{aligned} R_{k+1} := \ & \|A_1 x_1^{k+1} + A_2 x_2^{k+1} + A_3 x_3^{k+1} - b\|^2 \\ & + 2\|A_2 x_2^{k+1} - A_2 x_2^k\|^2 + 3\|A_3 x_3^{k+1} - A_3 x_3^k\|^2. \end{aligned}$$

- We can prove: $R_k = o(1/k)$

ADMM for $N \geq 3$: non-ergodic convergence rate

Theorem (Lin-Ma-Zhang-2014a)

If f_2 and f_3 are strongly convex, and

$$\gamma \leq \min\left\{\frac{\sigma_2}{2\lambda_{\max}(A_2^\top A_2)}, \frac{\sigma_3}{2\lambda_{\max}(A_3^\top A_3)}\right\},$$

then

$$\sum_{k=1}^{\infty} R_k < +\infty \quad \text{and} \quad R_k = o(1/k).$$

ADMM for $N \geq 3$: non-ergodic convergence rate

Theorem (Lin-Ma-Zhang-2014a)

If $f_2, \ldots, f_N$ are strongly convex, and

$$\gamma \leq \min_{2 \leq i \leq N-1} \left\{ \frac{2\sigma_i}{(2N-i)(i-1)\lambda_{\max}(A_i^\top A_i)}, \frac{2\sigma_N}{(N-2)(N+1)\lambda_{\max}(A_N^\top A_N)} \right\},$$

then

$$\sum_{k=1}^{\infty} R_k < +\infty \quad \text{and} \quad R_k = o(1/k),$$

where

$$R_{k+1} := \left\| \sum_{i=1}^{N} A_i x_i^{k+1} - b \right\|^2 + \sum_{i=2}^{N} \frac{(2N-i)(i-1)}{2} \|A_i x_i^k - A_i x_i^{k+1}\|^2.$$

ADMM for $N \geq 3$: global linear convergence

- Globally linear convergence of ADMM for $N \geq 3$ (Lin-Ma-Zhang-2014b)

	s.c.	Lipschitz	full row rank	full column rank
1	$f_2, \cdots, f_N$	∇f_N	A_N	—
2	$f_1, \cdots, f_N$	$\nabla f_1, \cdots, \nabla f_N$	—	—
3	$f_2, \cdots, f_N$	$\nabla f_1, \cdots, \nabla f_N$	—	A_1

Table: Three scenarios leading to global linear convergence

- Reduce to the conditions in (Deng-Yin-2012) when $N = 2$

Variants: Modified Multi-Block ADMM

- Proximal Jacobian ADMM (Deng-Lai-Peng-Yin-2014)

$$\left\{\begin{array}{rcl} x_1^{k+1} & := & \mathrm{argmin}_{x_1\in\mathcal{X}_1}\ \mathcal{L}_\gamma(x_1, x_2^k, \ldots, x_N^k; \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{P_1}^2 \\ x_2^{k+1} & := & \mathrm{argmin}_{x_2\in\mathcal{X}_2}\ \mathcal{L}_\gamma(x_1^k, x_2, x_3^k, \ldots, x_N^k; \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_{P_2}^2 \\ & \vdots & \\ x_N^{k+1} & := & \mathrm{argmin}_{x_N\in\mathcal{X}_N}\ \mathcal{L}_\gamma(x_1^k, x_2^k, \ldots, x_{N-1}^k, x_N; \lambda^k) + \frac{1}{2}\|x_N - x_N^k\|_{P_N}^2 \\ \lambda^{k+1} & := & \lambda^k - \alpha\gamma\left(\sum_{j=1}^N A_j x_j^{k+1} - b\right). \end{array}\right.$$

- Conditions for convergence:

$$\left\{\begin{array}{l} P_i \succ \gamma(1/\epsilon_i - 1)A_i^\top A_i, i = 1, 2, \ldots, N \\ \sum_{i=1}^N \epsilon_i < 2 - \alpha \end{array}\right.$$

- $o(1/k)$ convergence rate in non-ergodic sense

Variants: Modified Multi-Block ADMM

- Proximal Gauss-Seidel ADMM

$$\begin{cases} x_1^{k+1} & := & \mathrm{argmin}_{x_1 \in \mathcal{X}_1} \mathcal{L}_\gamma(x_1, x_2^k, x_3^k; \lambda^k) + \frac{1}{2}\|x_1 - x_1^k\|_{P_1}^2 \\ x_2^{k+1} & := & \mathrm{argmin}_{x_2 \in \mathcal{X}_2} \mathcal{L}_\gamma(x_1^{k+1}, x_2, x_3^k; \lambda^k) + \frac{1}{2}\|x_2 - x_2^k\|_{P_2}^2 \\ x_3^{k+1} & := & \mathrm{argmin}_{x_3 \in \mathcal{X}_3} \mathcal{L}_\gamma(x_1^{k+1}, x_2^{k+1}, x_3; \lambda^k) + \frac{1}{2}\|x_3 - x_3^k\|_{P_3}^2 \\ \lambda^{k+1} & := & \lambda^k - \alpha\gamma\left(\sum_{j=1}^3 A_j x_j^{k+1} - b\right). \end{cases}$$

Proximal Gauss-Seidel ADMM

Theorem (Lin-Ma-Zhang-2014c)

Ergodic $O(1/k)$ convergence rate in terms of both objective value and primal feasibility, under conditions: f_3 is strongly convex, and

$$\begin{cases} P_1 & \succ & \gamma(3+\frac{5}{\epsilon_1})A_1^\top A_1 \\ P_2 & \succ & \gamma(1+\frac{3}{\epsilon_2})A_2^\top A_2 \\ P_3 & \succ & \gamma(\frac{1}{\epsilon_3}-1)A_3^\top A_3 \\ 3(\epsilon_1+\epsilon_2+\epsilon_3) & < & 1-\alpha \\ \gamma & < & \frac{\sigma_3\epsilon_3}{2(\epsilon_3+1)\lambda_{\max}(A_3^\top A_3)} \end{cases}$$

Ongoing work. More coming soon.

The General Problem Formulation

- We consider the following convex optimization problem

$$
\begin{array}{ll}
\min & f(x) := g(x_1, \cdots, x_K) + \sum_{k=1}^K h_k(x_k) \\
\text{subject to} & E_1 x_1 + E_2 x_2 + \cdots + E_K x_K = q, \\
 & x_k \in X_k, \quad k = 1, 2, ..., K,
\end{array}
\tag{P}
$$

- $g(\cdot)$ a smooth convex function; h_k a nonsmooth convex function
- $x := (x_1^T, ..., x_K^T)^T \in \Re^n$ block variables
- $X := \prod_{k=1}^K X_k$ feasible set
- $E := (E_1, \cdots, E_K)$ and $h(x) := \sum_{k=1}^K h_k(x_k)$
- Augmented Lagrangian function

$$
L(x; y) = g(x) + h(x) + \langle y, q - Ex \rangle + \frac{\rho}{2}\|q - Ex\|^2
$$

Small dual step size and error bound condition

- ADMM for $N \geq 3$ without strong convexity assumption (Hong-Luo-2012)

$$\begin{cases} x_1^{k+1} & := & \operatorname{argmin}_{x_1 \in \mathcal{X}_1} \; L(x_1, x_2^k, \ldots, x_N^k; y^k) \\ x_2^{k+1} & := & \operatorname{argmin}_{x_2 \in \mathcal{X}_2} \; L(x_1^{k+1}, x_2, x_3^k, \ldots, x_N^k; y^k) \\ & \vdots & \\ x_N^{k+1} & := & \operatorname{argmin}_{x_N \in \mathcal{X}_N} \; L(x_1^{k+1}, x_2^{k+1}, \ldots, x_{N-1}^{k+1}, x_N; y^k) \\ y^{k+1} & := & y^k - \alpha \left(\sum_{j=1}^N A_j x_j^{k+1} - b \right). \end{cases}$$

- Do not assume strong convexity, but need other conditions.

Small dual step size and error bound condition

- Error bound condition: there exist positive scalars τ and δ such that the following error bound holds

$$dist(x, X(y)) \leq \tau \|\tilde{\nabla}_x L(x; y)\|,$$

for all (x, y) such that $\|\tilde{\nabla}_x L(x; y)\| \leq \delta$

- [Hong-Luo-2012]: Given that α is small enough (upper bounded by some constant related to τ and δ), the small step size variant of ADMM converges linearly.
- α is too small and not computable, thus not practical.

- A Block Successive Upper-bound Minimization Method of Multipliers
- Introduce the augmented Lagrangian function for problem (P)

$$L(x; y) = g(x) + h(x) + \langle y, q - Ex \rangle + \frac{\rho}{2}\|q - Ex\|^2$$

- y dual variable; $\rho \geq 0$ primal stepsize
- **Main idea: Primal update**
  1. Update the primal variables successively (Gauss-Seidel)
  2. Optimize some approximate version of $L(x, y)$
- **Main idea: Dual update**
  1. Inexact dual ascent + proper step size control

The BSUM-M Algorithm: Details

- At iteration $r+1$, a block variable x_k is updated by solving

$$\min_{x_k \in X_k} \quad u_k\left(x_k; x_1^{r+1}, \cdots, x_{k-1}^{r+1}, x_k^r, \cdots, x_K^r\right) + \langle y^{r+1}, q - E_k x_k \rangle + h_k(x_k)$$

- $u_k(\cdot\,;\, x_1^{r+1}, \cdots, x_{k-1}^{r+1}, x_k^r, \cdots, x_K^r)$: is an *upper-bound* of

$$g(x) + \frac{\rho}{2}\|q - Ex\|^2$$

at the current iterate $(x_1^{r+1}, \cdots, x_{k-1}^{r+1}, x_k^r, \cdots, x_K^r)$

The BSUM-M Algorithm: Details (cont.)

The BSUM-M Algorithm

At each iteration $r \geq 1$:

$$\begin{cases} y^{r+1} = y^r + \alpha^r(q - Ex^r) = y^r + \alpha^r \left(q - \sum_{k=1}^{K} E_k x_k^r \right), \\ x_k^{r+1} = \arg\min_{x_k \in X_k} u_k(x_k; w_k^{r+1}) - \langle y^{r+1}, E_k x_k \rangle + h_k(x_k), \ \forall\ k \end{cases}$$

where $\alpha^r > 0$ is the dual stepsize.

- To simplify notations, we have defined

$$w_k^{r+1} := (x_1^{r+1}, \cdots, x_{k-1}^{r+1}, x_k^r, x_{k+1}^r, \cdots, x_K^r),$$

The BSUM-M Algorithm: Randomized Version

- Select a vector $\{p_k > 0\}_{k=0}^{K}$ such that $\sum_{k=0}^{K} p_k = 1$
- Each iteration "t" only updates a single randomly selected primal or dual variable

At iteration $t \geq 1$, pick $k \in \{0, \cdots, K\}$ with probability p_k and

If $k = 0$

$$y^{t+1} = y^t + \alpha^t(q - Ex^t),$$

$$x_k^{t+1} = x_k^t, \;\; k = 1, \cdots, K.$$

Else If $k \in \{1, \cdots, K\}$

$$x_k^{t+1} = \text{argmin}_{x_k \in X_k} \; u_k(x_k; x^t) - \langle y^r, E_k x_k \rangle + h_k(x_k),$$

$$x_j^{t+1} = x_j^t, \; \forall \, j \neq k, \quad y^{t+1} = y^t.$$

End

Convergence Analysis: Assumptions

- **Assumption A** (on the problem)

(a) Problem (P) is convex problem

(b) $g(x) = \ell(Ax) + \langle x, b \rangle$; $\ell(\cdot)$ smooth strictly convex, A not necessarily full column rank

(c) Nonsmooth function h_k:

$$h_k(x_k) = \lambda_k \|x_k\|_1 + \sum_J w_J \|x_{k,J}\|_2,$$

where $x_k = (\cdots, x_{k,J}, \cdots)$ is a partition of x_k; $\lambda_k \geq 0$ and $w_J \geq 0$ are some constants.

(d) The feasible sets $\{X_k\}$ are compact polyhedral sets, and are given by $X_k := \{x_k \mid C_k x_k \leq c_k\}$.

Convergence Analysis: Assumptions

- **Assumption B** (on u_k)

(a) $u_k(v_k; x) \geq g(v_k, x_{-k}) + \frac{\rho}{2}\|E_k v_k - q + E_{-k}x_{-k}\|^2, \quad \forall\, v_k \in X_k, \ \forall\, x, k$ (upper-bound)

(b) $u_k(x_k; x) = g(x) + \frac{\rho}{2}\|Ex - q\|^2, \ \forall\, x, k$ (locally tight)

(c) $\nabla u_k(x_k; x) = \nabla_k \left(g(x) + \frac{\rho}{2}\|Ex - q\|^2\right), \ \forall\, x, k$

(d) For any given x, $u_k(v_k; x)$ is strongly convex in v_k

(e) For given x, $u_k(v_k; x)$ has Lipchitz continuous gradient

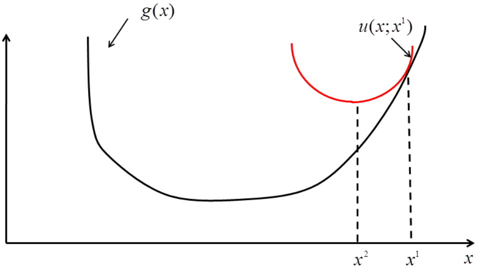

Figure: Illustration of the upper-bound.

The Convergence Result

Theorem (Hong, Chang, Wang, Razaviyanyn, Ma and Luo 2014)

Suppose Assumptions A-B hold. Assume the dual stepsize α^r satisfies

$$\sum_{r=1}^{\infty} \alpha^r = \infty, \quad \lim_{r \to \infty} \alpha^r = 0.$$

Then we have the following:

1. *For the BSUM-M, we have* $\lim_{r\to\infty} \|Ex^r - q\| = 0$*, and every limit point of* $\{x^r, y^r\}$ *is a primal and dual optimal solution.*
2. *For the RBSUM-M, we have* $\lim_{t\to\infty} \|Ex^t - q\| = 0$ *w.p.1. Further, every limit point of* $\{x^t, y^t\}$ *is a primal and dual optimal solution w.p.1.*

Counterexample for multi-block ADMM

- Recently [Chen-He-Ye-Yuan 13] shows (through an example) that applying ADMM to multi-block problem can diverge
- We show applying (R)BSUM-M to the same problem converges (diminishing dual step size)
- **Main message**: Dual stepsize control is crucial
- Consider the following linear systems of equations (unique solution $x_1 = x_2 = x_3 = 0$)

$$E_1 x_1 + E_2 x_2 + E_3 x_3 = 0,$$

$$\text{with} \quad [E_1 \; E_2 \; E_3] = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Counterexample for multi-block ADMM (cont.)

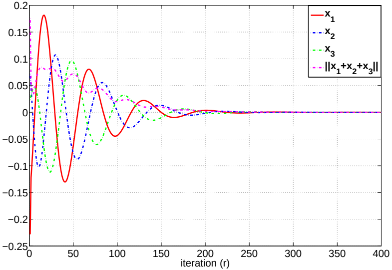

Figure: Iterates generated by the BSUM-M. Each curve is averaged over 1000 runs (with random starting points).

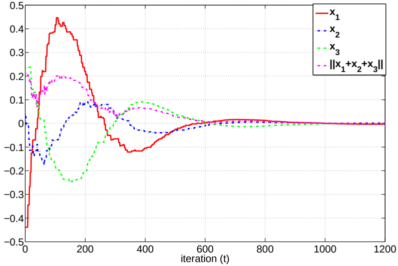

Figure: Iterates generated by the RBSUM-M algorithm. Each curve is averaged over 1000 runs (with random starting points)

Thank you for your attention !