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 $x \in C_1 \cap C_2 \longrightarrow$ 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.

$$\begin{aligned} &\underset{x}{\text{minimize}} \;\; f(x) \\ &\text{subject to } x \in C \end{aligned}$$

- **equivalent inclusion problem:**

$$0 \in N_C(x) + \nabla f(x)$$

  N_C is the normal cone

- **projected gradient method**:

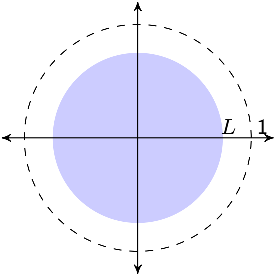
$$x^{k+1} \leftarrow \underbrace{\mathbf{proj}_C \circ (I - \gamma \nabla f)}_{T} x^k$$

Convergence

Contractive operator

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

$$\|Tx - Ty\| \le L\|x - y\|, \quad \forall x, y$$



Between $L = 1$ and $L < 1$

- $L < 1 \;\Rightarrow$ geometric convergence
- $L = 1 \;\Rightarrow$ 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, $Rx^* = 0 \;\Leftrightarrow\; x^* = Tx^*$
- **averaged operator:** from some $\eta > 0$,

$$\|Tx - Ty\|^2 \le \|x - y\|^2 - \eta\|Rx - Ry\|^2, \quad \forall x, y$$

- **interpretation:** set y as a fixed point, then distance to y improve by the amount of fixed-point residual
- **property**[1]**:** if T has a fixed point, then $x^{k+1} \leftarrow Tx^k$ converges weakly to a fixed point

[1] Krasnosel'skiĭ'57, Mann'56

Why called "averaged"?

lemma: For $\alpha \in (0,1)$, T is α-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:

$$T_1, T_2 \text{ nonexpansive} \Rightarrow\; T_1 \circ T_2 \text{ nonexpansive}$$

$$T_1, T_2 \text{ averaged} \Rightarrow\; T_1 \circ T_2 \text{ averaged}$$

(though the averagedness constants get worse.)

How to get an averaged-operator composition?

Forward-backward splitting

- **derive**:

$$\begin{aligned} 0 \in Ax + Bx &\iff x - Bx \in x + Ax \\ &\iff (I - B)x \in (I + A)x \\ &\iff \underbrace{\underbrace{(I + A)^{-1}}_{\text{backward}} \underbrace{(I - B)}_{\text{forward}}}_{\text{operator } T_{\mathrm{FBS}}} x = x \end{aligned}$$

- Although $(I + A)$ may be set-valued, $(I + A)^{-1}$ is single-valued!

- **forward-backward splitting (FBS) operator** (Mercier'79): for $\gamma > 0$

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

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

- **conclusion:** T_{FBS} is averaged, thus if a fixed-point exists,

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

  converges

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

[3] $\langle Bx - By, x - y\rangle \geq \beta\|Bx - By\|^2, \quad \forall x, y$

Major operator splitting schemes

$$0 \in Ax + Bx$$

- **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 $\circ A$)

  A is bounded linear operator

DRS for optimization

$$\underset{x}{\text{minimize}}\, f(x) + g(x)$$

- f, g are proper closed convex, may be non-differentiable
- DRS iteration: $z^{k+1} = T_{\mathrm{DRS}}(z^k) \iff$

$$\begin{aligned} x^{k+1/2} &= \mathbf{prox}_{\gamma f}(z^k) \\ x^{k+1} &= \mathbf{prox}_{\gamma g}(2x^{k+1/2} - z^k) \\ z^{k+1} &= z^k + (x^{k+1} - x^{k+1/2}) \end{aligned}$$

- $z^k \to z^*$ 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:**

$$p^\star = \min \ c^T x \quad \text{subject to} \ \underbrace{Ax = b}_{x \in \mathcal{L}}, \ \underbrace{x \geq 0}_{x \in \mathbb{R}+}$$

- every LP is in exactly one of the 3 cases:
  1) $p^\star$ finite $\Leftrightarrow \exists$ primal solution $\Leftrightarrow \exists$ primal-dual solution pair
  2) $p^\star = -\infty$: problem is feasible, unbounded $\Leftrightarrow \exists$ improving direction[4]
  3) $p^\star = +\infty$: problem is infeasible $\Leftrightarrow \mathrm{dist}(L, \mathbb{R}^+) > 0 \Leftrightarrow \exists$ 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

$$p^\star = \min\ c^T x \quad \text{subject to } \underbrace{Ax = b}_{x \in \mathcal{L}},\ x \in K$$

- every problem is in one of the **7 cases**:

  1) $p^\star$ finite: 1a) has PD sol pair, 1b) has P sol only, 1c) no P sol

  2) $p^\star = -\infty$: 2a) has improving direction, 2b) no improving direction

  3) $p^\star = +\infty$: 3a) $\operatorname{dist}(\mathcal{L}, K) > 0 \Leftrightarrow$ has strict separating hyperplane
  3b) $\operatorname{dist}(\mathcal{L}, K) = 0 \Leftrightarrow$ no strict separating hyperplane

- all "b" "c" cases are pathological

- even nearly pathological problems can fail existing solvers

Example 1

- **3-variable problem:**

$$\text{minimize } x_1 \quad \text{subject to } x_2 = 1,\ \underbrace{2x_2x_3 \geq x_1^2,\ x_2, x_3 \geq 0}_{\text{rotated second-order cone}}.$$

- belongs to **case 2b):**
  - feasible
  - $p^\star = -\infty$, by letting $x_3 \to \infty$ and $x_1 \to -\infty$
  - no improving direction[6]
- **existing solvers**[7]**:**
  - SDPT3: "Failed", $p^\star$ no reported
  - SeDuMi: "Inaccurate/Solved", $p^\star = -175514$
  - Mosek: "Inaccurate/Unbounded", $p^\star = -\infty$

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

[7] using their default settings

Example 2

- **3-variable problem:**

$$\text{minimize } 0 \quad \text{subject to } \underbrace{\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} x = \begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{x \in \mathcal{L}}, \ \underbrace{x_3 \geq \sqrt{x_1^2 + x_2^2}}_{x \in K}.$$

- belongs to **case 3b):**
  - infeasible[8]
  - $\mathrm{dist}(\mathcal{L}, K) = 0$ [9]
  - no strict separating hyperplane

- **existing solvers[10]:**
  - SDPT3: "Infeasible", $p^\star = \infty$
  - SeDuMi: "Solved", $p^\star = 0$
  - Mosek: "Failed", $p^\star$ not reported

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

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

[10] using their default settings

Conic DRS

$$\begin{aligned} &\text{minimize } c^T x \quad \text{subject to } Ax = b,\ x \in K \\ \Leftrightarrow\quad &\text{minimize } \underbrace{\left(c^T x + \delta_{A\cdot=b}(x)\right)}_{f(x)} + \underbrace{\delta_K(x)}_{g(x)} \end{aligned}$$

- cone K is nonempty closed convex[11], matrix A has full row rank
- each iteration: projection onto $A\cdot = b$, then projection onto K
- per-iteration cost: $O(n^2 + \text{cost}(\mathbf{proj}_K))$ 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?

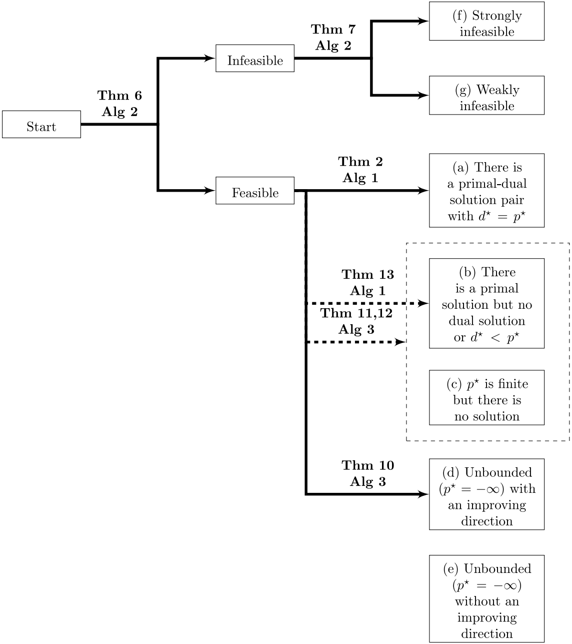
- **iteration:** $z^{k+1} = T(z^k)$, where T is averaged
- **general theorem**[12]**:** $z^k - z^{k+1} \to v = \mathrm{Proj}_{\overline{\mathrm{ran}(I-T)}}(\mathbf{0})$
- v is "the best approximation to a fixed point of T"

[12] Pazy'71, Baillon-Bruck-Reich'78

Our results (Liu-Ryu-Yin'17)

- **proof simplification**
- **new rate of convergence:** $\|z^k - z^{k+1}\| \le \|v\| + \epsilon + O(\frac{1}{\sqrt{k+1}})$
- for conic programs, **a workflow** using three simultaneous DRS:
  1) original DRS
  2) same DRS with $c = \mathbf{0}$
  3) same DRS with $b = \mathbf{0}$
- 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



Theorems

- Identifications are described in a series of theorems in the form

  Run DRS (one of three). If $\lim_k z^k - z^{k+1} = v$ *..., $\|z^k\|$ ..., or $\|z^{k+1} - z^k\|$ ..., then the problem is in case ... and ...*

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

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

- **example:** Theorem 10: If feasible, run Alg3. Let $z^k - z^{k+1} \to d$. Problem is 2a) if and only if $d \neq \mathbf{0}$. If $d \neq \mathbf{0}$, then it is an improving direction.

Weakly infeasible SDP set (Liu-Pataki'17)

	$m = 10$		$m = 20$	
	Clean	Messy	Clean	Messy
SeDuMi	0	0	1	0
SDPT3	0	0	0	0
Mosek	0	0	11	0
PP[13]+SeDuMi	100	0	100	0

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)

	$m = 10$		$m = 20$	
	Clean	Messy	Clean	Messy
Proposed	100	21	100	99

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

our percentage is way much better!

Strongly infeasible SDP set (Liu-Pataki'17)

	$m=10$		$m=20$	
	Clean	Messy	Clean	Messy
Proposed	100	100	100	100

(stopping: $\|z^{5e4} - z^{5e4+1}\|_2 \leq 10^{-3}$)

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 $f(x^{k+1/2}) + g(x^{k+1})$
- adaptation to ADMM

acknowledgements: NSF

report: `https://arxiv.org/abs/1706.02374`