Splitting methods for nonconvex feasibility problems

Ting Kei Pong
Assistant Professor
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong

Tutte Seminar
July 2015
(Joint work with Guoyin Li)

Feasibility Problem

- Given closed sets D_i, $i = 1, \ldots, m$, find a point

$$x \in \bigcap_{i=1}^{m} D_i.$$

- Example: Finding a solution of $Ax = b$ with $\|x\|_0 \leq r$.

Feasibility Problem

- Given closed sets D_i, $i = 1, \ldots, m$, find a point

$$x \in \bigcap_{i=1}^{m} D_i.$$

- Example: Finding a solution of $Ax = b$ with $\|x\|_0 \leq r$.
- The general problem can be reformulated as finding a point in

$$\{(x_1, \ldots, x_m) : \; x_1 = \cdots = x_m\} \cap (D_1 \times D_2 \times \cdots \times D_m).$$

- Only need to consider the intersection of a closed *convex* set C and a closed set D.

When D is convex

- Alternating projection:

$$x^{t+1} = P_D(P_C(x^t)).$$

- Splitting methods ($0 < \alpha \leq 2$):

$$\begin{cases} y^{t+1} = \arg\min_{y \in C} \left\{ \|y - x^t\| \right\}, \\ z^{t+1} = \arg\min_{z \in D} \left\{ \|2y^{t+1} - x^t - z\| \right\}, \\ x^{t+1} = x^t + \alpha(z^{t+1} - y^{t+1}). \end{cases}$$

When D is convex

- Alternating projection:

$$x^{t+1} = P_D(P_C(x^t)).$$

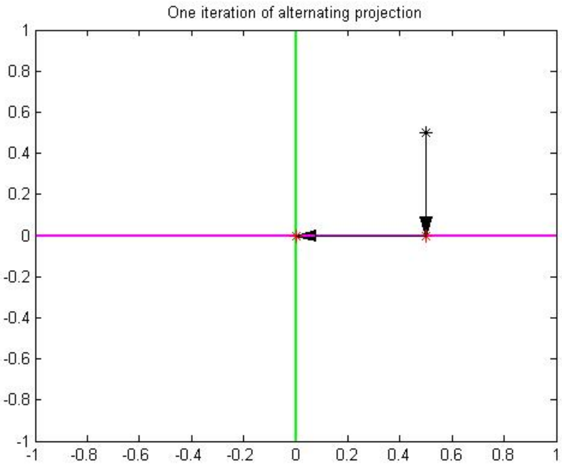
- Splitting methods ($0 < \alpha \leq 2$):

$$\begin{cases} y^{t+1} = \underset{y \in C}{\arg\min} \left\{ \|y - x^t\| \right\}, \\ z^{t+1} = \underset{z \in D}{\arg\min} \left\{ \|2y^{t+1} - x^t - z\| \right\}, \\ x^{t+1} = x^t + \alpha(z^{t+1} - y^{t+1}). \end{cases}$$

- Douglas-Rachford (DR): $\alpha = 1$.
- Peaceman-Rachford (PR): $\alpha = 2$.

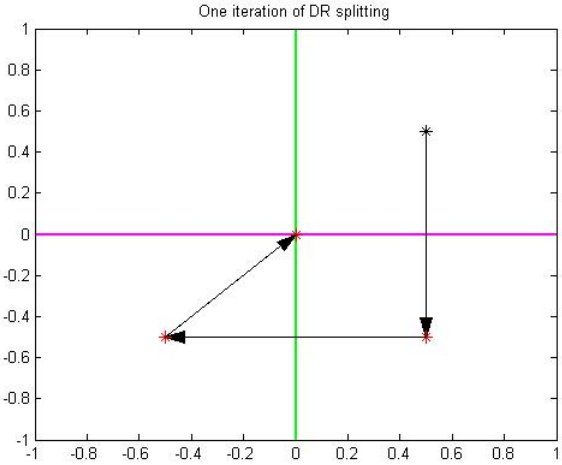
Behavior in convex case: AP

Finding the intersection of the axes, starting from $(0.5, 0.5)$.



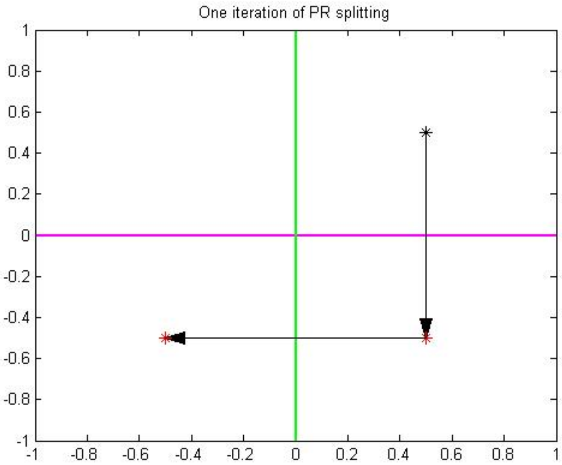
Behavior in convex case: DR

Showing the x-iterates: average after two successive reflections.
$P_C(x^t)$ will converge to the intersection.



Behavior in convex case: PR

Showing the x-iterate; not convergent.



When D is nonconvex

For the convergence of DR splitting:

- Mainly local convergence results.
- Require various regularity conditions on the sets.
- Local convergence for finding intersection of $Ax = b$ and $\|x\|_0 \leq r$. (Hesse, Luke, Neumann '13).

When D is nonconvex

For the convergence of DR splitting:

- Mainly local convergence results.
- Require various regularity conditions on the sets.
- Local convergence for finding intersection of $Ax = b$ and $\|x\|_0 \leq r$. (Hesse, Luke, Neumann '13).
- Global convergence shown for the intersection of a circle and a straight line in $\mathbb{R}^2$. (Artacho, Borwein '12)

Our DR splitting

- DR splitting: ($\gamma > 0$)

$$\begin{cases} y^{t+1} = \arg\min_{y} \left\{ \frac{1}{2} d_C^2(y) + \frac{1}{2\gamma}\|y - x^t\|^2 \right\}, \\ z^{t+1} \in \operatorname*{Arg\,min}_{z \in D} \left\{ \|2y^{t+1} - x^t - z\|^2 \right\}, \\ x^{t+1} = x^t + (z^{t+1} - y^{t+1}). \end{cases}$$

- The y-update is $\frac{1}{1+\gamma}(x^t + \gamma P_C(x^t))$.

Our DR splitting

- DR splitting: $(\gamma > 0)$

$$\begin{cases} y^{t+1} = \arg\min_{y} \left\{ \frac{1}{2} d_C^2(y) + \frac{1}{2\gamma}\|y - x^t\|^2 \right\}, \\ z^{t+1} \in \operatorname*{Arg\,min}_{z \in D} \left\{ \|2y^{t+1} - x^t - z\|^2 \right\}, \\ x^{t+1} = x^t + (z^{t+1} - y^{t+1}). \end{cases}$$

- The y-update is $\frac{1}{1+\gamma}(x^t + \gamma P_C(x^t))$.
- DR splitting applied to minimizing $\frac{1}{2}d_C^2 + \delta_D$.

DR Convergence result I

Fact 1 (Li, P '14): [Global convergence]
Suppose that $0 < \gamma < \sqrt{\frac{3}{2}} - 1$, and either C or D is compact.

Then the sequence $\{(y^t, z^t, x^t)\}$ generated from DR splitting is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_{u \in D} \frac{1}{2} d_C^2(u),$$

i.e., $0 \in y^* - P_C(y^*) + N_D(y^*)$.

DR Convergence result I

Fact 1 (Li, P '14): [Global convergence]

Suppose that $0 < \gamma < \sqrt{\frac{3}{2}} - 1$, and either C or D is compact.

Then the sequence $\{(y^t, z^t, x^t)\}$ generated from DR splitting is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_{u \in D} \ \frac{1}{2} d_C^2(u),$$

i.e., $0 \in y^* - P_C(y^*) + N_D(y^*)$.

- Clearly, if $d_C(y^*) = 0$, then y^* solves the feasibility problem.

DR Convergence result II

Fact 2 (Li, P '14): [Convergence of the whole sequence]

Suppose that $0 < \gamma < \sqrt{\frac{3}{2}} - 1$, C and D are semi-algebraic, and one of them is compact.

Then the sequence $\{(y^t, z^t, x^t)\}$ generated from DR splitting is bounded, and is convergent to some (y^*, z^*, x^*) satisfying $z^* = y^*$, with y^* being a stationary point of the problem $\min_{u \in D} \frac{1}{2} d_C^2(u)$. Furthermore,

$$\sum_{t=1}^{\infty} \|y^{t+1} - y^t\| < \infty.$$

DR Convergence result III

Fact 3 (Li, P '14): [Local convergence]
Let $C = \{x : \; Ax = b\}$ and D be a closed semi-algebraic set, $0 < \gamma < \sqrt{\frac{3}{2}} - 1$ and $\lim(y^t, z^t, x^t) = (y^*, z^*, x^*)$.
Suppose that $z^* \in C \cap D$ with

$$N_C(z^*) \cap -N_D(z^*) = \{0\}.$$

Then there exist $\eta \in (0, 1)$ and $\kappa > 0$ such that for all large t,

$$\mathrm{dist}\big(0, z^t - P_C(z^t) + N_D(z^t)\big) \leq \kappa\, \eta^t.$$

Our DR vs classical DR

- Example that the classical DR diverges: for $\eta \in (0, 1]$ (Bauschke, Noll '14)

$$C = \{x \in \mathbb{R}^2 : \ x_2 = 0\}$$
$$D = \{(0, 0), (7 + \eta, \eta), (7, -\eta)\}$$

Initialized at $x^0 = (7, \eta)$, the classical DR exhibits a discrete limit cycle.

Our DR vs classical DR

- Example that the classical DR diverges: for $\eta \in (0, 1]$ (Bauschke, Noll '14)

$$C = \{x \in \mathbb{R}^2 : \ x_2 = 0\}$$
$$D = \{(0, 0), (7 + \eta, \eta), (7, -\eta)\}$$

  Initialized at $x^0 = (7, \eta)$, the classical DR exhibits a discrete limit cycle.

- For our DR, with $\gamma \in (0, \sqrt{\frac{3}{2}} - 1)$, explicit computation shows that $y^* = z^* = (7 + \eta, \eta)$ and $x^* = (7 + \eta, (1 + \gamma)\eta)$ for this starting point.

For PR splitting

- Does not converge in general even if D is convex.
- Modifying as follows also cannot guarantee convergence even if both sets are convex:

$$
\begin{cases}
y^{t+1} = \arg\min_{y} \left\{ \frac{1}{2} d_C^2(y) + \frac{1}{2\gamma} \|y - x^t\|^2 \right\}, \\
z^{t+1} \in \operatorname{Arg\,min}_{z \in D} \left\{ \|2y^{t+1} - x^t - z\|^2 \right\}, \\
x^{t+1} = x^t + 2(z^{t+1} - y^{t+1}).
\end{cases}
$$

For PR splitting

- Does not converge in general even if D is convex.
- Modifying as follows also cannot guarantee convergence even if both sets are convex:

$$\begin{cases} y^{t+1} = \arg\min\limits_{y} \left\{ \frac{1}{2} d_C^2(y) + \frac{1}{2\gamma} \|y - x^t\|^2 \right\}, \\ z^{t+1} \in \operatorname{Arg}\min\limits_{z \in D} \left\{ \|2y^{t+1} - x^t - z\|^2 \right\}, \\ x^{t+1} = x^t + 2(z^{t+1} - y^{t+1}). \end{cases}$$

- Indeed, PR splitting applied to minimizing sum of convex functions $f + g$ converges when f is continuous and strictly convex. (Lions, Mercier '79)

Our PR splitting

- PR splitting: ($\gamma > 0$)

$$\begin{cases} y^{t+1} = \arg\min\limits_{y} \left\{ \frac{1}{2} d_C^2(y) + \frac{5}{2}\|y\|^2 + \frac{1}{2\gamma}\|y - x^t\|^2 \right\}, \\ z^{t+1} \in \operatorname*{Arg\,min}\limits_{z \in D} \left\{ -\frac{5}{2}\|z\|^2 + \frac{1}{2\gamma}\|2y^{t+1} - x^t - z\|^2 \right\}, \\ x^{t+1} = x^t + 2(z^{t+1} - y^{t+1}). \end{cases}$$

Our PR splitting

- PR splitting: $(\gamma > 0)$

$$\begin{cases} y^{t+1} = \arg\min\limits_{y} \left\{ \frac{1}{2} d_C^2(y) + \frac{5}{2}\|y\|^2 + \frac{1}{2\gamma}\|y - x^t\|^2 \right\}, \\ z^{t+1} \in \operatorname*{Arg\,min}\limits_{z \in D} \left\{ -\frac{5}{2}\|z\|^2 + \frac{1}{2\gamma}\|2y^{t+1} - x^t - z\|^2 \right\}, \\ x^{t+1} = x^t + 2(z^{t+1} - y^{t+1}). \end{cases}$$

- Closed form updates for $\gamma \in (0, \frac{1}{5})$:

$$y^{t+1} = \frac{1}{6\gamma + 1}\left[x^t + \gamma P_C\left(\frac{x^t}{5\gamma + 1}\right)\right], \quad z^{t+1} \in P_D\left(\frac{2y^{t+1} - x^t}{1 - 5\gamma}\right).$$

PR Convergence result

Fact 5 (Li, P '15): [Global convergence]
Suppose that $0 < \gamma < \frac{1}{12}$, and D is compact.

Then the sequence $\{(y^t, z^t, x^t)\}$ generated from PR splitting is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_{u \in D} \ \frac{1}{2} d_C^2(u),$$

i.e., $0 \in y^* - P_C(y^*) + N_D(y^*)$.

PR Convergence result

Fact 5 (Li, P '15): [Global convergence]
Suppose that $0 < \gamma < \frac{1}{12}$, and D is compact.

Then the sequence $\{(y^t, z^t, x^t)\}$ generated from PR splitting is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_{u \in D} \frac{1}{2} d_C^2(u),$$

i.e., $0 \in y^* - P_C(y^*) + N_D(y^*)$.

If both sets are in addition semi-algebraic, then the whole sequence is convergent.

More general settings

- Splitting methods for $\min_u \; f(u) + g(u)$, $\alpha = 1$ or 2:

$$\begin{cases} y^{t+1} \in \operatorname*{Arg\,min}_y \left\{ f(y) + \frac{1}{2\gamma}\|y - x^t\|^2 \right\}, \\ z^{t+1} \in \operatorname*{Arg\,min}_z \left\{ g(z) + \frac{1}{2\gamma}\|2y^{t+1} - x^t - z\|^2 \right\}, \\ x^{t+1} = x^t + \alpha(z^{t+1} - y^{t+1}). \end{cases}$$

- f has Lipschitz gradient whose continuity modulus is L, g is proper closed; $f + \frac{l}{2}\|\cdot\|^2$ is convex.

Convergence results for DR: $\alpha = 1$

Fact 6 (Li, P '14): [Global convergence]
Suppose that $(1+\gamma L)^2 + \frac{5\gamma l}{2} - \frac{3}{2} < 0$, f and g are bounded below and at least one of them is coercive.
Then the sequence generated is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_u \; f(u) + g(u),$$

i.e., $0 \in \nabla f(y^*) + \partial g(y^*)$.

Convergence results for DR: $\alpha = 1$

Fact 6 (Li, P '14): [Global convergence]
Suppose that $(1+\gamma L)^2 + \frac{5\gamma l}{2} - \frac{3}{2} < 0$, f and g are bounded below and at least one of them is coercive.
Then the sequence generated is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_u \; f(u) + g(u),$$

i.e., $0 \in \nabla f(y^*) + \partial g(y^*)$.
If, in addition, f and g are semi-algebraic, then the whole sequence is convergent.

Convergence proof?

- KEY: Makes use of

$$\mathfrak{D}_\gamma(y,z,x) := f(y) + g(z) + \frac{1}{2\gamma}\|x-y\|^2 - \frac{1}{2\gamma}\|x-z\|^2.$$

- Can show that for some $k_1, k_2 > 0$:

$$\begin{aligned}\mathfrak{D}_\gamma(y^t,z^t,x^t) - \mathfrak{D}_\gamma(y^{t+1},z^{t+1},x^{t+1}) &\geq k_1\|y^{t+1}-y^t\|^2;\\ \mathrm{dist}(0,\partial\mathfrak{D}_\gamma(y^t,z^t,x^t)) &\leq k_2\|y^{t+1}-y^t\|.\end{aligned}$$

Convergence results for PR: $\alpha = 2$

Fact 7 (Li, P '15): [Global convergence]
Suppose that f is strongly convex with strong convexity modulus $\sigma > 0$ and $3\sigma > 2L$, g is bounded below, and $\gamma \in (0, \frac{3\sigma - 2L}{L^2})$.
Then the sequence generated is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_u \; f(u) + g(u),$$

i.e., $0 \in \nabla f(y^*) + \partial g(y^*)$.

Convergence results for PR: $\alpha = 2$

Fact 7 (Li, P '15): [Global convergence]
Suppose that f is strongly convex with strong convexity modulus $\sigma > 0$ and $3\sigma > 2L$, g is bounded below, and $\gamma \in (0, \frac{3\sigma - 2L}{L^2})$.
Then the sequence generated is bounded, and any cluster point (y^*, z^*, x^*) satisfies $z^* = y^*$. Moreover, y^* is a stationary point of

$$\min_u \; f(u) + g(u),$$

i.e., $0 \in \nabla f(y^*) + \partial g(y^*)$.
If, in addition, f and g are semi-algebraic, then the whole sequence is convergent.

Numerical simulations

- Find a point in $Ax = b$ with $\|x\|_0 \leq r$ and $\|x\|_\infty \leq 10^6$.
- Consider random instances: generate an r-sparse vector $\tilde{x}$, an $m \times n$ matrix A, and set $b = A\tilde{x}$.
- Compare with alternating projection. Initialize all three algorithms at $x^0 = 0$.
- Terminate when successive changes are less than 10^{-8}.

Numerical simulations

- Find a point in $Ax = b$ with $\|x\|_0 \le r$ and $\|x\|_\infty \le 10^6$.
- Consider random instances: generate an r-sparse vector $\tilde{x}$, an $m \times n$ matrix A, and set $b = A\tilde{x}$.
- Compare with alternating projection. Initialize all three algorithms at $x^0 = 0$.
- Terminate when successive changes are less than 10^{-8}.
- For the splitting methods, start with a γ larger than the threshold, decrease γ if $\|y^{t+1} - y^t\|$ does not deteriorate quickly enough or $\|y^t\|$ becomes too large.

Numerical simulations

Over 50 trials for each m, n; sparsity is $\lceil \frac{m}{5} \rceil$; succ means $\mathrm{fval} < 10^{-12}$.

Data	DR: $fval = \frac{1}{2}d_C^2(z^t)$			PR: $fval = \frac{1}{2}d_C^2(z^t)$			Alt Proj: $fval = \frac{1}{2}d_C^2(x^t)$		
m, n	iter	$\mathrm{fval}_{\max}$	succ	iter	$\mathrm{fval}_{\max}$	succ	iter	$\mathrm{fval}_{\max}$	succ
100, 4000	1967	3e-02	30	491	7e-2	0	1694	8e-2	0
100, 5000	2599	2e-02	18	586	7e-2	0	1978	7e-2	0
100, 6000	2046	1e-02	12	684	5e-2	0	2350	5e-2	0
200, 4000	836	2e-15	50	310	2e-1	14	1076	3e-1	0
200, 5000	1080	3e-15	50	364	1e-1	2	1223	2e-1	0
200, 6000	1279	7e-02	43	431	1e-1	5	1510	2e-1	1
300, 4000	600	3e-15	50	223	2e-1	35	872	4e-1	3
300, 5000	710	4e-15	50	295	2e-1	25	1068	3e-1	3
300, 6000	812	3e-15	50	350	2e-1	21	1252	3e-1	1
400, 4000	520	2e-15	50	156	3e-1	47	818	6e-1	30
400, 5000	579	3e-15	50	213	3e-1	42	946	4e-1	12
400, 6000	646	4e-15	50	288	2e-1	38	1108	3e-1	4

Open questions

- Can the smoothness assumption on f be relaxed?
- Can line-search be incorporated?

Open questions

- Can the smoothness assumption on f be relaxed?
- Can line-search be incorporated? Why line search?

Open questions

- Can the smoothness assumption on f be relaxed?
- Can line-search be incorporated? Why line search?

$\mathrm{AP} + \mathrm{ls}$: Alternating projection with nonmonotone line search using Barzalai-Borwein stepsize.

Data		DR: $fval = \frac{1}{2}d_C^2(z^t)$			AP + ls: $fval = \frac{1}{2}d_C^2(x^t)$		
m	n	iter	$\mathrm{fval}_{\max}$	succ	iter	$\mathrm{fval}_{\max}$	succ
100	4000	1967	3e-02	30	81	2e-02	8
100	5000	2599	2e-02	18	78	2e-02	5
200	4000	836	2e-15	50	75	6e-02	28
200	5000	1080	3e-15	50	86	7e-02	15
300	4000	600	3e-15	50	60	7e-02	48
300	5000	710	4e-15	50	67	1e-01	36

Conclusion

- The DR splitting applied to $\min_{u \in D} \frac{1}{2} d_C^2(u)$ with a compact C or D generates a sequence that clusters at a stationary point.
- The PR splitting *suitably* applied to $\min_{u \in D} \frac{1}{2} d_C^2(u)$ with a compact D generates a sequence that clusters at a stationary point.
- Under semi-algebraicity assumption, the whole sequence converges.

Reference:

- G. Li and T. K. Pong.
  Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems.
  Available at *http://arxiv.org/abs/1409.8444*.
- G. Li and T. K. Pong.
  Peaceman-Rachford splitting for a class of nonconvex optimization problems.
  Available at *http://arxiv.org/abs/1507.00887*.

Thanks for coming! $\because$