Douglas-Rachford splitting for nonconvex feasibility problems

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

Workshop on Optimization and Data Analytics
May 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:

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

- Douglas-Rachford (DR) splitting:

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

When D is convex

- Alternating projection:

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

- Douglas-Rachford (DR) splitting:

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

Empirically, DR splitting is usually faster.

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 approach

- 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 approach

- 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 $\frac{1}{2} d_C^2 + \delta_D$.

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 $\{(y^t, z^t, x^t)\}$ 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., $0 \in y^* - P_C(y^*) + N_D(y^*)$.

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 $\{(y^t, z^t, x^t)\}$ 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.

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 $\{(y^t, z^t, x^t)\}$ 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,

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

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

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

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

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

Our DR vs classical DR

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

More general settings

- Douglas-Rachford splitting for $\min_u \; f(u) + g(u)$:

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

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

General convergence results

Fact 4 (Li, P '14): [Global convergence]
Suppose that $(1+\gamma L)^2 + \frac{5\gamma l}{2} - \frac{3}{2} < 0$ and a cluster point of $\{(y^t, z^t, x^t)\}$ exists.
Then 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.

General convergence results

Fact 4 (Li, P '14): [Global convergence]
Suppose that $(1+\gamma L)^2 + \frac{5\gamma l}{2} - \frac{3}{2} < 0$ and a cluster point of $\{(y^t, z^t, x^t)\}$ exists.
Then 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.

Boundedness of sequence

Fact 5 (Li, P '15): [Boundedness]
Suppose that $(1+\gamma L)^2 + \frac{5\gamma l}{2} - \frac{3}{2} < 0$.

Suppose in addition

- both functions f and g are bounded below; and
- at least one of them is coercive.

Then the sequence $\{(y^t, z^t, x^t)\}$ is bounded.

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 $k_1, k_2 > 0$:

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

Numerical simulations

- Find a point in $Ax = b$ with $\|x\|_0 \leq r$, $\|x\|_\infty \leq 10^6$.
- 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 both algorithms at $x^0 = 0$.
- Terminate when successive changes are less than 10^{-8}.

Numerical simulations

- Find a point in $Ax = b$ with $\|x\|_0 \leq r$, $\|x\|_\infty \leq 10^6$.
- 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 both algorithms at $x^0 = 0$.
- Terminate when successive changes are less than 10^{-8}.
- For DR splitting, start with a $\gamma > \sqrt{\frac{3}{2}} - 1$, decrease γ if $\|y^t\|$ gets too large or $\|y^{t+1} - y^t\|$ does not deteriorate quickly enough.

Numerical simulations

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

Data		DR: $fval = \frac{1}{2}d_C^2(z^t)$				Alt Proj: $fval = \frac{1}{2}d_C^2(x^t)$			
m	n	iter	$\mathrm{fval}_{\max}$	$\mathrm{fval}_{\min}$	succ	iter	$\mathrm{fval}_{\max}$	$\mathrm{fval}_{\min}$	succ
100	4000	1967	3e-02	6e-17	30	1694	8e-02	4e-03	0
100	5000	2599	2e-02	2e-16	18	1978	7e-02	5e-03	0
100	6000	2046	1e-02	1e-16	12	2350	5e-02	4e-05	0
200	4000	836	2e-15	2e-16	50	1076	3e-01	3e-05	0
200	5000	1080	3e-15	2e-16	50	1223	2e-01	2e-03	0
200	6000	1279	7e-02	1e-16	43	1510	2e-01	1e-13	1
300	4000	600	3e-15	2e-16	50	872	4e-01	6e-14	3
300	5000	710	4e-15	4e-16	50	1068	3e-01	9e-14	3
300	6000	812	3e-15	2e-16	50	1252	3e-01	1e-13	1
400	4000	520	2e-15	3e-17	50	818	6e-01	7e-14	30
400	5000	579	3e-15	5e-16	50	946	4e-01	9e-14	12
400	6000	646	4e-15	6e-16	50	1108	3e-01	1e-13	4

Conclusion

- The DR splitting applied to $\min_{u \in D} \frac{1}{2} d_C^2(u)$, with either C or D being compact, can be shown to generate a bounded sequence that clusters at a stationary point.
- Under semi-algebraicity assumption, the whole sequence can be shown to be convergent.

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*.

Thanks for coming!