Convergence of Nonconvex Douglas-Rachford Splitting and Nonconvex ADMM

Shuvomoy Das Gupta

Thales Canada - Transportation Solutions

56th Allerton Conference on Communication, Control and Computing, 2018

Outline

What is this talk about?

this talk is about ADMM and Douglas-Rachford splitting for nonconvex problems

- the alternating direction method of multipliers (ADMM)
  - originally designed to solve convex optimization problem
- Douglas-Rachford splitting algorithm
  - ADMM is its special case in a convex setup
- both guaranteed to converge for convex problems

Motivation

- nonconvex ADMM (**NC-ADMM**)— a heuristic based on ADMM — has become a popular heuristic to tackle nonconvex problems
- recently, NC-ADMM heuristic has been applied to
  - [Erseghe, 2014] optimal power flow problem, [Takapoui *et al.*, 2017] mixed integer quadratic optimization, [Iyer *et al.*, 2014] submodular minimization with nonconvex constraints ...
- [Diamond *et al.*, 2018] Python package `NCVX` implements ADMM heuristic (NC-ADMM)
  - extension of `CVXPY`: used by fortune 500 companies
  - NC-ADMM often produces lower objective values compared with exact solvers within a time limit
- nonconvex Douglas-Rachford splitting (**NC-DRS**): analogous nonconvex heuristic based on Douglas-Rachford splitting
- not much has been done to improve the theoretical understanding of such heuristics

Summary of the results

- NC-DRS
  - attacks the original problem directly
  - optimal solutions can be characterized via the NC-DRS operator
  - if deviation from a convex setup is bounded $\Rightarrow$ it will converge or oscillate in a compact connected set
- NC-ADMM
  - works on a modified dual problem, not the original nonconvex problem
  - not equivalent to NC-DRS, but there is a relationship between them
  - likely to produce a lower objective value

Problem in consideration

- minimize a convex cost function with nonconvex constraint set

$$\begin{array}{ll} \text{minimize}_x & f(x) \\ \text{subject to} & x \in \mathcal{C} \end{array} \tag{OPT}$$

- f is closed, proper, and convex
- $\mathcal{C}$ is compact, but not necessarily convex

Reformulation through indicator function

- indicator function of set $\mathcal{C}$:

$$\delta_{\mathcal{C}}(x) = \begin{cases} 0, & \text{if } x \in \mathcal{C} \\ \infty, & \text{if } x \notin \mathcal{C} \end{cases}$$

- $\delta_{\mathcal{C}}$ is closed and proper, but not necessarily convex
- we can write

$$\begin{pmatrix} \text{minimize}_x & f(x) \\ \text{subject to} & x \in \mathcal{C} \end{pmatrix} = \text{minimize}_x \; f(x) + \delta_{\mathcal{C}}(x)$$

NC-ADMM and NC-DRS:

- NC-ADMM:

$$
\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma f}(y_n - z_n) \\
y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}(x_{n+1} + z_n) \\
z_{n+1} &= z_n - y_{n+1} + x_{n+1}
\end{aligned}
$$

- NC-DRS:

$$
\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma f}(z_n) \\
y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}(2x_{n+1} - z_n) \\
z_{n+1} &= z_n + y_{n+1} - x_{n+1}
\end{aligned}
$$

- both have same subroutines, but different inputs

Proximal operator of f and projection onto $\mathcal{C}$

- both NC-DRS and NC-ADMM have same subroutines: first $\mathbf{prox}_{\gamma f}$, then $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ and finally $\sum$
- proximal operator of f evaluated at point x with parameter $\gamma > 0$:

$$\mathbf{prox}_{\gamma f}(x) = \operatorname{argmin}_y \left(f(y) + \frac{1}{2\gamma}\|y - x\|^2\right)$$

  - single-valued, continuous

- projection onto $\mathcal{C}$:

$$\mathbf{prox}_{\gamma \delta_{\mathcal{C}}}(x) = \mathbf{\Pi}_{\mathcal{C}}(x) = \operatorname{argmin}_{y \in \mathcal{C}} \left(\|y - x\|^2\right)$$

  - there can be multiple projections
  - one such projection is denoted by $\tilde{\mathbf{\Pi}}_{\mathcal{C}}(\cdot)$

When $\mathcal{C}$ is convex

- f is closed, proper, convex
- $\mathcal{C}$: convex $\Rightarrow x_n, y_n$ converge to an optimal solution for any initial condition
- but $\mathcal{C}$ is not necessarily convex in our setup
  - convergence conditions are messy

Why convergence conditions are messy?

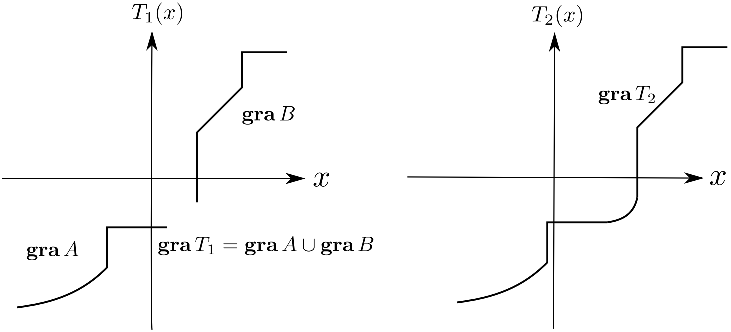
the convergence conditions are messy because:

- subdifferential operator of $\delta_{\mathcal{C}}$ is *monotone*, but ***not*** *maximally monotone*
- $\Rightarrow \tilde{\mathbf{\Pi}}_{\mathcal{C}}$: is *expansive i.e.,* ***not*** *nonexpansive*
- $\Rightarrow$ the underlying *reflection operator* is *expansive*

Outline

Monotone and maximally monotone operators

- T is monotone if for every $(x,u),(y,v) \in \mathbf{gra} T$
$$\langle x-y \mid u-v \rangle \geq 0$$
- T is maximally monotone if $\mathbf{gra}\, T$ is not properly contained by any other monotone operator's graph



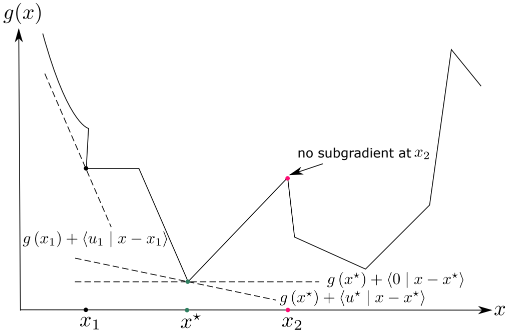
monotone, but not maximally monotone

maximally monotone

Subdifferential operator

- g: closed, proper, but not necessarily convex
- ∂g: subdifferential of g is monotone, but not maximally monotone

$$\partial g(x) = \{u \in \mathbf{R}^n \mid (\forall y \in \mathbf{R}^n)\; g(y) \geq g(x) + \langle u \mid y - x \rangle\}$$



Why convergence conditions are messy?

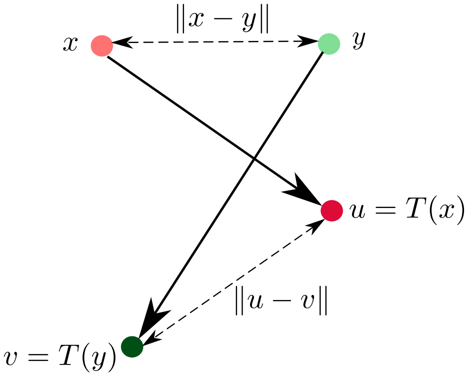
- our problem: $\mathrm{minimize}_x \, f(x) + \delta_{\mathcal{C}}(x)$
- ∂f: maximally monotone
- $\partial \delta_{\mathcal{C}}$: monotone, but **not** maximally monotone

  $\Rightarrow \tilde{\mathbf{\Pi}}_{\mathcal{C}}$: is *expansive (i.e., **not** nonexpansive)*

What is a nonexpansive operator?

T : single-valued operator on $\mathbf{R}^n$

- T is nonexpansive on $\mathbf{R}^n$ if for every x, y and $u = T(x), v = T(y)$ we have $\|u - v\| \le \|x - y\|$



Firmly nonexpansive operator

T : single-valued operator on $\mathbf{R}^n$

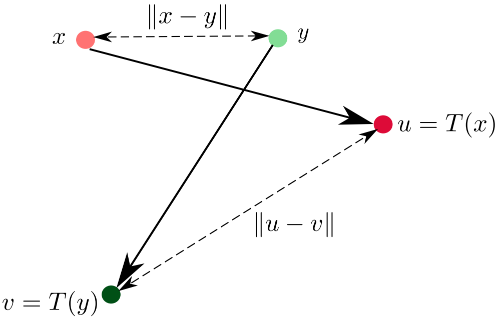
- T is firmly nonexpansive on $\mathbf{R}^n$ if $(2T - I_n)$ is nonexpansive
- a firmly nonexpansive operator is also nonexpansive
- $\mathbf{prox}_{\gamma f}$ is firmly nonexpansive

Operators that are expansive

T is a single-valued operator on $\mathbf{R}^n$

- T is expansive if there exist x, y and $u = T(x), v = T(y)$ such that

$$\|u - v\| > \|x - y\|$$



- $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ is expansive

Outline

Compact form NC-DRS

- NC-DRS:

$$\begin{aligned} x_{n+1} &= \mathbf{prox}_{\gamma f}(z_n) \\ y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}(2x_{n+1} - z_n) \\ z_{n+1} &= z_n + y_{n+1} - x_{n+1}. \end{aligned}$$

- $\tilde{T}$: nonconvex Douglas-Rachford operator

$$\tilde{T} = \tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(2\mathbf{prox}_{\gamma f} - I_n\right) + I_n - \mathbf{prox}_{\gamma f}.$$

- $\tilde{R}$: nonconvex reflection operator of $\tilde{T}$

$$\tilde{R} = 2\tilde{T} - I_n$$

- NC-DRS in compact form: $z_{n+1} = \tilde{T} z_n = \frac{1}{2}\left(\tilde{R} + I_n\right) z_n$
- $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$: expansive $\Rightarrow$ $\tilde{R}$: expansive $\Rightarrow$ root of all convergence issues

Characterization of minimizers

- $\operatorname{argmin}(f+\delta_{\mathcal{C}})$ is the set of minimizers of $\min._x f(x)+\delta_{\mathcal{C}}(x)$

$$\boxed{\mathbf{prox}_{\gamma f}(\mathbf{fix}\,\tilde{T}) \subseteq \operatorname{argmin}(f+\delta_{\mathcal{C}})}$$

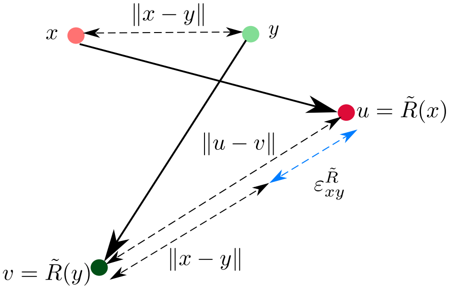
- underlying assumptions:
  1. $\mathbf{zer}\left(\partial f+\partial\delta_{\mathcal{C}}\right)$ is nonempty
  2. $\mathbf{fix}\,\tilde{T}$ is nonempty
  3. $\mathbf{fix}\left\{\left(2\mathbf{\Pi}_{\mathcal{C}}-I_n\right)\left(2\mathbf{prox}_{\gamma f}-I_n\right)\right\}$ is nonempty

Outline

Setup: expansiveness of $\tilde{R}$

- $\varepsilon_{xy}^{\tilde{R}}$: expansiveness of $\tilde{R}$ at x, y

$$\varepsilon_{xy}^{\tilde{R}} = \begin{cases} \|\tilde{R}(x) - \tilde{R}(y)\| - \|x - y\|, & \text{if } \|x - y\| < \|\tilde{R}(x) - \tilde{R}(y)\| \\ 0, & \text{else} \end{cases}$$



- $\sigma_{xy}^{\tilde{R}} = \sqrt{\varepsilon_{xy}^{\tilde{R}}}\sqrt{\|\tilde{R}(x) - \tilde{R}(y)\| + \|x - y\|}$

Conditions

- $(z_n)_{n\in\mathbf{N}}$: sequence of vectors generated for some chosen initial point z_0

if the following holds:

- there exists a $z \in \mathbf{fix}\,\tilde{T}$, such that $\sum_{n=0}^{\infty}\left(\sigma_{z_n z}^{\tilde{R}}\right)^2$ is bounded above, and $\|z_0 - z\|^2$ is finite
  - define $r := \sqrt{\|z_0 - z\|^2 + \sum_{n=0}^{\infty}\left(\sigma_{z_n z}^{\tilde{R}}\right)^2}$
  - $B(z;r)$: compact ball with center z and radius r

then...

Convergence results

then one of the following will happen:

1. **convergence to a point:** the sequence $(z_n)_{n\in\mathbf{N}}$ converges to a point $z^\star \in B(z;r)$
2. **cluster points form a continuum:** the set of cluster points of $(z_n)_{n\in\mathbf{N}}$ forms a nonempty compact connected set in $B(z;r)$

if situation 1 occurs and $\underline{\lim}_{n\to\infty}\left(\sigma^{\tilde{R}}_{z_n z^\star}\right)^2 = 0$, then $x_n = \mathbf{prox}_{\gamma f}\left(z_{n-1}\right)$ converges to an optimal solution

Some comments on convergence

- for convergence total deviation of $\tilde{R}$ from being a nonexpansive operator over the sequence $\{(z_n, z)\}_{n\in\mathbf{N}}$ is bounded
- depends on the initial point
- in our case $\mathcal{C}$ is not necessarily convex
- but, if we take $\mathcal{C}$ to be convex then
  - total deviation of $\tilde{R}$ from being a nonexpansive operator over the sequence $\{(z_n, z)\}_{n\in\mathbf{N}}$ is zero
  - our convergence proof coincides with known convergence results for convex setup

Outline

Constructing NC-ADMM

- original problem: $\text{minimize}_{x \in \mathcal{C}}\ f(x)$
- take dual and apply NC-DRS to the dual
- resultant algorithm is relaxed NC-ADMM

$$
\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma f}(y_n - z_n) \\
y_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}(z_n + x_{n+1}) \\
z_{n+1} &= z_n - y_{n+1} + x_{n+1}
\end{aligned}
$$

- relaxed NC-ADMM solves $\text{minimize}_{x \in \mathbf{conv}\, C}\ f(x)$

Constructing NC-ADMM (continued)

- relaxed NC-ADMM

$$
\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma f}\left(y_n - z_n\right)\\
y_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}\left(z_n + x_{n+1}\right)\\
z_{n+1} &= z_n - y_{n+1} + x_{n+1}
\end{aligned}
$$

  - solves $\mathrm{minimize}_{x \in \mathbf{conv} C}\; f(x)$
- replace $\mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}}$ with $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ to solve $\mathrm{minimize}_{x \in \mathcal{C}}\; f(x)$
- resultant algorithm is NC-ADMM:

$$
\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma f}\left(y_n - z_n\right)\\
y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}\left(x_{n+1} + z_n\right)\\
z_{n+1} &= z_n - y_{n+1} + x_{n+1}
\end{aligned}
$$

Constructing NC-ADMM

$$\text{minimize}_{x \in \mathcal{C}} \;\; (f(x))$$

original problem

take dual

$$\text{maximize}_{\nu} \;\; \left(-f^{\star\vee}(\nu) - \delta_{\mathcal{C}}^{\star}(\nu)\right)$$

apply NC-DRS

$$\begin{aligned}
\zeta_{n+1} &= \mathbf{prox}_{\gamma \delta_{\mathcal{C}}^{\star}}(\psi_n) \\
\xi_{n+1} &= \mathbf{prox}_{\gamma f^{\star\vee}}(2\zeta_{n+1} - \psi_n) \\
\psi_{n+1} &= \psi_n + \xi_{n+1} - \zeta_{n+1}
\end{aligned}$$

$$\delta_{\mathcal{C}}^{\star\star} = \delta_{\mathbf{conv}\, \mathcal{C}}$$

$$\mathbf{prox}_{\gamma \delta_{\mathcal{C}}^{\star}}(x) = x - \gamma \mathbf{\Pi}_{\mathbf{conv}\, \mathcal{C}}(\gamma^{-1} x)$$

$$\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\tilde{\gamma} f}(y_n - z_n) \\
y_{n+1} &= \mathbf{\Pi}_{\mathbf{conv}\, \mathcal{C}}(z_n + x_{n+1}) \\
z_{n+1} &= z_n - y_{n+1} + x_{n+1}
\end{aligned}$$

relaxed NC-ADMM: solves $\text{minimize}_{x \in \mathbf{conv}\, \mathcal{C}} \;\; (f(x))$

replace $\mathbf{\Pi}_{\mathbf{conv}\, \mathcal{C}}$ with $\tilde{\mathbf{\Pi}}_{\mathcal{C}}$ to solve $\text{minimize}_{x \in \mathcal{C}} \;\; (f(x))$

$$\begin{aligned}
x_{n+1} &= \mathbf{prox}_{\gamma f}(y_n - z_n) \\
y_{n+1} &= \tilde{\mathbf{\Pi}}_{\mathcal{C}}(x_{n+1} + z_n) \\
z_{n+1} &= z_n - y_{n+1} + x_{n+1}
\end{aligned}$$

NC-ADMM

If $\mathcal{C}$ is convex

if $\mathcal{C}$ is convex

- strong duality holds
- $\mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}} = \tilde{\mathbf{\Pi}}_{\mathcal{C}}$
- Moreau's decomposition can be applied to $\delta_{\mathcal{C}}$:

$$\mathbf{prox}_{\delta_{\mathcal{C}}} + \mathbf{prox}_{\delta_{\mathcal{C}}^{\star}} = I_n$$

then:

- no information loss in constructing NC-ADMM from NC-DRS
- NC-ADMM and NC-DRS are equivalent to each other

However...

when $\mathcal{C}$ is nonconvex

- NC-ADMM and NC-DRS are not equivalent
- reasons for the loss of equivalency:
  1. strict duality gap
  2. $\mathbf{\Pi}_{\mathbf{conv}\,\mathcal{C}} \neq \tilde{\mathbf{\Pi}}_{\mathcal{C}}$
  3. Moreau's decomposition does not hold
- NC-ADMM works on a modified dual, hence produces a lower objective value
- there is no relationship between the minimizers of the original problem and the fixed point set of the NC-ADMM operator

What if?

- is it possible to establish convergence NC-ADMM ignoring NC-DRS?
- our convergence analysis is established for
  - nonempty, compact, but not necessarily convex constraint sets
  - it is equally applicable to the smaller subclass of problems with convex constraint sets
  - in this smaller subclass of convex problems NC-ADMM and NC-DRS have the same convergence properties

Summary

- NC-DRS and NC-ADMM are very similar looking, but very different heuristics
- NC-DRS
  - attacks the original problem directly
  - optimal solutions can be characterized via the NC-DRS operator
  - will converge or oscillate in if its expansiveness of the nonconvex reflection operator is bounded
- NC-ADMM
  - works on a modified dual problem, not equivalent to NC-DRS
  - likely to produce a lower objective value

Limitations

- strong assumptions in minimizer characterization
- sufficient conditions cannot be checked directly
- numerical experiments to compare the performance of NC-ADMM with NC-DRS
- adaptive step size is not considered

End of talk

- if you are working on nonconvex problems using ADMM/DRS, please talk to me!

Thank you!

Questions?

Outline

What is a maximally monotone operator?

- set-valued operator T: takes a vector x and outputs a set $T(x)$ in $\mathbf{R}^n$
  - if $u \in T(x)$ then we say (x,u) is in $\mathbf{gra}\, T$
  - $\mathbf{gra}\, T = \{(x,u) \mid u \in T(x)\}$
- T is monotone if for every $(x,u),(y,v) \in \mathbf{gra} T$

$$\langle x - y \mid u - v \rangle \geq 0$$

- T is maximally monotone if $\mathbf{gra}\, T$ is not properly contained by any other monotone operator's graph

Some comments on convergence

- total deviation of $\tilde{R}$ from being a nonexpansive operator over the sequence $\{(z_n, z)\}_{n\in\mathbf{N}}$ is bounded
- depends on the initial point
- in our case $\mathcal{C}$ is not necessarily convex
- but, if we take $\mathcal{C}$ to be convex then
  - total deviation of $\tilde{R}$ from being a nonexpansive operator over the sequence $\{(z_n, z)\}_{n\in\mathbf{N}}$ is zero
  - our convergence proof coincides with known convergence results for convex setup

What if?

- is it possible to establish convergence NC-ADMM ignoring NC-DRS?
- our convergence analysis is established for
  - nonempty, compact, but not necessarily convex constraint sets
  - it is equally applicable to the smaller subclass of problems with convex constraint sets
  - in this smaller subclass of convex problems NC-ADMM and NC-DRS have the same convergence properties

Proof sketch

- $(z_n)_{n\in\mathbf{N}}$ stays in the compact set $B(z;r)$
- $\underline{\lim}_{n\to\infty} \|\tilde{R}z_n - z_n\| = 0$
- $(z_n)_{n\in\mathbf{N}}$
  1. either converges to a point
  2. or its set of cluster points forms a nonempty closed and connected set in $B(z;r)$
- if situation 1 occurs and $\underline{\lim}_{n\to\infty} \sigma^2_{z_n z^\star} = 0$, then x_n converges to an optimal solution

Diamond, S, Takapoui, R, & Boyd, S. 2018.
A general system for heuristic minimization of convex functions over non-convex sets.
Optimization Methods and Software, **33**(1), 165–193.

Erseghe, Tomaso. 2014.
Distributed optimal power flow using ADMM.
IEEE transactions on power systems, **29**(5), 2370–2380.

Iyer, Rishabh, Jegelka, Stefanie, & Bilmes, Jeff. 2014.
Monotone Closure of Relaxed Constraints in Submodular Optimization: Connections Between Minimization and Maximization.
Pages 360–369 of: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence.
UAI'14.
Arlington, Virginia, United States: AUAI Press.

Takapoui, Reza, Moehle, Nicholas, Boyd, Stephen, & Bemporad, Alberto. 2017.
A simple effective heuristic for embedded mixed-integer quadratic programming.
International Journal of Control, 1–11.