# Douglas-Rachford Splitting and ADMM for Pathological Convex Optimization

**Ernest K. Ryu** Yanli Liu Wotao Yin

ISMP
July 5, 2018

# Problem setup

Consider the primal problem

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

and its dual problem

$$\text{maximize} \quad -f^*(\nu) - g^*(-\nu) \tag{D}$$

Write $p^\star$ and $d^\star$ for the primal and dual optimal values.

# Douglas-Rachford splitting

DRS applied to (P):

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

DRS finds solutions to (P) and (D). Convergence depends on the status of both (P) and (D).

# DRS convergence (classical)

The classic literature says DRS “converges” when

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

However, are these assumptions actually necessary?

# DRS convergence (new)

DRS "works" when

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

Summary of this work: DRS essentially "works" when $p^\star = d^\star$.

# Pathology: definition

Problem pair (P) and (D) is not pathological if

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

Otherwise, it's pathological.

For a precise discussion, we need to classify pathologies into several cases. Let's not do that here.

# Pathology: examples

(P) is (weakly) infeasible

$$\underset{x\in\mathbb{R}}{\text{minimize}} \quad \underbrace{-\log x}_{=f(x)} + \underbrace{1/\sqrt{-x}}_{=g(x)} \qquad \text{(P)}$$

Note $\operatorname{dom} f = (0,\infty)$ and $\operatorname{dom} g = (-\infty, 0]$. Infeasible since $\operatorname{dom} f \cap \operatorname{dom} g = \emptyset$. Weakly infeasible since $\operatorname{dist}(\operatorname{dom} f, \operatorname{dom} g) = 0$.

# Pathology: examples

(P) is feasible but has no solution

$$\underset{(\nu_1,\nu_2)\in\mathbb{R}^2}{\text{minimize}} \quad \sqrt{\nu_1^2+\nu_2^2} - \nu_1 + \delta_{\{\nu_2=1\}}(-\nu_2)$$

(D) is

$$\underset{(x_1,x_2)\in\mathbb{R}^2}{\text{maximize}} \quad -\delta_{\{(x_1,x_2)\,|\,x_1^2+x_2^2\leq 1\}} - x_2 - \delta_{\{(x_1,x_2)\,|\,x_1=1\}} \qquad \text{(P)}$$

Nevertheless, $d^\star = p^\star = 0$.

# What do we want DRS to do?

We want DRS to find a point that is **approximately feasible** and (when applicable) **approximately optimal**.

E.g. if (P) is weakly infeasible, we want

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

E.g. if (P) is feasible but has no solution, we want

$$x^{k+1} - x^{k+1/2} \to 0$$
$$f(x^{k+1/2}) + g(x^{k+1}) \to p^\star$$

# DRS convergence (new): examples

Theorem

*If (P) is weakly infeasible and* $p^\star = d^\star = \infty$*, then*

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

Theorem

*If (P) is feasible but has no solution and* $p^\star = d^\star > -\infty$*.*

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

*and*

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

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

# Theoretical components

DRS has 2 goals: achieve feasibility and reduce function value.
We use 2 set of tools to show DRS achieves both goals.

**Operator theory and "fixed-point iteration" without a fixed point.** With this machinery, we show things like $x^{k+1} - x^k \to 0$ or $x^{k+1} - x^k \to v$, where we characterize $v$.

**Function-value analysis (i.e., subgradient inequalities).** With this machinery, we show things like $f(x^{k+1/2}) + g(x^{k+1}) \to p^\star$. This part needs the $d^\star = p^\star$ assumption.

# Prior work

There has been surprisingly little work studying DRS and ADMM under pathologies. Our understanding is still incomplete.

Results in specific pathological setups:

- Bauschke, Combettes, and Luke, 2004.
- Bauschke and Moursi, 2016, 2017,
- Liu, Ryu, and Yin, 2018.

Results on general setups:

- Bauschke, Hare, and Moursi, 2014, 2016, 2017

ADMM under specific pathological setups for conic programs:

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

# Outline

# Fixed point iteration with a fixed point

DRS is a fixed point iteration with a firmly nonexpansive operator $T$:

$$z^{k+1} = T(z^k).$$

Under non-pathology, $T : \mathbb{R}^d \to \mathbb{R}^d$ has a fixed point and

$$z^k \to z^\star$$

for some fixed point $z^\star$.

# Fixed point iteration without fixed points

The *infimal displacement vector* of $T$ is

$$v = P_{\overline{\mathrm{range}(I-T)}} 0.$$

## Lemma (Pazy 1971, Baillon, Bruck, Reich 1978)

*When $T$ is firmly nonexpansive and has no fixed point, then*

$$z^k = -kv + o(k), \qquad z^k - T(z^k) \to v.$$

# Fixed point iteration without fixed points

Under pathology, $T : \mathbb{R}^d \to \mathbb{R}^d$ has no fixed point, and

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

When $v = 0$, $z^{k+1} - z^k \to 0$, and therefore $x^{k+1} - x^k \to 0$. (DRS achieves approximate feasibility.)

When $v \neq 0$, we can understand the asymptotic behavior of DRS with $v$. (This $v$ turns out to be a certificate of infeasibility.)

# Characterization of $v$

Theorem (Bauschke, Hare, Moursi 2016)

*When $T$ is the DRS operator,*

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

# Characterization of $v$: Infeasible case

When (P) is infeasible,

$$v = \Pi_{\overline{\mathrm{dom}\, f - \mathrm{dom}\, g}}(0),$$

i.e., $v$ represents the shortest distance from $\mathrm{dom}\, g$ to $\mathrm{dom}\, f$.

This implies $\|x^{k+1} - x^{k+1/2}\| \to \mathrm{dist}(\mathrm{dom}\, f, \mathrm{dom}\, g)$,
i.e., $(x^{k+1/2}, x^{k+1})$ represents the best effort to achieve feasibility.

# Characterization of $v$: Other cases

When (P) feasible but has no solution, and (D) feasible, we know $z^k - T(z^k) \to 0$. For more details, we need more work.

When (P) is feasible, and (D) is strongly infeasible, we know $z^k - T(z^k) \to v$. To further understand what $v$ is, we need more work.

For other pathological cases, we also need more work to concretely understand the asymptotic behavior.

# Outline

# Improving direction

$d \neq 0$ is an *improving direction* for (P) if we have $C > 0$ such that

$$f(x+d) + g(x+d) \leq f(x) + g(x) - C$$

for all $x$.

When $d$ is an improving direction, we have

$$f(x+\alpha d) + g(x+\alpha d) = -C\alpha + o(\alpha)$$

as $\alpha \to \infty$ for all feasible $x$.

If (P) has an improving direction, then $p^\star = -\infty$. (This generalizes the notion of improving directions in conic programs.)

# Recession function

The recession function of $f$ is defined as

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

for any $x \in \operatorname{dom} f$.

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

# Recession function and improving direction

**Lemma**

*$d$ is an improving direction for (P) if and only if*

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

*and if and only if (D) is strongly infeasible.*

I.e., improving directions are closely related to recession functions.

# Characterization of $v$

Using duality relationships like $\operatorname{rec} f = (\sigma_{f^*})^*$, we can characterize $v$ with improving directions.

## Lemma

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

$$v = -d \neq 0$$

*for some improving direction $d$.*

Similar results hold for different pathologies.

# DRS with strong dual infeasibility

Theorem

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

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

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

Similar results hold for different pathologies.

# Outline

# Fixed-point analysis is not enough

Under certain pathologies, DRS iterates satisfy

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

However, this is not enough.

This is much alike the fact that

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

does not necessarily imply

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

even if $f$ is convex.

# Counter-example

If

$$f(x, y) = x^2/y$$

then

$$f(x, x^2) = 1$$

but

$$\nabla f(x, x^2) = (2/x, -1/x^4) \to 0$$

as $x \to \infty$.

Consequence: We must separately show the DRS iterates achieve approximately optimal function values.

# Key inequality and its consequence

## Lemma

$$f(x^{k+1/2}) - f(x) + g(x^{k+1}) - g(x) \leq (1/\gamma)(x^{k+1} - x^{k+1/2})^T(x - z^{k+1})$$

*For any $x$.*

With some work, we can use this inequality to show

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

This is an inequality, and we need the other direction.

# 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 (P) achieved with infinitesimal infeasibilities.

## Lemma

*When convex,*

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

(This is where $d^\star = p^\star$ enters the analysis.)

With this, we can show

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

# Outline

# Convergence results

Putting these pieces together, we can show things like

Theorem

*If (P) is feasible but has no solution, (D) is feasible, and $p^\star = d^\star$. Then*

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

*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 necessary?

DRS iteration has 2 goals: achieve feasibility and reduce function value.

Because DRS never arrives at feasibility, it can reduce the function value below $p^\star$ when strong duality fails.

## Conjecture

*When strong duality fails, DRS necessarily fails in that*

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

*In other words, DRS finds the wrong objective value.*

# Evidence for conjecture

The pathological problem

$$\underset{x\in\mathbb{R}^2}{\text{minimize}} \quad \underbrace{\exp(-\sqrt{x_1x_2})}_{f(x)} + \underbrace{\delta_{\{(x_1,x_2)\,|\,x_1=0\}}(x)}_{g(x)}$$

has $p^\star = 1$ but $d^\star = 0$. Experimentally, we observe

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

for all $\gamma > 0$.

# Evidence for conjecture

The pathological problem

$$\underset{X \in \mathbf{S}^3}{\text{minimize}} \quad \underbrace{\delta_{\mathbf{S}^3_+}(X)}_{f(X)} + \underbrace{X_{22} + \delta_{\{X \in \mathbf{S}^3 \mid X_{33}=0, X_{22}+2X_{13}=1\}}(X)}_{g(X)},$$

has $p^\star = 1$ but $d^\star = 0$. Experimentally, we observe

$$d^\star = \lim_{k \to \infty} f(x^{k+1/2}) + g(x^{k+1})$$

for $\gamma \geq 0.5$, and

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

for $0 < \gamma < 0.5$.

# Outline

# Setup

For ADMM, consider the primal problem

$$\begin{array}{ll} \text{minimize} & f(x) + g(y) \\ \text{subject to} & Ax + By = c, \end{array} \qquad \text{(P-ADMM)}$$

and its dual problem

$$\text{maximize} \quad -f^*(-A^T\nu) - g^*(-B^T\nu) - c^T\nu. \qquad \text{(D-ADMM)}$$

Write $p^\star$ and $d^\star$ for the primal and dual optimal values.

# Method

ADMM applied to this primal-dual problem pair is

$$\begin{aligned} x^{k+1} &\in \operatorname*{arg\,min}_{x\in\mathbb{R}^p} L_\rho(x, y^k, \nu^k) \\ y^{k+1} &\in \operatorname*{arg\,min}_{y\in\mathbb{R}^q} L_\rho(x^{k+1}, y, \nu^k) \\ \nu^{k+1} &= \nu^k + \rho(Ax^{k+1} + By^{k+1} - c). \end{aligned}$$

(We need to assume something to ensure the subproblems have a solution.)

# Pathological convergence examples

If $d^\star = p^\star \in [-\infty, \infty)$, primal problem is feasible but has no solution, then $Ax^k + By^k - c \to 0$ and

$$\lim_{k\to\infty} \frac{1}{k} \sum_{i=1}^{k} f(x^i) + g(y^i) = p^\star, \quad \liminf_{k\to\infty} f(x^k) + g(y^k) = p^\star.$$

# Pathological convergence examples

It $d^\star = p^\star = \infty$, problem problem is infeasible, then

$$\|Ax^k + By^k - c\| \to \inf_{\substack{x \in \mathrm{dom}\, f \\ y \in \mathrm{dom}\, g}} \|Ax + By - c\|.$$

# Outline

# Conclusion and future work

Conclusion of this work:

- With some caveats, DRS and ADMM work when strong duality holds.
- We conjectured that DRS necessarily fails when strong duality fails, and provided supporting evidence.

Open questions:

- What happens to DRS and ADMM in the absence of strong duality?
- DRS can be generalized with a relaxation parameter in $(0, 2)$. Our analysis generalizes to this setup. ADMM can be generalized with a relaxation parameter in $(0, 1.618)$. Our analysis does not immediately generalize to this setup.