A THEORY OF GENERALIZED PROXIMITY FOR ADMM

François D. Côté, Ioannis N. Psaromiligkos, and Warren J. Gross

Department of Electrical and Computer Engineering, McGill University, Montreal, Canada

francois.cote@mail.mcgill.ca; yannis@ece.mcgill.ca; warren.gross@mcgill.ca

In brief

ADMM, or the alternating direction method of multipliers, has become the most widely used proximal method in signal processing, but as this study reveals, the method turns out to be a deficient rearrangement of a more practical approach that we propose, which generalizes the Douglas-Rachford algorithm (DR).

Our method stems from a notion of proximity that we develop, which offers new findings on local inversion and infimal postcomposition.

Compared to ADMM, our approach enjoys not only a more reasonable form but also a more general convergence result.

Introduction

ADMM vs. DR Fundamentally, there is a close relationship between ADMM and the Douglas-Rachford algorithm (DR). Yet in the literature, they are growing apart—ADMM having seduced scientists and engineers, and DR, mathematicians. We draw two distinctions:

- ADMM is conducive to a more practical problem.
- DR enjoys a more general convergence result.

Notation

$\mathcal{F}$, $\mathcal{G}$, $\mathcal{H}$	Real Hilbert spaces
$\Gamma_0(\mathcal{H})$	Class of proper convex lower semicontinuous mappings $\mathcal{H} \to (-\infty, +\infty]$
$\mathcal{B}(\mathcal{G}, \mathcal{H})$	Space of bounded linear mappings $\mathcal{G} \to \mathcal{H}$
dom	Effective domain
cone	Conical hull
int, ri	Interior and relative interior

Problems ADMM handles problems of the form

$$\min_{(y,z)\in\mathcal{F}\times\mathcal{G}} f(y) + g(z) \quad \text{s.t.} \quad Ay + Bz = c,$$

which fits many modern, real-world situations, particularly distributed ones. On the other hand, DR handles problems of a simpler kind:

$$\min_{x\in\mathcal{H}} f(x) + g(x).$$

Convergence While such a result has remained vague for ADMM, DR converges under a result combining

- inexact updates,
- infinite dimensions,
- the acceleration technique of relaxation, and
- entirely primal (as opposed to dual) assumptions.

Result To have the best of both worlds, we generalize DR, re-tailoring it to the more practical problem, while keeping the general convergence result.

Formulation

Context Let f be a function in $\Gamma_0(\mathcal{F})$, and g one in $\Gamma_0(\mathcal{G})$. Let A be an operator in $\mathcal{B}(\mathcal{F},\mathcal{H})$, and B one in $\mathcal{B}(\mathcal{G},\mathcal{H})$. Finally, let c, d, and e be vectors in $\mathcal{H}$. Suppose that

$$f + \|A\cdot\|^2 \quad \text{and} \quad g + \|B\cdot\|^2 \quad \text{are strongly convex}$$

and that the following qualification condition holds:

$$\operatorname{cone}(A\operatorname{dom} f + B\operatorname{dom} g - c) \quad \text{is a closed subspace.}$$

Problem

P $$\min_{(y,z)\in\mathcal{F}\times\mathcal{G}} f(y) + g(z) \quad \text{s.t.} \quad Ay + Bz = c = d + e.$$

Proposition The qualification condition holds in each of the following cases:

- $\operatorname{int}(A\operatorname{dom} f) \cap (-B\operatorname{dom} g + c) \neq \varnothing$ or $(-\operatorname{int}(B\operatorname{dom} g)) \cap (A\operatorname{dom} f - c) \neq \varnothing$.
- f is finite and A is surjective, or similarly with g and B.
- $\mathcal{H}$ is finite dimensional and $\operatorname{ri}(A\operatorname{dom} f) \cap (-\operatorname{ri}(B\operatorname{dom} g) + c) \neq \varnothing$.

Related problem Notice that for any (y,z) satisfying the constraint, there exists an x such that

a $x = Ay - d$ and **b** $x = -Bz + e$.

Consider finding an x corresponding to a solution to **P**. Among all y values satisfying **a**, we can assign to each x, those minimizing f. This minimum is given by the infimal postcomposition

$$(A\cdot -d) \triangleright f(x) \overset{\text{def}}{=} \inf_{\substack{y\in\mathcal{F} \\ Ay-d=x}} f(y).$$

Having repeated the operation with (z, b, g), we can then find an x for which an associated (y,z) minimizes $f+g$:

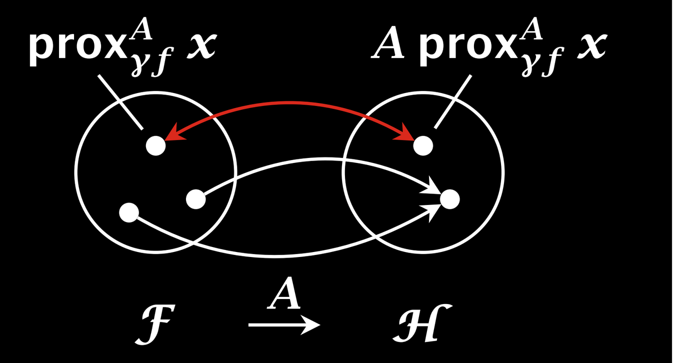
$$\min_{x\in\mathcal{H}} (A\cdot -d) \triangleright f(x) + (-B\cdot +e) \triangleright g(x).$$

Not only can DR solve this problem; quite remarkably, DR produces the optimal (y,z) as a byproduct!

Key result

Partial inversion

The generalized proximal mapping inherits many properties of the vanilla operator. But it also manifests new ones, like allowing for a local inversion of the possibly non-injective A. We provide this property in the following result:



Lemma In the context of generalized proximity, let y be a vector in $\mathcal{F}$, and suppose that

$$f(y) \le f(y + \Delta y) \quad \text{for every } \Delta y \text{ in the null space of } A.$$

Then, A is invertible at any proximal point that is equal to y, which is to say,

$$A\operatorname{prox}^A_{\gamma f} x = Ay \quad \Longrightarrow \quad \operatorname{prox}^A_{\gamma f} x = y.$$

Proximity

Context Fix a real number $\gamma > 0$. Let x be a vector in $\mathcal{H}$. Let f be a function in $\Gamma_0(\mathcal{F})$, and A an operator in $\mathcal{B}(\mathcal{F},\mathcal{H})$. Suppose that

$$f + \|A\cdot\|^2 \quad \text{is strongly convex.}$$

Functions resulting from infimal postcomposition belong to the class of functions that DR requires.

Theorem *In the context of generalized proximity,*

$$(A\cdot +c) \triangleright f \in \Gamma_0(\mathcal{H}).$$

Defining problem

$$\min_{y\in\mathcal{F}} \gamma f(y) + \frac{1}{2}\|Ay - x\|^2.$$

Generalized mapping We can prove that there must exist a unique solution to this problem—a solution, which we denote by

$$\operatorname{prox}^A_{\gamma f} x,$$

that generalizes the ordinary proximal operator and exhibits new properties, like indicating where A is invertible [*see the box at the top*]. With c a vector in $\mathcal{H}$, the prox relative to $(A\cdot +c) \triangleright f$ is well defined: we have found that infimal postcomposition can yield a function in $\Gamma_0(\mathcal{H})$ [*see the wrapped text above*]. In fact,

$$\operatorname{prox}_{\gamma((A\cdot+c)\triangleright f)} x = A\operatorname{prox}^A_{\gamma f}(x - c) + c.$$

Proposed algorithm

Context Fix a real number $\gamma > 0$ along with a sequence $(\lambda_n)_{n\in\mathbb{N}}$ in $(0,2)$. Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{F}$ and $\mathcal{G}$, respectively. Refer to **P** as well as its context. Suppose that

- the problem has at least one solution, (y^*, z^*);
- $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$;
- $\sum_{n\in\mathbb{N}} \lambda_n(\|Aa_n\| + \|Bb_n\|) < +\infty$.

Iterations Take any x_0 in $\mathcal{H}$, and for every n in $\mathbb{N}$, repeat

$$\begin{aligned} z_n &= \operatorname{prox}^{-B}_{\gamma g}(x_n - e) + b_n, \\ y_n &= \operatorname{prox}^{A}_{\gamma f}(2(-Bz_n + e) - x_n + d) + a_n, \quad \text{and} \\ x_{n+1} &= x_n + \lambda_n(Ay_n + Bz_n - c). \end{aligned}$$

Convergence It follows that

- $(x_n)_{n\in\mathbb{N}}$ converges weakly to some point x in $\mathcal{H}$,
- $\operatorname{prox}^{-B}_{\gamma g}(x - e) = z^*$ and $\operatorname{prox}^{A}_{\gamma f}(2(-Bz^* + e) - x + d) = y^*$,

and if $\mathcal{H}$ is finite dimensional and $a_n \to 0$ and $b_n \to 0$, then

- $y_n \to y^*$ and $z_n \to z^*$.

Recovering ADMM

How? Looking at the iterations of the proposed algorithm, we can define

$$u_n = \frac{1}{\gamma}(x_n + Bz_n - e), \quad n \ge 1,$$

and allow u_0 and z_0 to be chosen freely, fixing x_0 and b_0. Then, we can start the procedure at y_n and update u_n.

Why? Beyond a connection to the augmented Lagrangian when $\lambda_n = 1$ (and a merely superficial connection at that), there is no apparent advantage to ADMM.

Iterations Set (u_0, z_0) in $\mathcal{H}\times\mathcal{G}$. For every n in $\mathbb{N}$, repeat

$$y_n = \operatorname*{arg\,min}_{y\in\mathcal{F}}\Big(f(y) + \frac{1}{2\gamma}\|Ay + Bz_n - c + \gamma u_n\|^2\Big) + a_n;$$

$$z_{n+1} = \operatorname*{arg\,min}_{z\in\mathcal{G}}\Big(g(z) + \frac{1}{2\gamma}\|Ay_n + Bz - c + \gamma u_n + (\lambda_n - 1)(Ay_n + Bz_n - c)\|^2\Big) + b_{n+1};$$

$$u_{n+1} = u_n + \frac{1}{\gamma}\big(Ay_n + Bz_{n+1} - c + (\lambda_n - 1)(Ay_n + Bz_n - c)\big).$$

Observation ADMM suffers from a major shortcoming:

- It requires extra memory, requiring both z_n and z_{n+1} for the u_n update under relaxation ($\lambda_n \neq 1$).

Conclusion

Our novel concept of proximity allows us to generalize DR, enhancing it beyond ADMM—both in theory and practice. Using our version of DR, we have recovered ADMM, giving it the primal context and general convergence result of our algorithm, and have thus revealed that existing algorithms actually converge more generally than previously reported. In fact, validating the conditions for convergence of many versions of ADMM now requires no detour through duality, and the convergence of many versions no longer requires finite dimensions or injective operators.

Take-home message — The ADMM formalism is a theoretical and practical disservice; there is merit in breaking away from it and paying closer attention to the Douglas-Rachford algorithm through generalized proximity.