Douglas-Rachford Splitting for Pathological Problems

Ernest Ryu, Yanli Liu, and Wotao Yin (Presenter)

UCLA, Math Department

Int'l Workshop on Modern Optimization and Applications — June 2018

Problem setup

Consider a convex primal problem

$$\text{minimize} \;\; f(x) + g(x) \tag{P}$$

and its dual problem

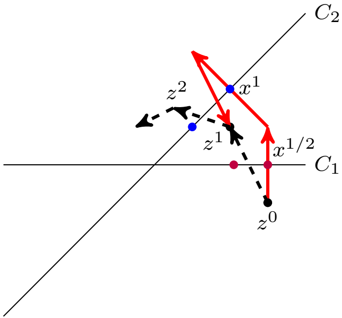
$$\text{maximize} \; -f^*(y) - g^*(-y). \tag{D}$$

f, g can be nonsmooth and indicator functions of constraints.

Douglas-Rachford splitting: "reflect, reflect, average"

Let us find a point in $C_1 \cap C_2$. We apply Douglas-Rachford splitting (DRS) to

$$\underset{x}{\text{minimize}} \;\; \delta_{C_1}(x) + \delta_{C_2}(x)$$



Proximal mapping

To minimize $f(x) + g(x)$, DRS replaces projection by proximal mapping.

The proximal mapping of a function f is defined as

$$\mathrm{prox}_{\gamma f}(x) = \operatorname*{arg\,min}_{v \in \mathbb{R}^n} \left\{ f(v) + \frac{1}{2\gamma} \|v - x\|^2 \right\}$$

When $f = \delta_C$ (indicator of a set C), $\mathrm{prox}_{\gamma f}$ is projection to C.

Douglas-Rachford splitting method[1]

DRS applied to (P):

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

Main sequence: $z^0, z^1, z^2, \ldots$

f is evaluated at the shadow sequence: $x^{1/2}, x^{1+1/2}, x^{2+1/2}, \ldots$

g is evaluated at the shadow sequence: $x^1, \quad x^2, \qquad x^3, \ldots$

[1] Lions-Mercier'79

DRS generalizes Spingarn's, classic ADMM, Prox-ADMM, Chambolle-Pock, and so on. (However, the transforms are not obvious.)

Once we understand the behavior of DRS, we can translate the results to other algorithms in principle.

DRS convergence (classical)

Classical fixed-point analysis:

- If DRS (more generally, a firmly nonexpansive operator) has a fixed point, then z^k converges to a fixed point;
- Otherwise, $\|z^k\| \to \infty$.

When we apply DRS to minimization

$$\begin{aligned} &\text{minimize} \quad f(x) + g(x) && (P) \\ &\text{maximize} \quad -f^*(y) - g^*(-y) && (D) \end{aligned}$$

Here, DRS has a fixed point if, and only if[2]

1. (P) has a solution,
2. (D) has a solution, *and*
3. strong duality holds, i.e., $p^\star = d^\star$.

If 1–3 are all satisfied, DRS converges and returns a primal-dual solution pair.

Otherwise, DRS diverges. In this case, we say (P) is pathological.

[2] Bauschke, Boț, Hare, Moursi'12

DRS convergence (new)[3]

DRS "still works" when

- ~~(P) has a solution,~~
- ~~(D) has a solution, and~~
- strong duality holds, i.e., $p^\star = d^\star$ ($-\infty$ and ∞ are allowed).

[3] Ryu, Liu, and Yin. arXiv:1801.06618.

Extended optimal value (definition)

$$p^\star := \begin{cases} \infty & \text{(P) is infeasible} \\ \inf_x f(x) + g(x) & \text{(P) is feasible, bounded} \\ -\infty & \text{(P) is feasible, unbounded} \end{cases}$$

Define $d^\star$ similarly for (D).

We always have $d^\star \le p^\star$. If "$<$", we lose strong duality.

Example 1

$$\underset{x \in \mathbb{R}^3}{\text{minimize}} \;\; x_1 \quad \text{subject to } x_2 = 1, \;\; \underbrace{2x_2x_3 \geq x_1^2, \;\; x_2, x_3 \geq 0}_{\text{rotated second-order cone } Q}$$

$$\iff \;\; \underset{x \in \mathbb{R}^3}{\text{minimize}} \;\; \underbrace{x_1 + \delta_{x_2=1}(x)}_{f(x)} + \underbrace{\delta_Q(x)}_{g(x)}$$

This problem is feasible and unbounded[4], but has no improving direction[5].

It dual problem is infeasible, so $p^\star = d^\star = -\infty$.

[4] by letting $x_3 \to \infty$ and $x_1 \to -\infty$

[5] **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).

Example 2

The primal problem:

$$\underset{x\in\mathbb{R}}{\text{minimize}} \;\; \underbrace{1/\sqrt{-x}}_{f(x)} \;\; \underbrace{-\log x}_{g(x)}$$

is weakly infeasible since $\mathrm{dom} f = (-\infty, 0]$ and $\mathrm{dom} g = (0, \infty)$.

The dual problem:

$$\underset{y\in\mathbb{R}}{\text{maximize}} \;\; c_1 y^{1/3} + 1 - \log(1/y),$$

where $c_1 > 0$, is feasible and unbounded.

We have $p^\star = d^\star = \infty$.

Example 3

The primal problem

$$\underset{y_1, y_2 \in \mathbb{R}}{\text{minimize}} \;\; \sqrt{y_1^2 + y_2^2} - y_1 + \delta_{y_2 = 1}(-y)$$

is feasible but has no solution to attain $p^\star = 0$.

This dual problem

$$\underset{x_1, x_2 \in \mathbb{R}}{\text{maximize}} \;\; -\delta_{x_1^2 + x_2^2 \leq 1}(x) - x_2 - \delta_{x_1 = 1}(x)$$

is feasible and has a solution $(1, 0)$, which attains $d^\star = p^\star = 0$.

What do we want DRS to do for pathological problems?

DRS generates iterates that are

- asymptotically feasible, if $\mathrm{dist}(\mathrm{dom} f, \mathrm{dom} g) = 0$,
- tracking an improving direction, if one exists,
- asymptotically optimal, if $p^\star = d^\star$.

DRS convergence (new): examples

Theorem

If (P) *is weakly infeasible, then*

$$x^{k+1} - x^{k+1/2} \to \mathbf{0}.$$

Theorem

If (P) *is feasible but does not have a solution and* $p^\star = d^\star \in [-\infty, \infty)$, *then*

$$x^{k+1} - x^{k+1/2} \to \mathbf{0}, \qquad \liminf_{k\to\infty} f(x^{k+1}) + g(x^{k+1/2}) = p^\star.$$

Theorem

If (P) *is feasible and unbounded, then*

$$\liminf_{k\to\infty} f(x^{k+1}) + g(x^{k+1/2}) = p^\star = -\infty.$$

Moreover, if there exists an improving direction d*, then* $x^{k+1} - x^k = d + o(1)$*.*

We can say something for all the pathological cases if $p^\star = d^\star$.

DRS for conic programming

Conic programming

$$
\begin{aligned}
\underset{x}{\text{minimize}}\ \ & c^T x \\
\text{subject to}\ & Ax = b \\
& x \in \underbrace{\text{closed convex cone}}_{K}
\end{aligned}
$$

generalizes many types of convex optimization: LP, convex QP/QCQP, SDP, ...

Conic programming pathologies

Joint constraints $Ax = b$ and $x \in K$ may be infeasible.

The objective can be unbounded $-\infty$.

Even worse, in these cases, a dual certificate and an unbounded direction may not exist. These are called *weak pathologies*.

Applying DRS to conic programming

Applying DRS[6] to

$$\underset{x}{\text{minimize}} \quad \underbrace{c^T x + \delta_{A\cdot=b}(x)}_{f(x)} + \underbrace{\delta_K(x)}_{g(x)}$$

can recognize infeasible, feasible, and unbounded problems.

Applying

$$\text{2nd DRS to } \underset{x}{\text{minimize}} \quad \underbrace{0^T x + \delta_{x:Ax=b}(x)}_{f(x)} + \underbrace{\delta_K(x)}_{g(x)},$$

$$\text{3rd DRS to } \underset{x}{\text{minimize}} \quad \underbrace{c^T x + \delta_{x:Ax=0}(x)}_{f(x)} + \underbrace{\delta_K(x)}_{g(x)},$$

further classifies almost all strong and week pathologies.

[6] Liu, Ryu, Yin, arXiv:1706.02374. Related to Wen, Goldfarb, Yin'2010.

Detection methods[7]: examples

Theorem

Run DRS. If $z^k - z^{k+1} \to v \neq \mathbf{0}$, then the conic program is infeasible and has a strict separating hyperplane

$$\left\{x : v^T x = (v^T x_0)/2\right\},$$

where $x_0 := A^T(AA^T)^{-1}b$.

Theorem

Run DRS and DRS 3. Assume DRS confirms feasibility. In DRS 3, if $z^k - z^{k+1} \to d \neq \mathbf{0}$, then the conic program is unbounded and $d \neq \mathbf{0}$ is an improving direction.

We have a flow to detect almost all cases.

For infeasible problems, we find a minimal change to restore feasibility.

[7] Liu, Ryu, Yin, arXiv:1706.02374

Other detection approaches

Self-dual embedding[8]:

- is a reformulation that is always feasible and can produce PD solutions
- can use facial reductions to identify weak pathologies

Facial reduction[9]:

- generates large but less pathological problems
- theoretically identify all cases
- no efficient numerical implementation yet

[8] Mizuno-Todd-Ye'93, Luo-Sturm-Zhang'99 and '00, Nesterov-Todd-Ye'99, Ye'11, Skajaa'Ye'12, etc.

[9] Methods: Borwein, Muramatsu, Pataki, Waki, Wolkowicz; numerical approaches: Lourenco-Muramatsu-Tsuchiya'15, Permenter-Friberg-Andersen'15

Weakly-infeasible SDP detection test

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

Percentage of success detections reported in Liu-Pataki'17

[10] PreProcessing by Permenter-Parilo'14

Weakly-infeasible SDP detection test

	$m = 10$		$m = 20$	
	Clean	Messy	Clean	Messy
Our triple-DRS	100	21	100	99

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

Our percentage is way much better!

In another strongly-infeasible SDP test, our detection is 100%.

Theoretical components

Prior work

There has been surprisingly little work studying DRS under pathologies.

Results on fixed-point setups:

- Pazy'71, Baillon-Bruck-Reich'78. ...
- Bauschke, Hare, and Moursi, 2014 and 2016

Results in specific pathological setups:

- Bauschke, Combettes, and Luke. Two closed convex sets. 2004.
- Bauschke and Moursi. Two affine subspaces 2016; Convex feasibility 2017.
- Liu, Ryu, and Yin. Conic programming, 2017.

ADMM under specific pathological setups for conic/quadratic programs:

- Raghunathan and Cairano, 2014.
- Stellato, Banjac, Goulart, Bemporad, and Boyd, 2017.
- Banjac, Goulart, Stellato, and Boyd. 2017.

We apply existing fixed-point analysis to get *asymptotic feasibility*.

Then, we add recession function analysis to get *improving directions*.

Next, we use objective value analysis to get *asymptotic optimality*.

Fixed-point analysis

Asymptotic behavior of fixed point iteration

The *infimal displacement vector*[11] is defined as

$$v := \mathrm{Proj}_{\overline{\mathrm{range}(I-T)}}\mathbf{0}.$$

Lemma (Pazy'71,Baillon-Bruck-Reich'78)

When T is firmly nonexpansive, then

$$z^k - T(z^k) \to v.$$

If $v = \mathbf{0}$, $z^k - z^{k+1} \to \mathbf{0}$ and $x^{k+1} - x^{k+1/2} \to \mathbf{0}$, so DRS is asymptotically feasible.

If $v \neq \mathbf{0}$, we can understand the limiting behavior z^k with v.

[11] name coined in Bauschke-Hare-Moursi'14

Characterization of v

Theorem (Bauschke, Hare, Moursi'16)

When T is the DRS operator,

$$\overline{\mathrm{range}(I-T)} = \overline{\mathrm{dom}f - \mathrm{dom}g} \cap \overline{\mathrm{dom}f^* + \mathrm{dom}g^*}$$

Consequence (example): if (P) is infeasible, then $v = \Pi_{\overline{\mathrm{dom}f - \mathrm{dom}g}}(\mathbf{0})$, i.e. v represents the shortest distance from $\mathrm{dom}g$ and $\mathrm{dom}f$. This implies

$$\|x^{k+1} - x^{k+1/2}\| \to \mathrm{dist}(\mathrm{dom}f, \mathrm{dom}g).$$

DRS makes an effort to achieve feasibility.

Unbounded problems and improving directions

Recession function

Let $f'(x;d)$ be d-directional derivative at x.

The recession function of f is defined as

$$\mathrm{rec}f(d) = \lim_{\alpha \to \infty} f'(x + \alpha d; d).$$

$\mathrm{rec}f$ characterizes the *asymptotic rate* of f as we go in direction d.

$\mathrm{rec}f(d)$ invariable for $x \in \mathrm{dom}f$, and is possibly ∞.

$\mathrm{rec}f(d)$ generalizes the *constant rate* of a linear program along direction d.

Recession function and improving direction

Lemma

d is an improving direction (P)*, i.e., for some $C > 0$,*

$$f(x+d) + g(x+d) \leq f(x) + g(x) - C, \quad \forall x \in \mathrm{dom} f \cap \mathrm{dom} g,$$

if and only if

$$\mathrm{rec} f(d) + \mathrm{rec} g(d) < 0,$$

and if and only if

(P) *is feasible and* (D) *is strongly infeasible.*

DRS under dual strong infeasibility

Using duality relationships such as $\mathrm{rec} f = (\sigma_{f^*})^*$, we can get:

Theorem

If (P) *is feasible and* (D) *is strongly infeasible, then*

$$d(x^{k+1/2}, \mathrm{dom}\, g) \to 0, \quad d(x^{k+1}, \mathrm{dom}\, f) \to 0$$

and $x^{k+1/2} - x^{k-1/2} = d + o(1)$ *for some improving direction* $d \neq \mathbf{0}$*.*

Similar results hold for different pathologies.

Objective value analysis

Fixed-point analysis is not enough

Under certain pathologies, DRS iterates satisfy

$$z^k - T(z^k) \to 0.$$

This is much alike the fact that

$$\nabla f(x^k) \to 0$$

does *not* necessarily imply

$$f(x^k) \to \inf_x f(x)$$

Example

Consider convex function

$$f(x,y) := x^2/y, \quad y > 0.$$

We have $\inf_{x,y} f(x,y) = 0$.

For $y := x^2$,

$$\begin{cases} f(x,x^2) \equiv 1 \\ \nabla f(x,x^2) = (2/x, -1/x^4) \to 0 \quad \text{as } x \to \infty \end{cases}$$

So, we must separately show DRS achieves approximately optimal objective.

Primal subvalue

Define the *primal subvalue* as

$$p^- := \lim_{\varepsilon \to 0^+} \inf_{\|x-y\| \leq \varepsilon} \{f(x) + g(y)\},$$

i.e., p^- is the optimal value of an infinitesimally perturbed (P).

Lemma

When convex,

$$d^* = p^- \leq p^*.$$

Convex problems with non-zero duality gap exist and are ill-posed.

Asymptotic objective convergence

Applying convex inequalities and primal subvalue analysis, we can get:

Theorem

If (P) *is feasible but has no solution and* (D) *is feasible, then*

$$x^{k+1/2} - x^k \to \mathbf{0}$$

and

$$\liminf_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) = p^\star.$$

We can say something for all pathological cases, so long as $p^\star = d^\star$.

Is strong duality $p^\star = d^\star$ necessary?

DRS can reduce the function value below $p^\star$ when strong duality fails.

In numerical examples, we observed DRS finds wrong objective:

$$\lim_{k\to\infty} f(x^{k+1/2}) + g(x^{k+1}) < p^\star.$$

A problem with $p^\star = 1$ but $d^\star = 0$:

$$\underset{x \in \mathbb{R}^2}{\text{minimize}} \ \ \exp(-\sqrt{x_1 x_2}) + \delta_{x_1=0}(x).$$

Another problem with $p^\star = 1$ but $d^\star = 0$:

$$\underset{X \in \boldsymbol{S}^3_+}{\text{minimize}} \ \delta_{\boldsymbol{S}^3_+}(X) + \Big(X_{22} + \delta_{\boldsymbol{S}^3, X_{33}=0, X_{22}+2X_{13}=1}(X)\Big).$$

For both, we observed: $\lim_k f(x^{k+1/2}) + g(x^{k+1}) \in [d^\star, p^\star)$.

Conjecture

If $d^\star < p^\star$, then DRS necessarily finds a wrong optimal value.

Summary

- DRS pathologies: arXiv:1801.06618 (under revision)
  Works whenever $p^\star = d^\star$, even for (strong/weak) infeasibility, unbounded (improving dir exists/not), and solution not attainable

- DRS for conic programming: arXiv:1706.02374 (accepted by MPA)
  Identify the pathologies of conic programs and generate certificates.
  "Rate of divergence." Numerically useful for weak pathologies.

Thank you!

Example 1

- **3-variable problem:**

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

- this problem is feasible, $p^\star = -\infty$ (by letting $x_3 \to \infty$ and $x_1 \to -\infty$), and has no improving direction[12]

- existing solvers[13]:
  - SDPT3: "Failed", $p^\star$ no reported
  - SeDuMi: "Inaccurate/Solved", $p^\star = -175514$
  - Mosek: "Inaccurate/Unbounded", $p^\star = -\infty$

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

[13] using their default settings

Example 2

- **3-variable problem:**

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

- this problem is infeasible[14], $\mathrm{dist}(\mathcal{L}, K) = 0$ [15], and has no strict separating hyperplane

- existing solvers[16]:
  - SDPT3: "Infeasible", $p^\star = \infty$
  - SeDuMi: "Solved", $p^\star = 0$
  - Mosek: "Failed", $p^\star$ not reported

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

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

[16] using their default settings