Douglas-Rachford Splitting for Infeasible, Unbounded, and Pathological Problems

Yanli Liu, Ernest Ryu, Wotao Yin
UCLA Math

US-Mexico Workshop Optimization and its Applications — Jan 8–12, 2018
Background
What is “splitting”? 

- Sun-Tzu: “远交近攻”，“各个击破” (400 BC)
- Caesar: “divide-n-conquer” (100–44 BC)
- Principle of computing: reduce a problem to simpler subproblems
- Example: find $x \in C_1 \cap C_2 \rightarrow$ project to $C_1$ and $C_2$ alternatively
Basic principles of splitting

split:

- x/y directions
- linear from nonlinear
- smooth from nonsmooth
- spectral from spatial
- convection from diffusion
- composite operators
  - \((I - \lambda(A + B))^{-1}\) to \((I - \lambda A)^{-1}\) and \((I - \lambda B)^{-1}\)

Also

- domain decomposition
- block-coordinate descent
- column generation, Bender’s decomposition, etc.
Operator splitting pipeline

1. Formulate

\[ 0 \in A(x) + B(x) \]

where \( A \) and \( B \) are operators, possibly set-valued

2. **operator splitting**: get a fixed-point operator \( T \):

\[ z^{k+1} \leftarrow Tz^k \]

Applying \( T \) reduces to computing \( A \) and \( B \) successively

3. Correctness and convergence:

- fixed-point \( z^* = Tz^* \) recovers a solution \( x^* \)
- \( T \) is **contractive** or, more weakly, **averaged**
Example: constrained minimization

- $C$ is a convex set. $f$ is a differentiable convex function.

\[
\minimize_x f(x)
\]

subject to $x \in C$

- equivalent inclusion problem:

\[
0 \in N_C(x) + \nabla f(x)
\]

$N_C$ is the normal cone

- projected gradient method:

\[
x^{k+1} \leftarrow \proj_C \circ (I - \gamma \nabla f) x^k
\]
Convergence
**Contractive operator**

- **definition:** $T$ is contractive if, for some $L \in [0, 1)$,

$$
\|Tx - Ty\| \leq L\|x - y\|, \quad \forall x, y
$$
Between $L = 1$ and $L < 1$

- $L < 1 \implies$ geometric convergence
- $L = 1 \implies$ iterates are bounded, but may diverge
- Some algorithms have $L = 1$ and still converge:
  - Alternative projection (von Neumann)
  - Gradient descent
  - Proximal-point algorithm
  - Operator splitting algorithms
Averaged operator

- **residual operator:** $R := I - T$. Hence, $Rx^* = 0 \iff x^* = Tx^*$

- **averaged operator:** from some $\eta > 0$,
  \[
  \|Tx - Ty\|^2 \leq \|x - y\|^2 - \eta \|Rx - Ry\|^2, \quad \forall x, y
  \]

- **interpretation:** set $y$ as a fixed point, then distance to $y$ improve by the amount of fixed-point residual

- **property**: if $T$ has a fixed point, then $x^{k+1} \leftarrow Tx^k$ converges weakly to a fixed point

---

1. Krasnosel'skii’57, Mann’56
Why called “averaged”?

**Lemma:** For \( \alpha \in (0, 1) \), \( T \) is \( \alpha \)-averaged if, and only if, there exists a nonexpansive (1-Lipschitz) map \( T' \) so that

\[
T = (1 - \alpha)I + \alpha T'.
\]
Composition of averaged operators

Useful theorem:

\[ T_1, T_2 \text{ nonexpansive} \implies T_1 \circ T_2 \text{ nonexpansive} \]

\[ T_1, T_2 \text{ averaged} \implies T_1 \circ T_2 \text{ averaged} \]

(though the averagedness constants get worse.)
How to get an averaged-operator composition?
Forward-backward splitting

- derive:

\[ 0 \in Ax + Bx \iff x - Bx \in x + Ax \]
\[ \iff (I - B)x \in (I + A)x \]
\[ \iff (I + A)^{-1} (I - B) x = x \]
\[ \text{backward forward operator } T_{\text{FBS}} \]

- Although \((I + A)\) may be set-valued, \((I + A)^{-1}\) is single-valued!
• forward-backward splitting (FBS) operator (Mercier’79): for $\gamma > 0$

$$T_{\text{FBS}} := (I + \gamma A)^{-1} \circ (I - \gamma B)$$

• key properties:
  • if $A$ is maximally monotone$^2$, then $(I + \gamma A)^{-1}$ is $\frac{1}{2}$-averaged
  • if $B$ is $\beta$-cocoercive$^3$ and $\gamma \in (0, 2\beta)$, then $(I - \gamma B)$ is averaged

• conclusion: $T_{\text{FBS}}$ is averaged, thus if a fixed-point exists,

$$x^{k+1} \leftarrow T_{\text{FBS}}(x^k)$$

converges

---

$^2 \langle Ax - Ay, x - y \rangle \geq 0, \forall x, y$

$^3 \langle Bx - By, x - y \rangle \geq \beta \|Bx - By\|^2, \forall x, y$
Major operator splitting schemes

0 ∈ Ax + Bx

- **forward-backward** (Mercier’79) for (maximally monotone) + (cocoercive)
- **Douglas-Rachford** (Lion-Mercier’79) for (maximally monotone) + (maximally monotone)
- **forward-backward-forward** (Tseng’00) for (maximally monotone) + (Lipschitz & monotone)
- **three-operator** (Davis-Yin’15) for (maximally monotone) + (maximally monotone) + (cocoercive)
- use non-Euclidean metric (Condat-Vu’13) for (maximally monotone o A)
  A is bounded linear operator
DRS for optimization

\[
\begin{align*}
\text{minimize } & f(x) + g(x) \\
\end{align*}
\]

- \( f, g \) are proper closed convex, may be non-differentiable
- DRS iteration: \( z^{k+1} = T_{\text{DRS}}(z^k) \iff \)
  \[
  x^{k+1/2} = \text{prox}_{\gamma f}(z^k) \\
x^{k+1} = \text{prox}_{\gamma g}(2x^{k+1/2} - z^k) \\
z^{k+1} = z^k + (x^{k+1} - x^{k+1/2})
  \]

- \( z^k \to z^* \) and \( x^k, x^{k+1/2} \to x^* \) if
  - primal dual solutions exist, and
  - \( -\infty < p^* = d^* < \infty \).

- otherwise, \( \|z^k\| \to \infty \).
New results
Overview

- **pathological conic programs**, even small ones, can cripple existing solvers

- **proposed**: use DRS
  - to identify infeasible, unbounded, pathological problems
  - to compute “certificates” if there is one
  - to “restore feasibility”

- **under the hood**: understanding divergent DRS iterates
Linear programming

- **standard-form:**
  \[
  p^* = \min c^T x \quad \text{subject to} \begin{cases} Ax = b, \quad x \geq 0 \\ x \in \mathcal{L} \end{cases}
  \]

- every LP is in exactly one of the 3 cases:
  1) \( p^* \) finite \( \iff \exists \) primal solution \( \iff \exists \) primal-dual solution pair
  2) \( p^* = -\infty \): problem is feasible, unbounded \( \iff \exists \) improving direction\(^4\)
  3) \( p^* = +\infty \): problem is infeasible \( \iff \text{dist}(L, \mathbb{R}^+) > 0 \iff \exists \) strict separating hyperplane\(^5\)

- cases (2) (3) arise, e.g., during branch-n-bound

- existing solvers are reliable

\(^4\) \( u \) is an *improving direction* if \( c^T u < 0 \) and \( x + \alpha u \) is feasible for all feasible \( x \) and \( \alpha > 0 \).

\(^5\) \{ \( x : h^T x = \beta \} \) strictly separates two sets \( L \) and \( K \) if \( h^T x < \beta < h^T y \) for all \( x \in \mathcal{L}, y \in K \).
Conic programming

- **standard-form**: \( K \) is a closed convex cone

\[
p^* = \min c^T x \quad \text{subject to } Ax = b, \ x \in K
\]

- every problem is in one of the 7 cases:
  1) \( p^* \) finite: 1a) has PD sol pair, 1b) has P sol only, 1c) no P sol
  2) \( p^* = -\infty \): 2a) has improving direction, 2b) no improving direction
  3) \( p^* = +\infty \): 3a) dist\((\mathcal{L}, K)\) > 0 ⇔ has strict separating hyperplane
     3b) dist\((\mathcal{L}, K)\) = 0 ⇔ no strict separating hyperplane

- all “b” “c” cases are pathological

- even nearly pathological problems can fail existing solvers
Example 1

- 3-variable problem:

  $$\text{minimize } x_1 \text{ subject to } x_2 = 1, \ 2x_2x_3 \geq x_1^2, \ x_2, x_3 \geq 0.$$  

- belongs to case 2b):
  - feasible
  - $$p^* = -\infty$$, by letting $$x_3 \to \infty$$ and $$x_1 \to -\infty$$
  - no improving direction\(^6\)

- existing solvers\(^7\):
  - SDPT3: “Failed”, $$p^*$$ no reported
  - SeDuMi: “Inaccurate/Solved”, $$p^* = -175514$$
  - Mosek: “Inaccurate/Unbounded”, $$p^* = -\infty$$

---

\(^6\text{reason: any improving direction } u \text{ has form } (u_1, 0, u_3), \text{ but by the cone constraint } 2u_2u_3 = 0 \geq u_1^2, \text{ so } u_1 = 0, \text{ which implies } c^T u_1 = 0 \text{ (not improving).}\)

\(^7\text{using their default settings}\)
Example 2

- 3-variable problem:

\[
\begin{array}{c}
\text{minimize } 0 \quad \text{subject to } \\
\begin{bmatrix}
0 & 1 & 1 \\
1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix} = \\
\begin{bmatrix}
0 \\
1
\end{bmatrix}, \quad \begin{bmatrix}
x_3 \geq \sqrt{x_1^2 + x_2^2} \\
x_3 \in K
\end{bmatrix}
\end{array}
\]

- belongs to case 3b):
  - infeasible\(^8\)
  - \(\text{dist}(L, K) = 0\) \(^9\)
  - no strict separating hyperplane

- existing solvers\(^{10}\):
  - SDPT3: “Infeasible”, \(p^* = \infty\)
  - SeDuMi: “Solved”, \(p^* = 0\)
  - Mosek: “Failed”, \(p^*\) not reported

---

\(^8\) \(x \in L\) imply \(x = [1, -\alpha, \alpha]^T, \alpha \in \mathbb{R}\), which always violates the second-order cone constraint.

\(^9\) \(\text{dist}(L, K) \leq \|[1, -\alpha, \alpha] - [1, -\alpha, (\alpha^2 + 1)^{1/2}]\|_2 \to \infty\) as \(\alpha \to \infty\).

\(^{10}\) using their default settings
Conic DRS

\[
\begin{align*}
\text{minimize} & \quad c^T x \quad \text{subject to} \quad Ax = b, \ x \in K \\
\iff \text{minimize} & \quad \left( c^T x + \delta_{A \cdot = b}(x) \right) + \delta_K(x) \\
\end{align*}
\]

- cone \( K \) is nonempty closed convex\(^\text{11} \), matrix \( A \) has full row rank
- each iteration: projection onto \( A \cdot = b \), then projection onto \( K \)
- per-iteration cost: \( O(n^2 + \text{cost(\text{proj}_K)}) \) with prefactorized \( AA^T \)
- prior work: Wen-Goldfarb-Yin’09 for SDP
- we know: if not case 1a), DRS diverges; but how?

\(^{11}\) not necessarily self-dual
What happens during divergence?

- **iteration**: \( z^{k+1} = T(z^k) \), where \( T \) is averaged

- **general theorem\(^{12}\)**: \( z^k - z^{k+1} \rightarrow v = \text{Proj}_{\text{ran}(I-T)}(0) \)

- \( v \) is “the best approximation to a fixed point of \( T \)”

---

\(^{12}\)Pazy’71, Baillon-Bruck-Reich’78
Our results (Liu-Ryu-Yin’17)

- proof simplification

- new rate of convergence: \[ \| z^k - z^{k+1} \| \leq \| v \| + \epsilon + O \left( \frac{1}{\sqrt{k+1}} \right) \]

- for conic programs, a workflow using three simultaneous DRS:
  1) original DRS
  2) same DRS with \( c = 0 \)
  3) same DRS with \( b = 0 \)

- most pathological cases are **identified**

- for unbounded problem 2a), **compute an improving direction**

- for infeasible problem 3a), **compute a strict separating hyperplane**

- for all infeasible problems, minimally alter \( b \) to **restore strong feasibility**
Decision flow

Start

- **Thm 6 Alg 2**
  - Infeasible
  - **Thm 7 Alg 2**
    - (f) Strongly infeasible
    - (g) Weakly infeasible

- Feasible
  - **Thm 2 Alg 1**
    - (a) There is a primal-dual solution pair with $d^* = p^*$

  - **Thm 13 Alg 1**
    - (b) There is a primal solution but no dual solution or $d^* < p^*$
  - **Thm 11,12 Alg 3**
    - (c) $p^*$ is finite but there is no solution

  - **Thm 10 Alg 3**
    - (d) Unbounded ($p^* = -\infty$) with an improving direction
    - (e) Unbounded ($p^* = -\infty$) without an improving direction
Theorems

- Identifications are described in a series of theorems in the form
  
  Run DRS (one of three). If \( \lim_k z^k \rightarrow z^{k+1} = v \ldots, \|z^k\| \ldots, \) or
  
  \( \|z^{k+1} - z^k\| \ldots, \) then the problem is in case \ldots and \ldots

- example: Theorem 7. Run Alg2. Let \( z^k \rightarrow z^{k+1} \rightarrow v. \) Problem is 3a) if and only if \( v \neq 0. \) If \( v \neq 0, \) we have the strict separating hyperplane:

  \[
  \{ x : v^T x = (v^T x_0)/2 \}.
  \]

- example: Theorem 10: If feasible, run Alg3. Let \( z^k \rightarrow z^{k+1} \rightarrow d. \) Problem is 2a) if and only if \( d \neq 0. \) If \( d \neq 0, \) then it is an improving direction.
Weakly infeasible SDP set (Liu-Pataki’17)

<table>
<thead>
<tr>
<th></th>
<th>$m = 10$</th>
<th></th>
<th>$m = 20$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clean</td>
<td>Messy</td>
<td>Clean</td>
<td>Messy</td>
</tr>
<tr>
<td>SeDuMi</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>SDPT3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mosek</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>$\text{PP}^{13} + \text{SeDuMi}$</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
</tr>
</tbody>
</table>

percentage of success detection on clean and messy examples in Liu-Pataki’17

$^{13}$PreProcessing by Permenter-Parilo’14
Weakly infeasible SDP set (Liu-Pataki’17)

<table>
<thead>
<tr>
<th></th>
<th>$m = 10$</th>
<th></th>
<th>$m = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clean</td>
<td>Messy</td>
<td>Clean</td>
</tr>
<tr>
<td>Proposed</td>
<td>100</td>
<td>21</td>
<td>100</td>
</tr>
</tbody>
</table>

(stopping: $\|z^{1e7}\|_2 \geq 800$)

*our percentage is way much better!*
Strongly infeasible SDP set (Liu-Pataki’17)

<table>
<thead>
<tr>
<th></th>
<th>$m = 10$</th>
<th></th>
<th>$m = 20$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>100</td>
<td>Clean</td>
<td>100</td>
</tr>
<tr>
<td>Messy</td>
<td>100</td>
<td>Messy</td>
<td>100</td>
</tr>
</tbody>
</table>

Proposed: \[ \| z^{5e4} - z^{5e4+1} \|_2 \leq 10^{-3} \]

our percentage is way much better!
Other approaches

- **homogeneous self-dual embedding**\(^{14}\):
  - is a reformulation that is always feasible and can produce PD solutions
  - can use facial reductions to identify “b” “c”

- **facial reduction**\(^{15}\):
  - generates bigger but less pathological problems
  - can theoretically identify all cases
  - no efficient numerical implementation yet
    - reduction is not cheap, also introduces new computational issues
    - generate cones that are intersections of original cones with linear subspaces, making IPM and DRS difficult to apply

---

\(^{14}\) Ye’11, Luo-Sturm-Zhang’00, Skajaa’Ye’12, etc.

\(^{15}\) Methods: Borwein, Muramatsu, Pataki, Waki, Wolkowicz; numerical approaches: Lourenco-Muramatsu-Tsuchiya’15, Permenter-Friberg-Andersen’15
Related work

Bauschke, Combettes, Hare, Luke, Moursi, and others recently did

- DRS for feasibility between two convex sets by
- Range of DRS and generalized solutions to $0 \in A + B$ where $A, B$ are maximally monotone
- Also, Moursi’s thesis on DRS in the possibly inconsistent case: Static properties and dynamic behaviour
summary:

- DRS iterates provide useful information even when they diverge
- easy to code it for conic programs

not covered:

- general convex problem $f(x) + g(x)$
- analysis of $f(x_{k+1/2}^k) + g(x_{k+1}^k)$
- adaptation to ADMM

acknowledgements: NSF