Nonconvex ADMM: Convergence and Applications

Yu Wang (Xi'an Jiaotong), Wotao Yin (UCLA), Jinshan Zeng (Jiangxi Normal)

SIAM Imaging'16 — May 26, 2016

1. Alternating Direction Method of Multipliers (ADMM): Background and Existing Work

Basic Formulation

$$\begin{aligned} \underset{x,y}{\text{minimize}} \;\; & f(x) + h(y) \\ \text{subject to } & Ax + By = b \end{aligned}$$

- functions f, h can take the extended value ∞, can be nonsmooth

ADMM

- Define the **augmented Lagrangian**

$$\mathcal{L}_\beta(x, y; w) = f(x) + h(y) + \langle w, Ax + By - b\rangle + \frac{\beta}{2}\|Ax + By - b\|_2^2$$

- **Algorithm:**
  - $x^{k+1} \in \arg\min_x \mathcal{L}_\beta(x, y^k; w^k)$
  - $y^{k+1} \in \arg\min_y \mathcal{L}_\beta(x^{k+1}, y; w^k)$
  - $w^{k+1} = w^k + \beta(Ax^{k+1} + By^{k+1} - b)$

- **Feature**: splits numerically awkward combinations of f and h

- Often, one or both subproblems are easy to solve

Brief history (convex by default)

- 1950s, Douglas-Rachford Splitting (DRS) for PDEs
- ADM (ADMM) Glowinski and Marroco'75, Gabay and Mercier'76
- Convergence proof: Glowinski'83
- ADMM=dual-DRS (Gabay'83), ADMM=DRS and ADMM=dual-ADMM (Eckstein'89, E.-Fukushima'94, Yan-Yin'14), ADMM=PPA (E.'92)
- if a subproblem is quadratic, equivalent under order swapping (Yan-Yin'14)
- Convergence rates (Monterio-Svaiter'12, He-Yuan'12, Deng-Yin'12, Hong-Luo'13, Davis-Yin'14, ...)
- Accelerations (Goldstein et al'11, Ouyang et al'13)
- Nonconvex (Hong-Luo-Raz...'14, Wang-Cao-Xu'14, Li-Pong'14, **this work**)

2. Nonconvex ADMM Applications

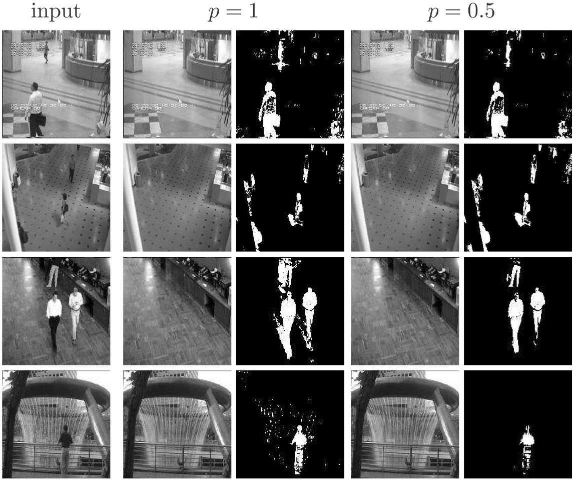
Background extraction from video

- From observation b of a video Z, decompose it into low-rank background L and sparse foreground S by

$$\begin{aligned} &\underset{Z,L,S}{\text{minimize}} \;\; \Psi(L) + \Phi(S) + \frac{1}{2}\|A(Z) - b\|_F^2 \\ &\text{subject to } L + S = Z. \end{aligned}$$

- Originally proposed by J.Wright et al. as Robust PCA
- Yuan-Yang'09 and Shen-Wen-Zhang'12 apply convex ADMM
- R.Chartrand'12 and Yang-Pong-Chen'14 use nonconvex regularization

Results of ℓ_p-minimization for S from Yang-Pong-Chen'14



Matrix completion with nonnegative factors

- From partial observations, recover a matrix $Z \approx XY$ where $X, Y \geq 0$
- Xu-Yin-Wen-Zhang'12 applies ADMM to the model

$$\begin{aligned} \underset{X,Y,Z,U,V}{\text{minimize}} \ & \frac{1}{2}\|XY - Z\|_F^2 + \iota_{\geq 0}(U) + \iota_{\geq 0}(V) \\ \text{subject to } & X - U = 0 \\ & Y - V = 0 \\ & \mathrm{Proj}_{\Omega}(Z) = \text{observation.} \end{aligned}$$

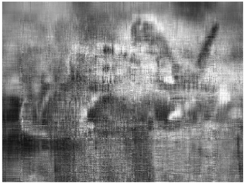
- The objective is nonconvex due to XY

Results from Xu-Yin-Wen-Zhang'12

Original images

Results from Xu-Yin-Wen-Zhang'12

Recovered images (SR: sample ratio)



ADM SR = 0.1

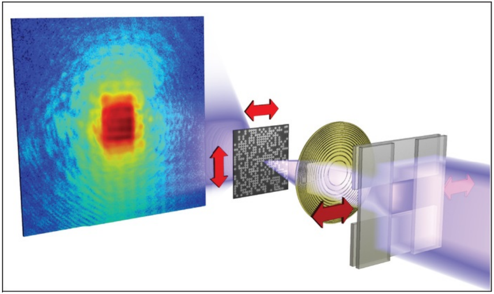
ADM SR = 0.1

ADM SR = 0.2

ADM SR = 0.15

Ptychographic phase retrieval

- Ptychography: a diffractive imaging technique that reconstructs an object from a set of diffraction patterns produced by a moving probe. The probe illuminates a portion of the object at a time.



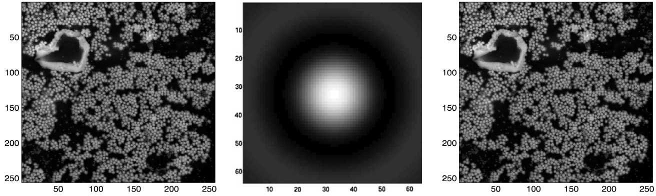
Thibault-Menzel'13

- Phaseless measurements: $b_i = |\mathcal{F} Q_i x|$, where x is the object and Q_i is an illumination matrix.

- let $|z|$ denote the amplitude vector of a complex vector z
- Wen-Yang-Liu-Marchesini'12 develops nonconvex ADMM for the model

$$\underset{x, z_1, \ldots, z_p}{\text{minimize}} \;\; \frac{1}{2}\Big\| |z_1| - b_1 \Big\|^2 + \cdots + \frac{1}{2}\Big\| |z_p| - b_p \Big\|^2$$

$$\text{subject to } z_i - \mathcal{F}Q_i x = 0, \quad i = 1, \ldots, p.$$



(a) (b) (c)

|original "gold ball"| |prob| |recovered "gold ball"|

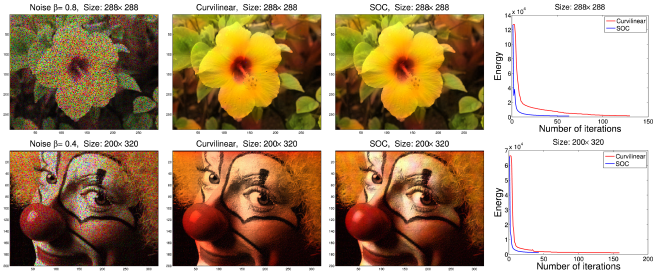
Optimization on spherical and Stiefel manifolds

- Lai-Osher'12 develops nonconvex ADMM for

$$\begin{aligned} \underset{X,P}{\text{minimize}} \;\; & f(X) + \iota_{\mathcal{P}}(P) \\ \text{subject to} \;\; & X - P = 0. \end{aligned}$$

- Examples of $\mathcal{P}$
  - Spherical manifold $\mathcal{P} = \{P : \|P(:,i)\|_2 = 1\}$
  - Stiefel manifold $\mathcal{P} = \{P : P^T P = I\}$

Chromatic-noise removal results from Lai-Osher'12



- "Curvilinear" is a feasible algorithm for manifold optimization from Wen-Yin'10

Mean-ρ-Basel portfolio optimization

- Goal: allocate assets for expected return, Basel regulation, and low risk
- Wen-Peng-Liu-Bai-Sun'13 applies nonconvex ADMM to solve this problem

$$\begin{aligned} \underset{u,x,y}{\text{minimize}} \;\; & \iota_{\mathcal{U}}(u) + \iota\rho_{\text{Basel}<C}(x) + \rho(y) \\ \text{subject to}\;\; & x + Ru = 0 \\ & y + Yu = 0. \end{aligned}$$

- $\mathcal{U} = \{u \geq 0 : \mu^T u \geq r, \mathbf{1}^T u = 1\}$
- $\rho_{\text{Basel}<C}(-Ru)$ is Basel Accord requirement, calculated on certain regulated dataset R
- $\rho(-Yu)$ is the risk measure, such as variance, VaR, CVaR
- Their results are **reportedly better than MIPs solved by CPLEX**

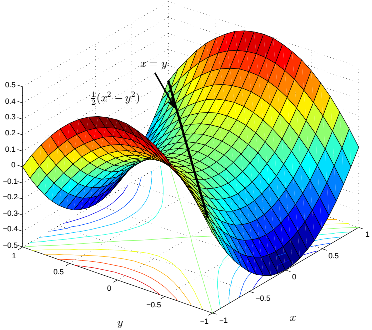
Other applications

- tensor factorization (Liavas-Sidiropoulos'14)
- compressive sensing (Chartrand-Wohlberg'13)
- optimal power flow (You-Peng'71)
- direction fields correction, global conformal mapping (Lai-Osher'14)
- image registration (Bouaziz-Tagliasacchi-Pauly'13)
- network inference (Miksik et al'14)

3. A simple example

A simple example

$$\underset{x,y\in\mathbb{R}}{\text{minimize}} \ \frac{1}{2}(x^2 - y^2)$$
$$\text{subject to } x - y = 0$$
$$x \in [-1, 1]$$



- **augmented Lagrangian**

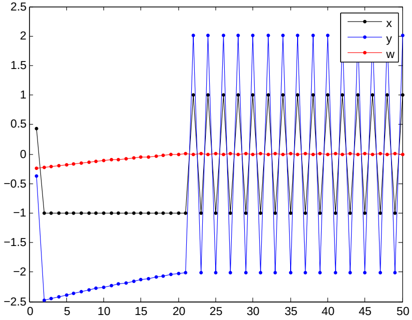
$$L_\beta(x, y, w) := \frac{1}{2}(x^2 - y^2) + \iota_{[-1,1]}(x) + w(x - y) + \frac{\beta}{2}\left|x - y\right|^2$$

- **ALM diverges** for any fixed β (but will converge if $\beta \to \infty$)
- **ADMM converges** for any fixed $\beta > 1$

Numerical ALM

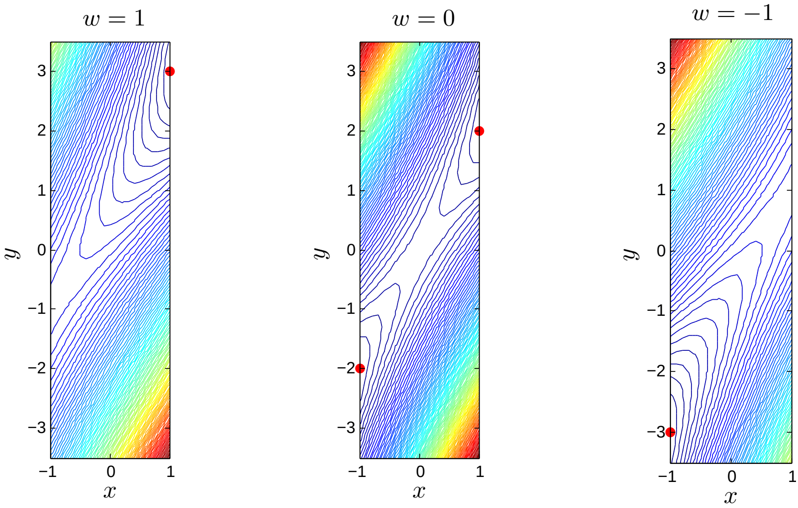
- set $\beta = 2$, initialize x, y, w as iid `randn`
- ALM iteration:

$$(x^{k+1}, y^{k+1}) = \arg\min_{x,y} L_\beta(x, y, w^k);$$
$$w^{k+1} = w^k + \beta(x^{k+1} - y^{k+1});$$



x^k, y^k oscillate, w^k also does in a small amount

why ALM diverges: $(x, y) = \arg\min_{x,y} L_\beta(x, y, w)$ is too sensitive in w



Contours of $L_\beta(x, y, w)$ for $\beta = 2$ and varying w

ADMM

- ADMM following the order $x \to y \to w$:

$$\begin{cases} x^{k+1} = \arg\min_x L_\beta(x, y^k, w^k) \\ y^{k+1} = \arg\min_y L_\beta(x^{k+1}, y, w^k) \\ w^{k+1} = w^k + \alpha\beta(x^{k+1} - y^{k+1}) \end{cases}$$

  or the order $y \to x \to w$:

$$\begin{cases} y^{k+1} = \arg\min_y L_\beta(x^k, y, w^k) \\ x^{k+1} = \arg\min_x L_\beta(x, y^{k+1}, w^k) \\ w^{k+1} = w^k + \alpha\beta(x^{k+1} - y^{k+1}) \end{cases}$$

- when $\beta > 1$, both x- and y-subproblems are (strongly) convex, so their solutions are stable

ADMM following the order $x \to y \to w$

$$\begin{cases} x^{k+1} = \mathbf{proj}_{[-1,1]}\left(\frac{1}{\beta+1}(\beta y^k - w^k)\right) \\ y^{k+1} = \frac{1}{\beta-1}(\beta x^{k+1} + w^k) \\ w^{k+1} = w^k + \alpha\beta(x^{k+1} - y^{k+1}) \end{cases}$$

- supposing $\alpha = 1$ and eliminating $y^k \equiv -w^k$, we get

$$\begin{cases} x^{k+1} = \mathbf{proj}_{[-1,1]}(-w^k) \\ w^{k+1} = \frac{-1}{\beta-1}\left(\beta x^{k+1} + w^k\right) \end{cases} \quad \Rightarrow \quad w^{k+1} = \frac{-1}{\beta - 1}(\beta \mathbf{proj}_{[-1,1]}(-w^k) + w^k)$$

- pick $\beta > 2$ and change variable $\beta\bar{w}^k \leftarrow w^k$
  - if $w^k \in [-1,1]$, then $\mathbf{proj}_{[-1,1]}(-w^k) = -w^k$ and $w^{k+1} = w^k$
  - o.w., $\bar{w}^{k+1} = \frac{1}{\beta-1}(\mathrm{sign}(\bar{w}^k) - \bar{w}^k)$ so $|\bar{w}^{k+1}| = \frac{1}{\beta-1}\Big||\bar{w}^k| - 1\Big|$

  $\{x^k, y^k, w^k\}$ **converges geometrically with finite termination**

ADMM following the order $y \to x \to w$

$$\begin{cases} y^{k+1} = \frac{1}{\beta-1}(\beta x^k + w^k) \\ x^{k+1} = \mathbf{proj}_{[-1,1]}\big(\frac{1}{\beta+1}(\beta y^{k+1} - w^k)\big) \\ w^{k+1} = w^k + \alpha\beta(x^{k+1} - y^{k+1}) \end{cases}$$

- set $\alpha = 1$ and introduce $z^k = \frac{1}{\beta^2-1}(\beta^2 x^k + w^k)$; we get

  $$z^{k+1} = \frac{1}{\beta - 1}\big(\beta \mathbf{proj}_{[-1,1]}(z^k) - z^k\big),$$

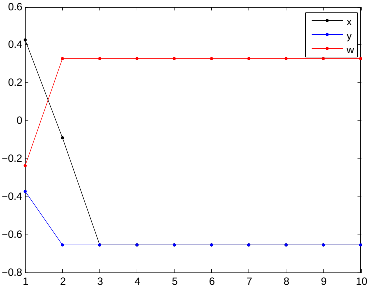
  which is similar to w^{k+1} in ADMM $x \to y \to w$.

- $x^{k+1} = \mathbf{proj}_{[-1,1]}(z^k)$ and $w^{k+1} = \beta x^{k+1} - (\beta + 1)z^k$

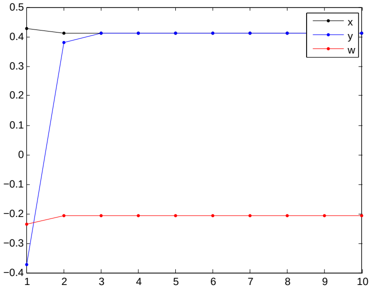
- $\{x^k, y^k, w^k\}$ **converges geometrically with finite termination**

Numerical test: finite convergence

ADMM $x \to y \to w$



ADMM $y \to x \to w$



Both iterations converge to a global solution in 3 steps

Why ADMM converges? Reduces to convex coordinate descent

- For this problem, we can show $y^k \equiv -w^k$ for ADMM $x \to y \to w$
- Setting $w = -y$ yields a convex function:

$$
\begin{aligned}
L_\beta(x, y, w)\big|_{w=-y} &= \frac{1}{2}(x^2 - y^2) + \iota_{[-1,1]}(x) - y(x-y) + \frac{\beta}{2}\big|x - y\big|^2 \\
&= \frac{\beta+1}{2}|x-y|^2 + \iota_{[-1,1]}(x) \\
&=: f(x, y)
\end{aligned}
$$

- ADMM $x \to y \to w$ $=$ coordinate descent to the convex $f(x, y)$:

$$
\begin{cases}
x^{k+1} = \arg\min_x f(x, y^k) \\
y^{k+1} = y^k - \rho \frac{\mathrm{d}}{\mathrm{d}y} f(x^{k+1}, y^k)
\end{cases}
$$

where $\rho = \frac{\beta}{\beta^2 - 1}$

4. New convergence results

The generic model

$$\begin{aligned} \underset{x_1,\ldots,x_p,y}{\text{minimize}} \quad & \phi(x_1,\ldots,x_p,y) \\ \text{subject to} \quad & A_1x_1 + \cdots + A_px_p + By = b, \end{aligned} \tag{1}$$

- we single out y because of its unique role: "locking" the dual variable w^k

Notation:

- $\mathbf{x} := [x_1; \ldots; x_p] \in \mathbb{R}^n$
- $\mathbf{x}_{<i} := [x_1; \ldots; x_{i-1}]$
- $\mathbf{x}_{>i} := [x_{i+1}; \ldots; x_p]$
- $\mathbf{A} := [A_1 \ \cdots \ A_p] \in \mathbb{R}^{m \times n}$
- $\mathbf{Ax} := \sum_{i=1}^{p} A_i x_i \in \mathbb{R}^m$.
- **Augmented Lagrangian**:

$$L_\beta(x_1, \ldots, x_p, y, w) = \phi(x_1, \ldots, x_p, y) + \langle w, \mathbf{Ax} + By - b \rangle + \frac{\beta}{2} \left\| \mathbf{Ax} + By - b \right\|^2$$

The Gauss-Seidel ADMM algorithm

0. initialize $\mathbf{x}^0, y^0, w^0$
1. for $k = 0, 1, ...$ do
2. for $i = 1, \ldots, p$ do
3. $x_i^{k+1} \leftarrow \arg\min_{x_i} L_\beta(x_{<i}^{k+1}, x_i, x_{>i}^k, y^k, w^k)$;
4. $y^{k+1} \leftarrow \arg\min_y L_\beta(\mathbf{x}^{k+1}, y, w^k)$;
5. $w^{k+1} \leftarrow w^k + \beta\left(\mathbf{A}\mathbf{x}^{k+1} + By^{k+1} - b\right)$;
6. if stopping conditions are satisfied, **return** $x_1^k, \ldots, x_p^k$ and y^k.

The overview of analysis

- Loss of convexity $\Rightarrow$ no Fejer-monotonicity, or VI based analysis.
- Choice of Lyapunov function is critical. Following Hong-Luo-Razaviyayn'14, we use the augmented Lagrangian.
- The last block y plays an important role.

ADMM is better than ALM for a class of nonconvex problems

- ALM: nonsmoothness generally requires $\beta \to \infty$;
- ADMM: works with a finite β if the problem has the y-block (h, B) where h is smooth and $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$, even if the problem is nonsmooth
- in addition, ADMM has simpler subproblems

Analysis keystones

P1 **(boundedness)** $\{\mathbf{x}^k, y^k, w^k\}$ is bounded, $L_\beta(\mathbf{x}^k, y^k, w^k)$ is lower bounded;

P2 **(sufficient descent)** for all sufficiently large k, we have

$$\begin{aligned}
&L_\beta(\mathbf{x}^k, y^k, w^k) - L_\beta(\mathbf{x}^{k+1}, y^{k+1}, w^{k+1}) \\
&\geq C_1\Big(\|B(y^{k+1} - y^k)\|^2 + \sum_{i=1}^{p} \|A_i(x_i^k - x_i^{k+1})\|^2\Big),
\end{aligned}$$

P3 **(subgradient bound)** exists $d^{k+1} \in \partial L_\beta(\mathbf{x}^{k+1}, y^{k+1}, w^{k+1})$ such that

$$\|d^{k+1}\| \leq C_2\Big(\|B(y^{k+1} - y^k)\| + \sum_{i=1}^{p} \|A_i(x_i^{k+1} - x_i^k)\|\Big).$$

Similar to coordinate descent but treats w^k in a special manner

Proposition

Suppose that the sequence $(\mathbf{x}^k, y^k, w^k)$ *satisfies P1–P3.*

(i) It has at least a limit point $(\mathbf{x}^*, y^*, w^*)$*, and any limit point* $(\mathbf{x}^*, y^*, w^*)$ *is a stationary solution. That is,* $0 \in \partial L_\beta(\mathbf{x}^*, y^*, w^*)$*.*

(ii) The running best rates [a] *of* $\{\|B(y^{k+1} - y^k)\|^2 + \sum_{i=1}^p \|A_i(x_i^k - x_i^{k+1})\|^2\}$ *and* $\{\|d^{k+1}\|^2\}$ *are* $o(\frac{1}{k})$*.*

(iii) If L_β *is a KŁ function, then converges globally to the point* $(\mathbf{x}^*, y^*, w^*)$*.*

[a] A nonnegative sequence a_k induces its running best sequence $b_k = \min\{a_i : i \le k\}$; therefore, a_k has running best rate of $o(1/k)$ if $b_k = o(1/k)$.

The proof is rather standard.

y^k controls w^k

- **Notation**: $\cdot^+$ denotes $\cdot^{k+1}$
- **Assumption**: β is sufficiently large but fixed
- By combining y-update and w-update
  (plugging $w^k = w^{k-1} + \beta(\mathbf{A}\mathbf{x}^k + By^k - b)$ into the y-optimality cond.)

$$0 = \nabla h(y^k) + B^T w^k, \quad k = 1, 2, \ldots$$

- Assumption $\{b\} \cup \mathrm{Im}(A) \subseteq \mathrm{Im}(B) \;\Rightarrow\; w^k \in \mathrm{Im}(B)$
- Then, with additional assumptions, we have

$$\|w^+ - w^k\| \leq O(\|By^+ - By^k\|)$$

  and

$$L_\beta(x^+, y^k, w^k) - L_\beta(x^+, y^+, w^+) \geq O(\|By^+ - By^k\|^2)$$

  (see the next slide for detailed steps)

Detailed steps

- Bound Δw by ΔBy:

  $$\|w^+ - w^k\| \le C\|B^T(w^+ - w^k)\| = O(\|\nabla h(y^+) - \nabla h(y^k)\|) \le O(\|By^+ - By^k\|)$$

  where $C := \lambda_{++}^{-1/2}(B^T B)$, the 1st "$\le$" follows from $w^+, w^k \in \mathrm{Im}(B)$, and the 2nd "$\le$" follows from the assumption of Lipschitz sub-minimization path (see later)

- Then, smooth h leads to sufficient decent during the y- and w-updates:

  $$\begin{aligned}
  &L_\beta(x^+, y^k, w^k) - L_\beta(x^+, y^+, w^+) \\
  =&\big(h(y^k) - h(y^+) + \langle w^+, By^k - By^+\rangle\big) + \frac{\beta}{2}\|By^+ - By^k\|^2 - \frac{1}{\beta}\|w^+ - w^k\|^2 \\
  \ge& - O(\|By^+ - By^k\|^2) + \frac{\beta}{2}\|By^+ - By^k\|^2 - O(\|By^+ - By^k\|) \\
  &\text{(with suff. large } \beta\text{)} \\
  =&O(\|By^+ - By^k\|^2)
  \end{aligned}$$

  where the "$\ge$" follows from the assumption of Lipschitz sub-minimization path (see later)

x^k-subproblems: fewer conditions on f, A

We only need conditions to ensure monotonicity and sufficient decent like

- $L_\beta(x^+_{<i}, x^k_i, x^k_{>i}, y^k, w^k) \geq L_\beta(x^+_{<i}, x^+_i, x^k_{>i}, y^k, w^k))$
- and sufficient descent:

  $$L_\beta(x^+_{<i}, x^k_i, x^k_{>i}, y^k, w^k) - L_\beta(x^+_{<i}, x^+_i, x^k_{>i}, y^k, w^k)) \geq O(\|A_i x^k_i - A_i x^+_i\|^2)$$

  For Gauss-Seidel updates, the proof is inductive $i = p, p-1, \ldots, 1$

A sufficient condition for what we need:

$f(x_1, \ldots, x_p)$ has the form: smooth + separable-nonsmoooth

Remedy of nonconvexity: Prox-regularity

- A convex function f has subdifferentials in $\mathrm{int}(\mathrm{dom} f)$ and satisfies

$$f(y) \geq f(x) + \langle d, y - x \rangle, \quad x, y \in \mathrm{dom} f, d \in \partial f(x)$$

- A function f is *prox-regular* if $\exists\ \gamma$ such that

$$f(y) + \frac{\gamma}{2}\|x - y\|^2 \geq f(x) + \langle d, y - x \rangle, \quad x, y \in \mathrm{dom} f, d \in \partial f(x)$$

  where ∂f is the *limiting subdifferential*.

- **Limitation**: not satisfied by functions with sharps, e.g., $\ell_{1/2}$, which are often used in sparse optimization.

Restricted prox-regularity

- **Motivation**: your points do not land on the steep region around the sharp, which we call the *exclusion set*
- **Exclusion set**: for $M > 0$, define

  $$S_M := \{x \in \mathrm{dom}(\partial f) : \|d\| > M \text{ for all } d \in \partial f(x)\}$$

  idea: points in S_M are never visited (for a suff. large M)
- A function is *restricted prox-regular* if $\exists\ M, \gamma > 0$ such that $S_M \subseteq \mathrm{dom}(\partial f)$ and any bounded $T \in \mathrm{dom}(f)$

  $$f(y)+\frac{\gamma}{2}\|x-y\|^2 \geq f(x)+\langle d, y-x\rangle, \quad x, y \in T \setminus S_M,\ d \in \partial f(x),\ \|d\| \leq M.$$
- Example: ℓ_q quasinorm, Schattern$-q$ quasinorm, indicator function of compact smooth manifold

Main theorem 1

Assumptions: $\phi(x_1, \ldots, x_n, y) = f(\mathbf{x}) + h(y)$

A1. the problem is feasible, the objective is feasible-coercive[1]

A2. $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$

A3. $f(\mathbf{x}) = g(\mathbf{x}) + f_1(x_1) + \cdots + f_n(x_n)$, where

- g is Lipschitz differentiable
- f_i is either restricted prox-regular, or continuous and piecewise linear[2]

A4. $h(y)$ is Lipschitz differentiable

A5. x and y subproblems have Lipschitz sub-minimization paths

Results: subsequential convergence to a stationary point from any start point; if L_β is KŁ, then whole-sequence convergence.

[1] For **feasible** points $(x_1, \ldots, x_p, y)$, if $\|(x_1, \ldots, x_n, y)\| \to \infty$, then $\phi(x_1, \ldots, x_n, y) \to \infty$.

[2] e.g., anisotropic total variation, sorted ℓ_1 function (nonconvex), $(-\ell_1)$ function, continuous piece-wise linear approximation of a function

Necessity of assumptions A2 A4

- Assumptions **A2 A4** apply to the last block (h, B)
- **A2** cannot be completely dropped.
  Counter example: the 3-block divergence example by Chen-He-Ye-Yuan'13
- **A4** cannot be completely dropped.
  Counter example:

$$\begin{aligned} \underset{x,y}{\text{minimize}} \quad & -|x| + |y| \\ \text{subject to } & x - y = 0, \; x \in [-1, 1]. \end{aligned}$$

  ADMM generates the alternating sequence $\pm(\frac{2}{\beta}, 0, 1)$

Lipschitz sub-minimization path

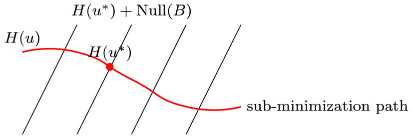
- ADMM subproblem has the form

$$y^k \in \arg\min_y \, h(y) + \frac{\beta}{2}\|By + \text{constants}\|^2$$

- Let $u = By^k$. Then y^k is also the solution to

$$\operatorname*{minimize}_y \, h(y) \quad \text{subject to } By = u.$$

- We assume a Lipschitz subminimization path



- Sufficient conditions: (i) smooth h + full col-rank B, (ii) smooth and strongly convex h; (iii) not above but your subprob solver warmstarts and finds a nearby solution.

Main theorem 2

Assumptions: $\phi(x_1, \ldots, x_n, y)$ can be fully coupled

- Feasible, the objective is feasible-coercive
- $\mathrm{Im}(A) \subseteq \mathrm{Im}(B)$
- ϕ is Lipschitz differentiable
- x and y subproblems have Lipschitz sub-minimization paths

Results: subsequential convergence to a stationary point from any start point; if L_β is KL, then whole-sequence convergence.

5. Comparison with Recent Results

Compare to Hong-Luo-Razaviyayn'14

- Their assumptions are strictly stronger, e.g., only smooth functions
  - $f = \sum_i f_i$, where f_i Lipschitz differentiable or convex
  - h Lipschitz differentiable
  - A_i has full col-rank and $B = I$
- Applications in consensus and sharing problems.

Compare to Li-Pong'14

- Their assumptions are strictly stronger
  - $p = 1$ and f is l.s.c.
  - $h \in C^2$ is Lipschitz differentiable and strongly convex
  - $A = I$ and B has full row-rank
  - h is coercive and f is lower bounded.

Compare to Wang-Cao-Xu'14

- Analyzed Bregman ADMM, which reduces to ADMM with vanishing aux. functions.
- Their assumptions are strictly stronger
  - B is invertible
  - $f(x) = \sum_{i=1}^{p}$, where f_i is strongly convex
  - h is Lipschitz differentiable and lower bounded.

6. Applications of Nonconvex ADMM with Convergence Guarantees

Application: statistical learning

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \;\; r(x) + \sum_{i=1}^{p} l_i(A_i x - b_i)$$

- r is regularization, l_i's are fitting measures
- ADMM-ready formulation

$$\begin{aligned} \underset{x, \{z_i\}}{\text{minimize}} \;\; & r(x) + \sum_{i=1}^{p} l_i(A_i z_i - b_i) \\ \text{subject to } & x = z_i, \;\; i = 1, \ldots, p. \end{aligned}$$

- ADMM will converge if
  - $r(x) = \|x\|_q^q = \sum_i |x_i|^q$, for $0 < q \leq 1$, or piecewise linear
  - $r(x) + \sum_{i=1}^{p} l_i(A_i x - b_i)$ is coercive
  - $l_1, \ldots, l_p$ are Lipschitz differentiable

Application: optimization on smooth manifold

$$\underset{x}{\text{minimize}}\ J(x) \quad \text{subject to } x \in S.$$

- ADMM-ready formulation

$$\begin{aligned} &\underset{x,y}{\text{minimize}}\ \iota_S(x) + J(y) \\ &\text{subject to } x - y = 0. \end{aligned}$$

- ADMM will converge if
  - S is a compact smooth manifold, e.g., sphere, Stiefel, and Grassmann manifolds
  - J is Lipschitz differentiable

Application: matrix/tensor decomposition

$$\underset{X,Y,Z}{\text{minimize}} \;\; r_1(X) + r_2(Y) + \|Z\|_F^2$$

$$\text{subject to } X + Y + Z = \text{Input}.$$

- Video decomposition: background + foreground + noise
- Hyperspectral decomposition: background + foreground + noise
- ADMM will converge if r_1 and r_2 satisfy our assumptions on f

6. Summary

Summary

- ADMM indeed works for some nonconvex problems!
- The theory indicates that ADMM works better than ALM when the problem has a block $(h(y), B)$ where h is smooth and $\mathrm{Im}(B)$ is dominant
- Future directions: weaker conditions, numerical results

Thank you!

Acknowledgements: Wei Shi, Ting Kei Pong, Qing Ling, Xin Liu, Yangyang Xu, and Yiqiu Dong

Supports: NSF

Reference: Yu Wang, Wotao Yin, and Jinshan Zeng. UCLA CAM 15-62.