ADMM for monotone operators: convergence analysis and rates

Ernö Robert Csetnek
University of Vienna

joint work with
Radu Ioan Boț (University of Vienna)

Splitting Algorithms, Modern Operator Theory and Applications (17w5030) -
Oaxaca, Mexico
September 17-22, 2017

Consider the optimization problem

$$\min_{x\in\mathcal{H}}\{f(x)+g(Lx)\} \quad (1)$$

- $f:\mathcal{H}\to\mathbb{R}\cup\{+\infty\}$ and $g:\mathcal{G}\to\mathbb{R}\cup\{+\infty\}$ are proper, convex and lower semicontinuous functions
- $\mathcal{H}$ and $\mathcal{G}$ are real Hilbert spaces
- $L:\mathcal{H}\to\mathcal{G}$ is a linear continuous operator.

If (1) has an optimal solution $\overline{x}\in\mathcal{H}$ and

$$0\in\operatorname{sqri}(\operatorname{dom} g-L(\operatorname{dom} f)),$$

the optimality conditions read:

$$0\in\partial f(\overline{x})+L^*\partial g(L\overline{x})$$

hence there exists $\overline{v}\in\mathcal{G}$ such that:

$$-L^*\overline{v}\in\partial f(\overline{x}) \text{ and } \overline{v}\in\partial g(L\overline{x}) \quad (2).$$

If (2) holds, $\overline{x}$ is solves (1) and $\overline{v}$ solves the Fenchel dual problem:

$$\max_{v\in\mathcal{G}}\{-f^*(-L^*v)-g^*(v)\}.$$

Methods for solving the optimization problem

$$\min_{x\in\mathcal{H}}\{f(x)+g(Lx)\} \quad (1).$$

Primal-dual splitting algorithms (Combettes, Chambolle, Pock, Condat, Vu, Pesquet, Boț, etc.):

$$\begin{aligned} x^{k+1} &= \operatorname{prox}_{\tau f}\left(x^k-\tau L^*(2y^k-y^{k-1})\right) \\ y^{k+1} &= \operatorname{prox}_{\sigma g^*}\left(y^k+\sigma Lx^{k+1}\right). \end{aligned}$$

- the nosnmooth functions are evaluated separately through their proximal operators

$$\begin{aligned} \operatorname{prox}_{\tau f}(x) &= \operatorname*{argmin}_{y\in\mathcal{H}}\left\{f(y)+\frac{1}{2\tau}\|y-x\|^2\right\} \\ &= (\mathrm{Id}+\tau\partial f)^{-1}(x). \end{aligned}$$

- the algorithm solves both primal and Fenchel dual problem

ADMM (alternating direction method of multipliers):

$$\begin{aligned} x^{k+1} &\in \operatorname*{argmin}_{x\in\mathbb{R}^n} L_c(x, z^k, y^k) = \operatorname*{argmin}_{x\in\mathbb{R}^n} \left\{ f(x) + \frac{c}{2}\|Lx - z^k + c^{-1}y^k\|^2 \right\} \\ z^{k+1} &= \operatorname*{argmin}_{z\in\mathbb{R}^m} L_c(x^{k+1}, z, y^k) = \operatorname*{argmin}_{z\in\mathbb{R}^m} \left\{ g(z) + \frac{c}{2}\|Lx^{k+1} - z + c^{-1}y^k\|^2 \right\} \\ y^{k+1} &= y^k + c(Lx^{k+1} - z^{k+1}). \end{aligned}$$

where

$$L_c(x, z, y) = f(x) + g(z) + \langle y, Lx - z\rangle + \frac{c}{2}\|Lx - z\|^2.$$

is the augmented Lagrangian associated to (1).

- Notice that the first minimization is not a proximal step, due to L.
- in very simple situations, like $f(x) = (1/2)\|x\|^2$, the first minimization requires $(\mathrm{Id} + L^*L)^{-1}$

Proximal ADMM overcomes these limitations: add some extra proximal terms in the above minimizations.

Main results presented in this talk:

- a unifying scheme: an algorithm for solving monotone inclusions which recovers many of the algorithms mentioned above
  - convergence analysis
  - several algorithms from the literature as special instances
- convergence rates for the iterates: variable metric techniques with strategies based on suitable choice of dynamical step sizes
  - convergence rates
  - several algorithms from the literature as particular cases

Problem formulation

The aim is to solve the primal monotone inclusion

$$\text{find } x \in \mathcal{H} \text{ such that } 0 \in Ax + (L^* \circ B \circ L)x + Cx,$$

together with its dual monotone inclusion

$$\text{find } v \in \mathcal{G} \text{ such that } \exists x \in \mathcal{H} : -L^*v \in Ax + Cx \text{ and } v \in B(Lx).$$

- $A : \mathcal{H} \rightrightarrows \mathcal{H}$ and $B : \mathcal{G} \rightrightarrows \mathcal{G}$ are maximally monotone operators
- $C : \mathcal{H} \to \mathcal{H}$ is η-cocoercive: $\langle x - y, Cx - Cy \rangle \geq \eta \|Cx - Cy\|^2$
- $L : \mathcal{H} \to \mathcal{G}$ is linear and continuous

We are looking for a primal-dual solution $(x, v) \in \mathcal{H} \times \mathcal{G}$:

$$-L^*v \in Ax + Cx \text{ and } v \in B(Lx).$$

Algorithm

$$
\begin{aligned}
x^{k+1} &= \left(cL^*L + M_1^k + A\right)^{-1}\left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k - Cx^k\right] \\
z^{k+1} &= \left(\mathrm{Id} + c^{-1}M_2^k + c^{-1}B\right)^{-1}\left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k\right] \\
y^{k+1} &= y^k + c(Lx^{k+1} - z^{k+1}).
\end{aligned}
$$

- $M_1^k \in \mathcal{S}_+(\mathcal{H})$, $M_2^k \in \mathcal{S}_+(\mathcal{G})$ for all k
- $\mathcal{S}_+(\mathcal{H})$: the operators $U : \mathcal{H} \to \mathcal{H}$ which are linear, continuous, self-adjoint and positive semidefinite
- $cL^*L + M_1^k \in \mathcal{P}_{\alpha_k}(\mathcal{H})$ for all k, with $\alpha_k > 0$
- $\mathcal{P}_\alpha(\mathcal{H}) := \{U \in \mathcal{S}_+(\mathcal{H}) : U \succcurlyeq \alpha\,\mathrm{Id}$ i.e. $\langle Ux, x\rangle \geq \alpha\|x\|^2\ \forall x \in \mathcal{H}\}$.

Algorithm

$$\begin{aligned} x^{k+1} &= \left(cL^*L + M_1^k + A\right)^{-1}\left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k - Cx^k\right] \\ z^{k+1} &= \left(\mathrm{Id} + c^{-1}M_2^k + c^{-1}B\right)^{-1}\left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k\right] \\ y^{k+1} &= y^k + c(Lx^{k+1} - z^{k+1}). \end{aligned}$$

The algorithm is well defined:

$$\text{if } U \in \mathcal{P}_\alpha(\mathcal{H})\ \alpha > 0, A : \mathcal{H} \rightrightarrows \mathcal{H} \text{ maximally monotone,}$$

then:

$$\forall\ x \in \mathcal{H},\ \exists! p \in \mathcal{H} \text{ such that } p = (U + A)^{-1}x.$$

This follows from

$$(U + A)^{-1} = (\mathrm{Id} + U^{-1}A)^{-1} \circ U^{-1}$$

and

$$U^{-1}A \text{ is maximally monotone in } (\mathcal{H}, \langle\cdot,\cdot\rangle_U)$$

where

$$\langle x, y\rangle_U := \langle x, Uy\rangle\ \forall x, y \in \mathcal{H}.$$

The first two relations are equivalent to

$$0 \in A(x^{k+1}) + cL^*(Lx^{k+1} - z^k + c^{-1}y^k) + M_1^k(x^{k+1} - x^k) + C(x^k),$$
$$0 \in Bz^{k+1} + c(-Lx^{k+1} + z^{k+1} - c^{-1}y^k) + M_2^k(z^{k+1} - z^k).$$

Particular cases: Proximal ADMM and classical ADMM

Take the variational case

$$A = \partial f, B = \partial g \text{ and } C = \nabla h.$$

$$0 \in \partial f(x^{k+1}) + cL^*(Lx^{k+1} - z^k + c^{-1}y^k) + M_1^k(x^{k+1} - x^k) + \nabla h(x^k)$$

is equivalent to

$$x^{k+1} = \operatorname*{argmin}_{x \in \mathcal{H}} \left\{ f(x) + \langle x - x^k, \nabla h(x^k) \rangle + \frac{c}{2} \|Lx - z^k + c^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|^2_{M_1^k} \right\}.$$

while

$$0 \in \partial g(z^{k+1}) + c(-Lx^{k+1} + z^{k+1} - c^{-1}y^k) + M_2^k(z^{k+1} - z^k)$$

is equivalent to

$$z^{k+1} = \operatorname*{argmin}_{z \in \mathcal{G}} \left\{ g(z) + \frac{c}{2}\|Lx^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2}\|z - z^k\|^2_{M_2^k} \right\}.$$

This particular case leads to

$$x^{k+1} = \operatorname{argmin}_{x\in\mathcal{H}} \left\{ f(x)+\langle x-x^k, \nabla h(x^k)\rangle + \frac{c}{2}\|Lx-z^k+c^{-1}y^k\|^2 + \frac{1}{2}\|x-x^k\|^2_{M_1^k} \right\}$$

$$z^{k+1} = \operatorname{argmin}_{z\in\mathcal{G}} \left\{ g(z) + \frac{c}{2}\|Lx^{k+1} - z + c^{-1}y^k\|^2 + \frac{1}{2}\|z - z^k\|^2_{M_2^k} \right\}$$

$$y^{k+1} = y^k + c(Lx^{k+1} - z^{k+1}),$$

in connection with

$$\min_{x\in\mathcal{H}}\{f(x) + g(Lx) + h(x)\} \quad (1).$$

- $h = 0$ and $M_1^k = M_2^k = 0$ leads to the **classical ADMM**
- $h = 0$ and M_1^k, M_2^k constant leads to the **Proximal ADMM**: Shefi-Teboulle 2014, Toh, Sun, etc.
- the general case as above: Banert, Boț, C., 2017

The role of M_1

A special choice of M_1 induces a proximal step in the minimization

$$x^{k+1} = \text{argmin}_{x\in\mathcal{H}} \left\{ f(x) + \langle x - x^k, \nabla h(x^k)\rangle + \frac{c}{2}\|Lx - z^k + c^{-1}y^k\|^2 + \frac{1}{2}\|x - x^k\|^2_{M_1^k} \right\}$$

Take

$$M_1^k := \frac{1}{\tau}\,\text{Id} - cL^*L \text{ for } \tau > 0$$

then one obtains the proximal step:

$$x^{k+1} = (\text{Id} + \tau\partial f)^{-1}\left[\tau cL^*(z^k - c^{-1}y^k) + x^k - \tau cL^*Lx^k - \tau\nabla h(x^k)\right].$$

Primal-dual algorithms as special cases

Algorithm

$$x^{k+1} = \left(cL^*L + M_1^k + A\right)^{-1}\left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k - Cx^k\right]$$

$$z^{k+1} = \left(\mathrm{Id} + c^{-1}M_2^k + c^{-1}B\right)^{-1}\left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k\right]$$

$$y^{k+1} = y^k + c(Lx^{k+1} - z^{k+1}).$$

For $M_1^k := \frac{1}{\tau}\,\mathrm{Id} - cL^*L$ for $\tau > 0$ and $M_2^k = 0$ we get

$$y^{k+1} = J_{cB^{-1}}\left(y^k + cLx^{k+1}\right)$$

$$x^{k+2} = J_{\tau A}\left(x^{k+1} - \tau Cx^{k+1} - \tau L^*(2y^{k+1} - y^k)\right)$$

- Vũ 2013
- the case $C = 0$: Boț, C., Heinrich 2013
- the variational case: Condat 2013, Cambolle-Pock 2011

Algorithm

$$
\begin{aligned}
x^{k+1} &= \left(cL^*L + M_1^k + A\right)^{-1}\left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k - Cx^k\right] \\
z^{k+1} &= \left(\mathrm{Id} + c^{-1}M_2^k + c^{-1}B\right)^{-1}\left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k\right] \\
y^{k+1} &= y^k + c(Lx^{k+1} - z^{k+1}).
\end{aligned}
$$

Convergence result: assume

- the set of primal-dual solutions is nonempty
- $M_1^k - \frac{1}{2\eta}\,\mathrm{Id} \in \mathcal{S}_+(\mathcal{H})$
- $M_1^k \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$

Suppose that one of the following assumptions holds:

(I) $M_1^k - \frac{1}{2\eta}\,\mathrm{Id} \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ with $\alpha_1 > 0$ for all $k \geq 0$;

(II) $L^*L \in \mathcal{P}_{\alpha}(\mathcal{H})$ and $M_2^k \in \mathcal{P}_{\alpha_2}(\mathcal{G})$, $\alpha, \alpha_2 > 0$ for all $k \geq 0$.

Then $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to (x, Lx, v), where $-L^*v \in Ax + Cx$ and $v \in B(Lx)$.

The case $C = 0$

Algorithm

$$\begin{aligned} x^{k+1} &= \left(cL^*L + M_1^k + A\right)^{-1}\left[cL^*(z^k - c^{-1}y^k) + M_1^k x^k\right] \\ z^{k+1} &= \left(\mathrm{Id} + c^{-1}M_2^k + c^{-1}B\right)^{-1}\left[Lx^{k+1} + c^{-1}y^k + c^{-1}M_2^k z^k\right] \\ y^{k+1} &= y^k + c(Lx^{k+1} - z^{k+1}). \end{aligned}$$

Convergence result: assume

- the set of primal-dual solutions is nonempty
- $M_1^k \in \mathcal{S}_+(\mathcal{H}), M_1^k \succcurlyeq M_1^{k+1}$, $M_2^k \in \mathcal{S}_+(\mathcal{G}), M_2^k \succcurlyeq M_2^{k+1}$

Suppose that one of the following assumptions holds:

(I) $M_1^k \in \mathcal{P}_{\alpha_1}(\mathcal{H})$ with $\alpha_1 > 0$ for all $k \geq 0$;

(II) $L^*L \in \mathcal{P}_{\alpha}(\mathcal{H})$ and $M_2^k \in \mathcal{P}_{\alpha_2}(\mathcal{G})$, $\alpha, \alpha_2 > 0$ for all $k \geq 0$,

(III) $L^*L \in \mathcal{P}_{\alpha}(\mathcal{H})$ with $\alpha > 0$ and $2M_2^{k+1} \succcurlyeq M_2^k \succcurlyeq M_2^{k+1}$ for all k.

Then $(x^k, z^k, y^k)_{k\geq 0}$ converges weakly to (x, Lx, v), where $-L^*v \in Ax$ and $v \in B(Lx)$.

The role of variable M_2^k

induces dynamic step sizes in the algorithm and allows to accelerate the convergence behavior.

Algorithm

(accelerated version)

$$
\begin{aligned}
y^{k+1} &= \left(\tau_k LL^* + M_2^k + B^{-1}\right)^{-1}\left[-\tau_k L(z^k - \tau_k^{-1}x^k) + M_2^k y^k\right] \\
z^{k+1} &= \left(\frac{\theta_k}{\lambda} - 1\right) L^* y^{k+1} + \frac{\theta_k}{\lambda} C x^k \\
&\quad + \frac{\theta_k}{\lambda}\left(\mathrm{Id} + \lambda\tau_{k+1}^{-1} A^{-1}\right)^{-1}\left[-L^* y^{k+1} + \lambda\tau_{k+1}^{-1} x^k - C x^k\right] \\
x^{k+1} &= x^k + \frac{\tau_{k+1}}{\theta_k}\left(-L^* y^{k+1} - z^{k+1}\right),
\end{aligned}
$$

where

- $\lambda, \tau_k, \theta_k > 0$ *for all* $k \geq 0$
- $\tau_k LL^* + M_2^k \in \mathcal{P}_{\alpha_k}(\mathcal{G})$ *for* $\alpha_k > 0$ *for all* $k \geq 0$.

Particular instances (accelerated primal-dual algorithms)

The choice

$$\tau_k LL^* + M_2^k = \sigma_k^{-1}\,\mathrm{Id} \;\; \forall k \geq 0$$

leads to

$$\begin{aligned} x^{k+1} &= J_{(\tau_{k+1}/\lambda)A}\left[x^k + \frac{\tau_{k+1}}{\lambda}\left(-L^* y^{k+1} - Cx^k\right)\right] \\ y^{k+2} &= J_{\sigma_{k+1}B^{-1}}\left[y^{k+1} + \sigma_{k+1}L\left(x^{k+1} + \theta_{k+1}(x^{k+1} - x^k)\right)\right] \end{aligned}$$

- Boț, C., Heinrich, Hendrich 2015
- variational case and $C = 0$: Chambolle-Pock 2011

Particular instances (accelerated proximal ADMM)

The variational case

$$A = \partial f, B = \partial g, C = 0$$

leads to

$$\begin{aligned}
y^{k+1} &= \operatorname*{argmin}_{y \in \mathcal{G}} \left[g^*(y) + \frac{\tau_k}{2} \left\| L^* y + z^k - \tau_k^{-1} x^k \right\|^2 + \frac{1}{2} \|y - y^k\|^2_{M_2^k} \right] \\
z^{k+1} &= \theta_k \operatorname*{argmin}_{z \in \mathcal{H}} \left[f^*(z) + \frac{\tau_{k+1}}{2} \left\| -L^* y^{k+1} - z + \tau_{k+1}^{-1} x^k \right\|^2 \right] \\
&\quad + (\theta_k - 1) L^* y^{k+1} \\
x^{k+1} &= x^k + \frac{\tau_{k+1}}{\theta_k} \left(-L^* y^{k+1} - z^{k+1} \right).
\end{aligned}$$

$\mathcal{O}(\frac{1}{k})$ convergence rate for the sequence $(x^k)_{k\in\mathbb{N}}$

Assume

- the set of primal-dual solutions is nonempty
- $A+C$ is γ-strongly monotone, $\gamma > 0$
- $\mu\tau_1 < 2\gamma$, $\lambda \geq \mu + 1$, $\sigma_0\tau_1\|L\|^2 \leq 1$,
- $\theta_k = \frac{1}{\sqrt{1+\tau_{k+1}\lambda^{-1}(2\gamma-\mu\tau_{k+1})}}$ for all k
- $\tau_{k+2} = \theta_k\tau_{k+1}$, $\sigma_{k+1} = \theta_k^{-1}\sigma_k$ for all k
- $\tau_k LL^* + M_2^k \succcurlyeq \sigma_k^{-1}\,\mathrm{Id}$ for all k
- $\frac{\tau_k}{\tau_{k+1}}LL^* + \frac{1}{\tau_{k+1}}M_2^k \succcurlyeq \frac{\tau_{k+1}}{\tau_{k+2}}LL^* + \frac{1}{\tau_{k+2}}M_2^{k+1}$ for all k.

Then there exists $\widetilde{c} > 0$ such that

$$\|x^k - x\| \leq \frac{\widetilde{c}}{k} \ \forall k \geq 2,$$

where x is the unique solution of the inclusion:

$$0 \in Ax + (L^* \circ B \circ L)x + Cx.$$

References

[1] R.I. Boţ, E.R. Csetnek, *ADMM for monotone operators: convergence analysis and rates*, arXiv:1705.01913 , 2017

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Foundations and Trends in Machine Learning 3, 1–12, 2010

[3] J. Eckstein, D.P. Bertsekas, *On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming 55, 293–318, 1992

[4] R. Shefi, M. Teboulle, *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM Journal on Optimization 24(1), 269–297, 2014

[5] S. Banert, R.I. Boţ, E.R. Csetnek, *Fixing and extending some recent results on the ADMM algorithm*, arXiv:1612.05057, 2016

[6] A. Chambolle, T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision 40(1), 120–145, 2011

[7] L. Condat, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications 158(2), 460–479, 2013

[8] B.C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators*, Advances in Computational Mathematics 38(3), 667–681, 2013

[9] R.I. Boţ, E.R. Csetnek, A. Heinrich, C. Hendrich, *On the convergence rate improvement of a primal-dual splitting algorithm for solving monotone inclusion problems*, Mathematical Programming 150(2), 251–279, 2015