Uniqueness of DRS as the 2 Operator Resolvent-Splitting and Impossibility of 3 Operator Resolvent-Splitting

Ernest K. Ryu

June 12, 2018
DIMACS Workshop on ADMM and
Proximal Splitting Methods in Optimization

Douglas-Rachford splitting

Consider the monotone inclusion problem

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A + B)x$$

Douglas-Rachford splitting (DRS) elegantly solves this problem with

$$z^{k+1} = (1/2)z^k + (1/2)(2J_{\alpha A} - I)(2J_{\alpha B} - I)z^k$$

P. L. Lions and B. Mercier, Splitting Algorithms ..., 1979.

Other splitting methods?

Given the success of DRS, we ask:

1. Are there other 2 operator splittings?
2. Can we generalize DRS to 3 operators?

Question 2 has been a long-standing open problem.
"[T]he convergence seems difficult to prove ... in the case of a sum of 3 operators." — Lions & Mercier

Answer: no and no (in a certain sense).

Rules of the game

What counts as a generalization of DRS?

DRS has the following key properties:

1. Only uses scalar multiplication, addition, and resolvents. (Resolvent-splitting)
2. Only uses $J_{\alpha A}$ and $J_{\alpha B}$ once. (Frugal)
3. Converges for any maximal monotone operators A and B. (Unconditional convergence)
4. Does not enlarge the problem size, i.e., $T : \mathbb{R}^d \to \mathbb{R}^d$ where $x \in \mathbb{R}^d$. (No lifting)

Let's look for generalizations of DRS that satisfy these 4 properties.

Other splitting methods: FBS

FBS solves

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A+B)x$$

with

$$x^{k+1} = (I + \alpha A)^{-1}(I - \alpha B)(z^x)$$

Conditionally convergent and not a resolvent-splitting.

Other splitting methods: PDHG

PDHG solves

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A+B)x$$

with

$$\begin{aligned} x^{k+1} &= J_A(x^k - \alpha u^k) \\ u^{k+1} &= (I - J_B)(u^k + \alpha(2x^{k+1} - x^k)) \end{aligned}$$

Uses lifting, since it maps $(x^k, u^k) \mapsto (x^{k+1}, u^{k+1})$.

Other splitting methods: DYS

DYS solves

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A + B + C)x$$

with

$$z^{k+1} = (I - J_{\alpha B} + J_{\alpha A} \circ (2J_{\alpha B} - I - \alpha C \circ J_{\alpha B}))z^k$$

Not a resolvent-splitting.

Other splitting methods

Other methods, such as

- FBFS
- PDFP^{2}O/PAPC
- Condat-Vũ
- GFBS
- PD3O
- PDFP
- AFBA
- FDRS
- FBHFS
- projective splitting
- method of Bricenõ-Arias and Combettes (2011)
- method of Combettes, Condat, Pesquet, Vũ (2014)

do not satisfy the 4 properties.

Other splitting methods

These other splitting methods have certainly provided great value.

Many of them include DRS as a special case and therefore are generalizations of DRS, in that sense.

However, they do not satisfy the 4 stated properties and therefore are not generalizations of DRS, in this sense.

(Proximal point method satisfies the 4 properties, but PPM is for 1 operator.)

Goal

Study generalizations of DRS by seeking splitting methods that satisfy these 4 properties.

This is somewhat of an arbitrary requirement. The other splittings are useful, even though they do not comply with these 4 requirements.

However, limiting the class of methods we study will allow us to characterize what is and is not possible.

Outline

Definitions

(T, S) is a fixed-point encoding for

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A + B)x \tag{2op}$$

if

$$\exists z^\star \text{ such that } \begin{pmatrix} T(A, B, z^\star) & = z^\star \\ S(A, B, z^\star) & = x^\star \end{pmatrix} \quad \Leftrightarrow \quad 0 \in (A + B)(x^\star).$$

For notational simplicity, drop the dependency on A and B.

T is the fixed-point mapping and S is the solution mapping.

Definitions

(T, S) is a resolvent-splitting if it is a fixed-point encoding constructed with resolvents of A and B, addition, and scalar multiplication

(T, S) is without lifting if $T : \mathbb{R}^d \to \mathbb{R}^d$ and $S : \mathbb{R}^d \to \mathbb{R}^d$.

(T, S) is frugal if it uses $J_{\alpha A}$ and $J_{\beta B}$ once.

Definition: equivalence

We can scale a fixed point iteration to get another one that is essentially the same:

$$z^{k+1} = T(z^k) \quad \Leftrightarrow \quad az^{k+1} = aT(a^{-1}az^k)$$

for any $a \neq 0$. We can swap the role of A and B of a splitting to get another one that is conceptually no different:

$$(T(A, B, \cdot), S(A, B, \cdot)) \quad \Leftrightarrow \quad (T(B, A, \cdot), S(B, A, \cdot)).$$

Two splittings are equivalent if one can be obtained from the other through scaling or swapping A and B.

All 2 operator resolvent-splitting

Theorem

Up to equivalence, any frugal resolvent-splitting without lifting for (2op) *can be expressed as*

$$\begin{aligned} x_1 &= J_{\alpha A} z \\ x_2 &= J_{\beta B}((1+\beta/\alpha)x_1 - (\beta/\alpha)z) \\ T(z) &= z + \theta(x_2 - x_1) \end{aligned}$$

for $\alpha, \beta > 0$, $\theta \neq 0$. *Also,*

$$S(z) = \eta x_1 + (1-\eta)x_2$$

and $\eta \in \mathbb{R}$.

Definition: unconditional convergence

A fixed-point encoding (T, S) converges unconditionally if

$$ST^k z^0 \to x^\star \in \mathbf{zer}(A + B)$$

for any $z^0 \in \mathbb{R}^d$ and $A, B \in \mathcal{M}$.

DRS is unique

Theorem

A fixed-point encoding of the form

$$\begin{aligned} x_1 &= J_{\alpha A} z \\ x_2 &= J_{\beta B}((1+\beta/\alpha)x_1 - (\beta/\alpha)z) \\ T(z) &= z + \theta(x_2 - x_1) \\ S(z) &= \eta x_1 + (1-\eta)x_2 \end{aligned}$$

converges unconditionally if and only if $\alpha = \beta > 0$ *and* $\theta \in (0, 2)$*.*

Proof sketch

Consider the problem

$$\underset{x \in \mathbb{R}^2}{\text{find}} \quad 0 = (A + B)x$$

where

$$A = \begin{bmatrix} 0 & \tan(\omega)/\alpha \\ -\tan(\omega)/\alpha & 0 \end{bmatrix} \qquad B = \begin{bmatrix} 0 & -\tan(\omega)/\beta \\ \tan(\omega)/\beta & 0 \end{bmatrix}$$

Proof sketch

With basic algebra, we can show that

$$T(z) = \begin{bmatrix} 1 & \theta(\alpha-\beta)\cos(\omega)\sin(\omega) \\ -\theta(\alpha-\beta)\cos(\omega)\sin(\omega) & 1 \end{bmatrix} z$$

With basic eigenvalue computation, we get

$$|\lambda_1|^2 = |\lambda_2|^2 = 1 + \left(\theta(1-\beta/\alpha)\cos(\omega)\sin(\omega)\right)^2$$

For $\beta \neq \alpha$ the iteration diverges. When $\beta = \alpha$, the splitting is DRS.

Outline

Uniqueness of 2 operator resolvent-splitting

Impossibility of 3 operator resolvent-splitting

Attainment of 3 operator resolvent-splitting with minimal lifting

Conclusion

Definitions

A pair of functions (T, S) is a fixed-point encoding of the problem

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A + B + C)(x) \tag{3op}$$

if

$$T(z^\star) = z^\star,\, x^\star = S(z^\star) \quad \Leftrightarrow \quad 0 \in (A + B + C)(x^\star)$$

(T, S) is a resolvent-splitting if it is a fixed-point encoding constructed with (finitely many) resolvents of A, B, and C, addition, and scalar multiplication

(T, S) is without lifting if $T : \mathbb{R}^d \to \mathbb{R}^d$ and $S : \mathbb{R}^d \to \mathbb{R}^d$.

Impossibility result

Theorem

There is no resolvent-splitting **without lifting** *for* (3op).

Clarification: Assume $T : \mathbb{R}^d \to \mathbb{R}^d$ and $S : \mathbb{R}^d \to \mathbb{R}^d$ are constructed with finitely many resolvents,

$$J_{\alpha_1 A}, J_{\alpha_2 A}, \ldots, J_{\alpha_{n_A} A}$$
$$J_{\beta_1 B}, J_{\beta_2 B}, \ldots, J_{\beta_{n_B} B}$$
$$J_{\gamma_1 C}, J_{\gamma_2 C}, \ldots, J_{\gamma_{n_C} C}$$

with possibly distinct scaling parameters α_i, β_j, γ_k. (T, S) will fail to be a fixed-point encoding for some A, B, and C.

Impossibility result

Theorem

There is no resolvent-splitting **without lifting** *for* (3op).

Clarification: The set of operators constructed with resolvents, identity operator, and scalar multiplication forms a "near-ring". No element of this near-ring is a fixed-point encoding for (3op).

Outline

Frugal 3 operator resolvent-splitting with lifting

To solve

$$\underset{x \in \mathbb{R}^d}{\text{find}} \quad 0 \in (A + B + C)x$$

a standard trick is to "copy" variables

$$\underset{x_1, x_2, x_3 \in \mathbb{R}^d}{\text{find}} \quad 0 \in \begin{bmatrix} Ax_1 \\ Bx_2 \\ Cx_3 \end{bmatrix} + N_{\{(x_1, x_2, x_3) \,|\, x_1 = x_2 = x_3\}}(x_1, x_2, x_3),$$

and apply DRS.

Frugal 3 operator resolvent-splitting with lifting

The DRS iteration is

$$\begin{aligned}\bar{z} &= (1/3)(z_A + z_B + z_C)\\ T_A(\boldsymbol{z}) &= z_A + J_{\alpha A}(2\bar{z} - z_A) - \bar{z}\\ T_B(\boldsymbol{z}) &= z_B + J_{\alpha B}(2\bar{z} - z_B) - \bar{z}\\ T_C(\boldsymbol{z}) &= z_C + J_{\alpha C}(2\bar{z} - z_C) - \bar{z}\end{aligned} \tag{1}$$

which also coincides with Spingarn's method.

This is a resolvent-splitting with lifting, since $\boldsymbol{T} : \mathbb{R}^{3d} \to \mathbb{R}^{3d}$.

With lifting it's possible. However, how much do we need to lift?

Definition: ℓ-fold lifting

We say a resolvent-splitting $(\boldsymbol{T}, S)$ uses *ℓ-fold lifting* if

$$\boldsymbol{T} : \mathbb{R}^{\ell d} \to \mathbb{R}^{\ell d} \qquad S : \mathbb{R}^{\ell d} \to \mathbb{R}^{d}.$$

1-fold lifting corresponds to no lifting.

Minimal lifting

Unconditionally convergent frugal resolvent-splitting with 3-fold lifting is possible.

Question: can we do this with 2-fold lifting?

Answer: yes.

Since 1-fold lifting is impossible, we call 2-fold lifting the minimal lifting.

Frugal 3 operator resolvent-splitting with minimal lifting

Theorem

The operator $\boldsymbol{T} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ *defined as*

$$\begin{aligned} x_1 &= J_{\alpha A}(z_1) \\ x_2 &= J_{\alpha B}(z_2 + x_1) \\ x_3 &= J_{\alpha C}\left(x_1 - z_1 + x_2 - z_2\right) \\ T_1(\boldsymbol{z}) &= z_1 + \theta x_3 - \theta x_1 \\ T_2(\boldsymbol{z}) &= z_2 + \theta x_3 - \theta x_2 \end{aligned}$$

for $\alpha > 0$ *and* $\theta \in (0, 1)$ *is unconditionally convergent.*

Proof is done with first principles, since this does not reduce to any known splitting (to the best of my knowledge).

Numerical example

Whether this splitting fast or efficient is somewhat besides the point, as its purpose is to establish attainment of minimal lifting.

Nevertheless, let's try it out to see if it works well.

Numerical example

Consider the Markowitz portfolio optimization problem

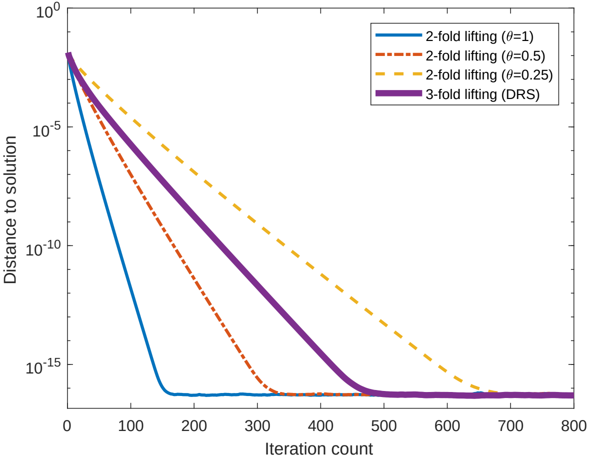
$$\begin{array}{ll} \underset{x\in\mathbb{R}^d}{\text{minimize}} & (1/n)\sum_{i=1}^n (a_i^T x - b)^2 \\ \text{subject to} & x \in \Delta \\ & \mu^T x \geq b, \end{array}$$

where d is the number of assets, $a_1, \ldots, a_n \in \mathbb{R}^d$ are n realizations of the returns on the assets, Δ is the standard simplex for portfolios with no short positions, $\mu \in \mathbb{R}^d$ is the average return of the assets, and $b \in \mathbb{R}$ is the desired expected return. We reformulate this problem as

$$\underset{x\in\mathbb{R}^d}{\text{minimize}} \quad \frac{1}{n}\sum_{i=1}^n (a_i^T x - b)^2 + \delta_\Delta(x) + \delta_{\{x \,|\, \mu^T x \geq b\}}(x)$$

We used synthetic data with $n = 30000$ and $d = 10000$, which make the data approximately 2GB in size.

Numerical example



The spitting with minimal lifting works well.

Outline

Conclusion

This work shows:

1. DRS is the unique frugal, unconditionally convergent resolvent-splitting without lifting.
2. There is no resolvent-splitting without lifting for 3 operators.
3. 2-fold lifting is the minimal lifting necessary for 3 operators.

The proofs are based on:

1. parameterizing the splitting and simplifying it based a technique inspired by Gaussian elimination and
2. showing when the Gaussian elimination step fails, we can construct counter examples. (This is the harder part.)