20 Reflectors

Goal

Reflectors, Douglas-Rachford, Peaceman-Rachford, primal-dual convergence, convergence rate, ADMM.

Alert 20.1: Convention

We continue our discussion on splitting methods and focus on reflectors.

Gray boxes are not required hence can be omitted for unenthusiastic readers.

This note is likely to be updated again soon.

We use the ordering $R_{\mathsf{B}}R_{\mathsf{A}}$ to be consistent with FB splitting.

Definition 20.2: Reflector

Recall that the resolvent for a map T is defined as $J_{\mathsf{T}}^{\eta} = (\mathrm{Id} + \eta\mathsf{T})^{-1}$. Similarly, we define the reflector of T:

$$R_{\mathsf{T}}^{\eta} = 2J_{\mathsf{T}}^{\eta} - \mathrm{Id} \subseteq (\mathrm{Id} - \eta\mathsf{T})(\mathrm{Id} + \eta\mathsf{T})^{-1}, \quad R_{\mathsf{T}} := R_{\mathsf{T}}^{1},$$

where the containment reduces to equality if T is monotone and single-valued (see Exercise 17.8). In other words, the reflector amounts to performing a backward (i.e. proximal) step $(\mathrm{Id}+\eta\mathsf{T})^{-1}$ followed by a forward (i.e. gradient) step $(\mathrm{Id} - \eta\mathsf{T})$.

We know from Theorem 17.6 that T is monotone iff its reflector R_{T} is nonexpansive (over its domain), and T is maximal monotone iff R_{T} is nonexpansive over the entire space. It is clear that an operator R is nonexpansive (over the entire space) iff $J := \frac{\mathrm{Id}+R}{2}$ is firmly nonexpansive (over the entire space) iff it is a reflector of some (maximal) monotone map (see Corollary 17.7). Obviously,

$$\mathbf{w} \in R_{\mathsf{T}}^{\eta}\mathbf{w} \iff \mathbf{w} \in J_{\mathsf{T}}^{\eta}\mathbf{w} \iff \mathbf{0} \in \mathsf{T}\mathbf{w}.$$

When $\mathsf{T} = \mathcal{N}_C$ is the normal cone of a closed convex set $C \subseteq \mathbb{R}^d$, the resolvent J_{T} is the familiar projection onto C while R_{T} is the reflection of the input w.r.t. C (hence explaining the name).

Alert 20.3: Non-commutativity

We remind that in general

$$(\mathrm{Id} - \eta\mathsf{T})(\mathrm{Id} + \eta\mathsf{T})^{-1} \neq (\mathrm{Id} + \eta\mathsf{T})^{-1}(\mathrm{Id} - \eta\mathsf{T}).$$

For instance, take $\mathsf{T} = \partial|\cdot|$ and $w = 1.5, \eta = 1$. However, when T is maximal monotone, we have

$$(\mathrm{Id} + \eta\mathsf{T})^{-1}(\mathrm{Id} - \eta\mathsf{T}) = (\mathrm{Id} + \eta\mathsf{T})^{-1}(\mathrm{Id} - \eta\mathsf{T})(\mathrm{Id} + \eta\mathsf{T})^{-1}(\mathrm{Id} + \eta\mathsf{T}).$$

Thus, upon an appropriate change-of-variable $\mathbf{w} \leftarrow (\mathrm{Id} + \eta\mathsf{T})\mathbf{w}$ we reduce the forward-backward map $(\mathrm{Id} + \eta\mathsf{T})^{-1}(\mathrm{Id} - \eta\mathsf{T})$ to the backward-forward map $(\mathrm{Id} - \eta\mathsf{T})(\mathrm{Id} + \eta\mathsf{T})^{-1}$. Conversely, if $\mathbf{z} = R_{\mathsf{T}}^{\eta}\mathbf{w} \in (\mathrm{Id} - \eta\mathsf{T})(\mathrm{Id} + \eta\mathsf{T})^{-1}\mathbf{w}$, then

$$J_{\mathsf{T}}^{\eta}\mathbf{z} \in (\mathrm{Id} + \eta\mathsf{T})^{-1}(\mathrm{Id} - \eta\mathsf{T})J_{\mathsf{T}}^{\eta}\mathbf{w}.$$

Proposition 20.4: Reflection-projection iteration (Bauschke and Kruk 2004)

Let $\mathsf{T} : \mathbb{R}^d \to \mathbb{R}^d$ be averaged and $\mathsf{R} : \mathbb{R}^d \to \mathbb{R}^d$ be nonexpansive and idempotent. Suppose $\mathsf{F} := \mathrm{Fix}\mathsf{T} \cap$

$\operatorname{Fix}\mathsf{R} \neq \emptyset$. *Then, the iterate*

$$\mathbf{w}_{t+1} := \mathsf{RT}\mathbf{w}_t + \boldsymbol{\epsilon}_t, \quad \textit{where} \quad \sum_t \|\boldsymbol{\epsilon}_t\|_2 < \infty,$$

converges to $\mathbf{w}_\infty \in \mathsf{F}$.

Proof: It is clear that $\{\mathbf{w}_t\}$ is quasi-Fejér monotone w.r.t. F: For any $\mathbf{w} \in \mathsf{F}$,

$$\|\mathbf{w}_{t+1} - \mathbf{w}\|_2 \leq \|\mathsf{T}\mathbf{w}_t - \mathbf{w}\|_2 + \|\boldsymbol{\epsilon}_t\|_2 \leq \|\mathbf{w}_t - \mathbf{w}\|_2 + \|\boldsymbol{\epsilon}_t\|_2.$$

Since T is say α-averaged, we have

$$\|\mathsf{T}\mathbf{w}_t - \mathbf{w}\|_2^2 + (\alpha^{-1} - 1)\|\mathbf{w}_t - \mathsf{T}\mathbf{w}_t\|_2^2 \leq \|\mathbf{w}_t - \mathbf{w}\|_2^2.$$

Combining the two inequalities, we know for some constant c (that may depend on $\mathbf{w}$),

$$(\alpha^{-1} - 1)\|\mathbf{w}_t - \mathsf{T}\mathbf{w}_t\|_2^2 \leq \|\mathbf{w}_t - \mathbf{w}\|_2^2 - \|\mathbf{w}_{t+1} - \mathbf{w}\|_2^2 + c\|\boldsymbol{\epsilon}_t\|_2,$$

whence follows that $\mathbf{w}_t - \mathsf{T}\mathbf{w}_t \to \mathbf{0}$. Let $\mathbf{w}_\infty$ be a limit point of $\{\mathbf{w}_t\}$. We clearly have $\mathbf{w}_\infty \in \operatorname{rge}\mathsf{R} = \operatorname{Fix}\mathsf{R}$ and $\mathbf{w}_\infty \in \operatorname{Fix}\mathsf{T}$. Therefore, the entire quasi-Fejér sequence $\{\mathbf{w}_t\}$ converges to $\mathbf{w}_\infty \in \mathsf{F}$. ■

We call an operator $\mathsf{R} : \mathbb{R}^d \to \mathbb{R}^d$ idempotent if $\mathsf{R}^2 := \mathsf{R} \circ \mathsf{R} = \mathsf{R}$. For such an operator R, $\operatorname{rge}\mathsf{R} = \operatorname{Fix}\mathsf{R}$. For instance, a convex cone K is obtuse (i.e. $\mathsf{K}^* \subseteq \mathsf{K}$) iff its reflector is idempotent. Obviously, orthogonal projectors are idempotent.

Bauschke and Kruk (2004) proved this theorem for the case where $\mathsf{T} = \prod_i \mathrm{P}_{C_i}$ and R is the reflector of an obtuse cone.

Needless to say, swapping the order to TR is immaterial and we obtain the same convergence. However, we cannot accommodate two or more idempotent operators: Let R_1 and R_2 be reflectors w.r.t. the y-axis. The resulting iterate $\mathbf{w}_{t+1} = \mathsf{R}_1\mathsf{R}_2\mathbf{w}_t$ halts in one iteration but halts at y-axis (i.e. fixed point) only if we start from there.

Bauschke, H. H. and S. G. Kruk (2004). "Reflection-Projection Method for Convex Feasibility Problems with an Obtuse Cone". *Journal of Optimization Theory and Applications*, vol. 120, no. 3, pp. 503–531.

Exercise 20.5: Idempotent reflector

Prove that

- the reflector R_K of a closed convex cone K is idempotent iff K is obtuse.
- if the reflector R_f^η is idempotent for all $\eta > 0$, then f must be the indicator of some closed convex set C. Does C have to be an obtuse convex cone?

Theorem 20.6: When reflector is contractive (Lions and Mercier 1979)

Let $\mathsf{T} : \mathbb{R}^d \to \mathbb{R}^d$ *be* σ*-maximal monotone and* L*-Lipschitz continuous. Then, its reflector* R_T *is* $\sqrt{1 - \frac{4\sigma}{(1+\mathsf{L})^2}}$*-contractive.*

Proof: For a σ-maximal monotone map T, we verify that

$$\begin{aligned}\|R_\mathsf{T}\mathbf{w} - R_\mathsf{T}\mathbf{z}\|_2^2 &= \|2(J_\mathsf{T}\mathbf{w} - J_\mathsf{T}\mathbf{z}) - (\mathbf{w} - \mathbf{z})\|_2^2 \\ &= -4\,\langle J_\mathsf{T}\mathbf{w} - J_\mathsf{T}\mathbf{z}; (\mathrm{Id} - J_\mathsf{T})\mathbf{w} - (\mathrm{Id} - J_\mathsf{T})\mathbf{z}\rangle + \|\mathbf{w} - \mathbf{z}\|_2^2 \qquad (20.1) \\ &\leq -4\sigma\|J_\mathsf{T}\mathbf{w} - J_\mathsf{T}\mathbf{z}\|_2^2 + \|\mathbf{w} - \mathbf{z}\|_2^2.\end{aligned}$$

When T is L-Lipschitz continuous, we have

$$\|\mathbf{w}-\mathbf{z}\|_2 \le \|J_{\mathsf{T}}\mathbf{w}-J_{\mathsf{T}}\mathbf{z}\|_2 + \|(\mathrm{Id}-J_{\mathsf{T}})\mathbf{w}-(\mathrm{Id}-J_{\mathsf{T}})\mathbf{z}\|_2 \le (1+\mathsf{L})\|J_{\mathsf{T}}\mathbf{w}-J_{\mathsf{T}}\mathbf{z}\|_2\,. \tag{20.2}$$

Combining the two inequalities completes the proof. ■

The equality (20.1) gives a more refined picture for any nonexpansion R_{T} (recall that the first term is nonpositive). The inequality (20.2) "inverts" the $(1+\mathsf{L})$-Lipschitz continuity of $\mathrm{Id}+\mathsf{T}$. Without L-Lipschitzness, the reflector R_{T} may not be contractive even when the resolvant J_{T} is, see Exercise 20.8 below.

It is clear that $\frac{\mathrm{Id}+NR_{\mathsf{T}}}{2}$ is $\sqrt{1-\frac{2\sigma}{(1+\mathsf{L})^2}}$-contractive for any nonexpansion N, since

$$\tfrac{1}{2}+\tfrac{1}{2}\sqrt{1-\tfrac{4\sigma}{(1+\mathsf{L})^2}} \le \sqrt{1-\tfrac{2\sigma}{(1+\mathsf{L})^2}}.$$

We know from Theorem 17.24 that J_{T} is $\frac{1}{1+\sigma}$-contractive. Note that it is possible for the reflector to be more contractive than the resolvant (but not when $\sigma \ge 2$)!

Lions, P.-L. and B. Mercier (1979). "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979.

Corollary 20.7: Contractive reflector $\implies$ strong convexity

Let $R_{\partial f}$ be the reflector of some convex function f. Then, $R_{\partial f}$ is L-contractive $\implies$ f is (at least) $\frac{1-\mathsf{L}}{1+\mathsf{L}}$-strongly convex.

Proof: $R_{\partial f}$ is L-contractive $\implies$ P_f is $\frac{1+\mathsf{L}}{2}$-contractive. Apply Theorem 17.24. ■

Exercise 20.8: Reflector may fail to be contractive even for strongly convex functions

Let $f(x)=\frac{4}{3}|x|^{3/2}$ for $|x|\le 1$. Verify the following:

- ∇f is not Lipschitz continuous.
- f is 1-strongly convex.
- $\mathrm{P}_f(x) = (x+2-2\sqrt{x+1})\wedge 1$ for $x \ge 0$ (and symmetric for $x \le 0$).
- $R_{\partial f}x = (x+4-4\sqrt{x+1})\wedge(2-x)$ for $x\ge 0$.
- $\mathrm{P}_f(x)$ is $(1-\frac{1}{\sqrt{2}})$-contractive while $R_{\partial f}$ is not contractive.

Algorithm 20.9: Splitting based on reflectors

For any map A and B we have the following implications:

$$\begin{aligned}\mathbf{0}\in(\mathsf{A}+\mathsf{B})\mathbf{z}=\mathsf{A}\mathbf{z}+\mathsf{B}\mathbf{z} &\iff \exists\mathbf{w}\in\mathbf{z}+\eta\mathsf{A}\mathbf{z},\ \text{equivalently } \mathbf{z}\in J^\eta_{\mathsf{A}}\mathbf{w} \quad \text{s.t.} \quad (\mathbf{z}+\eta\mathsf{B}\mathbf{z})\ni 2\mathbf{z}-\mathbf{w}\in R^\eta_{\mathsf{A}}\mathbf{w}\\ &\implies \exists\mathbf{w}\quad\text{s.t.}\quad \mathbf{z}\in J^\eta_{\mathsf{A}}\mathbf{w}\cap J^\eta_{\mathsf{B}}R^\eta_{\mathsf{A}}\mathbf{w} \iff \mathbf{z}\in J^\eta_{\mathsf{A}}\mathbf{w} \text{ and } 2\mathbf{z}-\mathbf{w}\in 2J^\eta_{\mathsf{B}}R^\eta_{\mathsf{A}}\mathbf{w}-\mathbf{w}\\ &\implies \exists\mathbf{w}\quad\text{s.t.}\quad \mathbf{w}\in R^\eta_{\mathsf{B}}R^\eta_{\mathsf{A}}\mathbf{w} \text{ and } \mathbf{z}\in J^\eta_{\mathsf{A}}\mathbf{w},\end{aligned} \tag{20.3}$$

whereas all are equivalent when A and B are monotone (so that both J and R are single-valued, resolving ambiguity). Naturally, we may attempt to apply fixed-point iteration to find $\mathbf{w}$ hence $\mathbf{z}$. We remark that although the equivalence (20.3) holds as long as A and B are monotone, we need both to be maximal monotone to guarantee the well-definedness of the iterates.

Following Varga (2009, p. 264) and Eckstein and Bertsekas (1992) we add relaxation and arrive at a general splitting algorithm based on reflectors:

Algorithm: A general splitting algorithm based on reflectors

Input: $\mathbf{w}_0$

for $t = 0, 1, \ldots$ **do**

  choose step size $\eta_t \geq 0$ and relaxation size $\gamma_t \in [0, 1]$

  $\mathbf{w}_{t+1} = (1-\gamma_t)\mathbf{w}_t + \gamma_t R_{\mathsf{B}}^{\eta_t} R_{\mathsf{A}}^{\eta_t}(\mathbf{w}_t) + \boldsymbol{\epsilon}_t$ // allow error $\boldsymbol{\epsilon}_t$

return $\mathbf{z} \in J_{\mathsf{A}}^{\eta}\mathbf{w}$ // assuming the for-loop returns $\mathbf{w}$ and $\eta_t \equiv \eta$

In particular, we have the following special cases:

- $\gamma_t \equiv 1$ reduces to the Peaceman-Rachford (PR) splitting algorithm studied initially by Peaceman and Rachford (1955). See also Kellogg (1969).
- $\gamma_t \equiv \frac{1}{2}$ reduces to the Douglas-Rachford (DR) splitting algorithm studied initially by Douglas and Rachford (1956). DR is exactly Krasnosel'skiĭ's averaging algorithm (see Remark 16.7) applied to PR.

Both PR and DR, in their current forms and generality, are due to Lions and Mercier (1979). Since

$$\|\hat{R}_{\mathsf{B}}^{\eta}\hat{R}_{\mathsf{A}}^{\eta} - R_{\mathsf{B}}^{\eta}R_{\mathsf{A}}^{\eta}\|_2 \leq \|\hat{R}_{\mathsf{B}}^{\eta}\hat{R}_{\mathsf{A}}^{\eta} - R_{\mathsf{B}}^{\eta}\hat{R}_{\mathsf{A}}^{\eta}\|_2 + \|R_{\mathsf{B}}^{\eta}\hat{R}_{\mathsf{A}}^{\eta} - R_{\mathsf{B}}^{\eta}R_{\mathsf{A}}^{\eta}\|_2 \leq \|\hat{R}_{\mathsf{B}}^{\eta}\hat{R}_{\mathsf{A}}^{\eta} - R_{\mathsf{B}}^{\eta}\hat{R}_{\mathsf{A}}^{\eta}\|_2 + \|\hat{R}_{\mathsf{A}}^{\eta} - R_{\mathsf{A}}^{\eta}\|_2$$
$$R_{\mathsf{B}}^{\eta}R_{\mathsf{A}}^{\eta} = 2J_{\mathsf{B}}^{\eta}(2J_{\mathsf{A}}^{\eta} - \mathrm{Id}) - 2J_{\mathsf{A}}^{\eta} + \mathrm{Id},$$

it suffices to evaluate reflectors R_{A}^{η} and R_{B}^{η} (or equivalently resolvants J_{A}^{η} and J_{B}^{η}) with summable errors.

As already argued in Alert 18.14, locally we would prefer over-relaxation (corresponding to $\gamma_t \in [\frac{1}{2}, 1]$) than under-relaxation (corresponding to $\gamma_t \in (0, \frac{1}{2})$).

Varga, R. S. (2009). "Matrix Iterative Analysis". 2nd. Springer.

Eckstein, J. and D. P. Bertsekas (1992). "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators". *Mathematical Programming*, vol. 55, pp. 293–318.

Peaceman, D. W. and H. H. Rachford Jr. (1955). "The Numerical Solution of Parabolic and Elliptic Differential Equations". *Journal of the Society for Industrial and Applied Mathematics*, vol. 3, no. 1, pp. 28–41.

Kellogg, R. B. (1969). "A Nonlinear Alternating Direction Method". *Mathematics of Computation*, vol. 23, no. 105, pp. 23–27.

Douglas Jr., J. and H. H. Rachford Jr. (1956). "On the Numerical Solution of Heat Conduction Problems in Two and Three Space Variables". *Transactions of the American Mathematical Society*, vol. 82, no. 2, pp. 421–439.

Lions, P.-L. and B. Mercier (1979). "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979.

Proposition 20.10: Maximality induces closedness (Eckstein and Svaiter 2008)

Let $\mathsf{A}, \mathsf{B} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ *be maximal monotone. Then,*

- *the map* $\mathsf{S} : \mathbb{R}^d \times \mathbb{R}^d \rightrightarrows \mathbb{R}^d \times \mathbb{R}^d$, $(\mathbf{x}, \mathbf{x}^*) \mapsto (\mathsf{B}\mathbf{x} + \mathbf{x}^*, \mathsf{A}^{-1}\mathbf{x}^* - \mathbf{x})$ *is maximal monotone.*
- *if* $(\mathbf{b}, \mathbf{b}^*) \leftharpoonup (\mathbf{b}_t, \mathbf{b}_t^*) \in \operatorname{gph}\mathsf{B}$ *and there exist* $(\mathbf{a}_t, \mathbf{a}_t^*) \in \operatorname{gph}\mathsf{A}$ *such that* $\mathbf{b}_t - \mathbf{a}_t \to \mathbf{0}$ *and* $\mathbf{b}_t^* + \mathbf{a}_t^* \to \mathbf{0}$, *then* $\mathbf{b} \in (\mathsf{A}+\mathsf{B})^{-1}\mathbf{0}$, $\mathbf{b}^* \in [-\mathsf{A}^{-1} \circ (-\mathrm{Id}) + B^{-1}]^{-1}\mathbf{0}$, $(\mathbf{a}_t, \mathbf{a}_t^*) \rightharpoonup (\mathbf{a}, \mathbf{a}^*) = (\mathbf{b}, -\mathbf{b}^*) \in \operatorname{gph}\mathsf{A}$ *and* $(\mathbf{b}, \mathbf{b}^*) \in \operatorname{gph}\mathsf{B}$.

Proof: Clearly, $\mathsf{S} = \mathsf{B} \times \mathsf{A}^{-1} + \mathsf{T}$ where $\mathsf{T}(\mathbf{x}, \mathbf{x}^*) = (\mathbf{x}^*, -\mathbf{x})$ and its maximality follows from Corollary 17.19.

Since $(\mathbf{b}_t^* + \mathbf{a}_t^*, \mathbf{a}_t - \mathbf{b}_t) \in \mathsf{S}(\mathbf{b}_t, \mathbf{a}_t^*)$, $(\mathbf{b}_t, \mathbf{a}_t^*) \rightharpoonup (\mathbf{b}, -\mathbf{b}^*)$ and $(\mathbf{b}_t^* + \mathbf{a}_t^*, \mathbf{a}_t - \mathbf{b}_t) \to (\mathbf{0}, \mathbf{0})$, it follows from the closedness of $\operatorname{gph}\mathsf{S}$ that $(\mathbf{0}, \mathbf{0}) \in \mathsf{S}(\mathbf{b}, -\mathbf{b}^*) = (\mathsf{B}\mathbf{b} - \mathbf{b}^*, \mathsf{A}^{-1}(-\mathbf{b}^*) - \mathbf{b})$. ■

The slick proof above also builds on an observation due to Bauschke (2009), which frees us from assuming $\mathsf{A}+\mathsf{B}$ is maximal monotone. See also Alves (2020) and Eckstein and Svaiter (2009).

Eckstein, J. and B. F. Svaiter (2008). "A family of projective splitting methods for the sum of two maximal monotone operators". *Mathematical Programming*, vol. 111, pp. 173–199.
Bauschke, H. H. (2009). "A Note on the Paper by Eckstein and Svaiter on "General Projective Splitting Methods for Sums of Maximal Monotone Operators"". *SIAM Journal on Control and Optimization*, vol. 48, no. 4, pp. 2513–2515.
Alves, M. M. (2020). "Another proof and a generalization of a theorem of H. H. Bauschke on monotone operators". *Optimization*.
Eckstein, J. and B. F. Svaiter (2009). "General Projective Splitting Methods for Sums of Maximal Monotone Operators". *SIAM Journal on Control and Optimization*, vol. 48, no. 2, pp. 787–811.

Exercise 20.11: A primal-dual perspective

Let $\mathsf{A},\mathsf{B}:\mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximal monotone, $\mathsf{T} := \mathsf{A}+\mathsf{B}$, and $\mathsf{T}^* := -\mathsf{B}^{-1}\circ(-\mathrm{Id}) + \mathsf{A}^{-1}$. Prove the following:

- $\mathsf{T}^{-1}\mathbf{0} = J^\eta_{\mathsf{A}}(\mathrm{Fix} R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}})$ (for any $\eta > 0$).
- $\mathsf{T}^{*-1}\mathbf{0} = {}^\eta\mathsf{A}(\mathrm{Fix} R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}})$, where recall that ${}^\eta\mathsf{A} := (\mathrm{Id} - J^\eta_{\mathsf{A}})/\eta$.
- $\mathsf{T}^{-1}\mathbf{0} \neq \emptyset \iff \mathsf{T}^{*-1}\mathbf{0} \neq \emptyset \iff \mathrm{Fix} R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \neq \emptyset$.
- $J_{\mathsf{A}^{-1}} = \mathrm{Id} - J_{\mathsf{A}}$ hence $R_{\mathsf{A}^{-1}} = -R_{\mathsf{A}}$.
- $-\mathsf{B}^{-1}-$ is maximal monotone and $J_{-\mathsf{B}^{-1}-} = \mathrm{Id} + J_{\mathsf{B}}\circ(-\mathrm{Id})$ hence $R_{-\mathsf{B}^{-1}-} = R_{\mathsf{B}}\circ(-\mathrm{Id})$.
- $R_{\mathsf{B}} R_{\mathsf{A}} = R_{-\mathsf{B}^{-1}-} R_{\mathsf{A}^{-1}}$. In other words, with $\eta \equiv 1$, the splitting Line 4 applied to the primal (19.1) and the dual (19.2) yields the same iterates (Gabay 1983, p. 323)!
- More generally, $R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} = R_{\eta\mathsf{B}} R_{\eta\mathsf{A}} = R_{-\mathsf{B}^{-1}\eta^{-1}-} R_{\mathsf{A}^{-1}\eta^{-1}} = \eta R^{\eta^{-1}}_{-\mathsf{B}^{-1}-} R^{\eta^{-1}}_{\mathsf{A}^{-1}} \eta^{-1}$, since $J_{\mathsf{A}^{-1}\eta^{-1}} = \eta J^{\eta^{-1}}_{\mathsf{A}^{-1}}\eta^{-1}$ hence $R_{\mathsf{A}^{-1}\eta^{-1}} = \eta R^{\eta^{-1}}_{\mathsf{A}^{-1}}\eta^{-1}$. In other words, applying step size η to the primal (19.1) is in some sense equivalent to applying step size η^{-1} to the dual (19.2).

Gabay, D. (1983). "Applications of the Method of Multipliers to Variational Inequalities". In: *Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems*. Vol. 15. 9, pp. 299–331.

Exercise 20.12: Primal-dual with linear composition

Let $\mathsf{A}:\mathbb{R}^d \rightrightarrows \mathbb{R}^d, \mathsf{B}:\mathbb{R}^p \rightrightarrows \mathbb{R}^p$, $L:\mathbb{R}^d \to \mathbb{R}^p$ a (continuous) linear map, $\mathsf{T} = \mathsf{A} + L^\top \mathsf{B} L$, and $\mathsf{T}^* = -L\mathsf{A}^{-1}(-L^\top) + \mathsf{B}^{-1}$. Prove the following:

$$\mathbf{0} \in \mathsf{T}\mathbf{a} \iff \mathbf{0} \in \mathsf{T}^*\mathbf{b}^*, \quad \text{where} \quad \text{we may choose } \mathbf{b}^* \in \mathsf{B}L\mathbf{a} \text{ and } -L^\top\mathbf{b}^* \in \mathsf{A}\mathbf{a}.$$

If both A and B are maximal monotone, then

- the map $\mathsf{S}:\mathbb{R}^d\times\mathbb{R}^p \rightrightarrows \mathbb{R}^d\times\mathbb{R}^p$, $(\mathbf{a},\mathbf{b}^*) \mapsto (\mathsf{A}\mathbf{a} + L^\top\mathbf{b}^*, \mathsf{B}^{-1}\mathbf{b}^* - L\mathbf{a})$ is maximal monotone.
- if $(\mathbf{a}_t,\mathbf{a}^*_t) \in \mathrm{gph}\,\mathsf{A}$ and $(\mathbf{b}_t,\mathbf{b}^*_t) \in \mathrm{gph}\,\mathsf{B}$ such that $(\mathbf{a}_t,\mathbf{b}^*_t) \rightharpoonup (\mathbf{a},\mathbf{b}^*)$, $\mathbf{a}^*_t + L^\top\mathbf{b}^*_t \to \mathbf{0}$ and $L\mathbf{a}_t - \mathbf{b}_t \to \mathbf{0}$, then $\mathbf{a}\in\mathsf{T}^{-1}\mathbf{0}$, $\mathbf{b}^*\in\mathsf{T}^{*-1}\mathbf{0}$, $(\mathbf{a}, -L^\top\mathbf{b}^*) \in \mathrm{gph}\,\mathsf{A}$ and $(L\mathbf{a},\mathbf{b}^*) \in \mathrm{gph}\,\mathsf{B}$.

(Hint: Prove directly or reduce to Exercise 20.11 through the product space trick.)

Theorem 20.13: Convergence of DR (Svaiter 2011)

Let $\mathsf{A}, \mathsf{B} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ *be maximal monotone and* $\mathsf{T} = \mathsf{A} + \mathsf{B}$. *Consider Line 4 with* $\sum_t \gamma_t(1 - \gamma_t) = \infty$, $\eta_t \equiv \eta > 0$, *and* $\sum_t \|\boldsymbol{\epsilon}_t\|_2 < \infty$.

- *If* $\mathsf{T}^{-1}\mathbf{0} = (\mathsf{A} + \mathsf{B})^{-1}\mathbf{0} \neq \emptyset$, *then* $\mathbf{w}_t \rightharpoonup \mathbf{w}_\infty = R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_\infty$, $\mathbf{z}_t := J^\eta_{\mathsf{A}} \mathbf{w}_t \rightharpoonup J^\eta_{\mathsf{A}} \mathbf{w}_\infty \in \mathsf{T}^{-1}\mathbf{0}$ *and* $\mathbf{z}^*_t := {}^\eta\mathsf{A}\mathbf{w}_t \rightharpoonup {}^\eta\mathsf{A}\mathbf{w}_\infty \in \mathsf{T}^{*-1}\mathbf{0}$, *where recall that* $\mathsf{T}^* := -\mathsf{B}^{-1} \circ (-\mathrm{Id}) + \mathsf{A}^{-1}$ *and* ${}^\eta\mathsf{A} := (\mathrm{Id} - J^\eta_{\mathsf{A}})/\eta$.
- *If* $\mathsf{T}^{-1}\mathbf{0} = (\mathsf{A} + \mathsf{B})^{-1}\mathbf{0} = \emptyset$, *then* $\{\mathbf{w}_t\}$ *is unbounded.*

Proof: The result on $\mathbf{w}_t$ readily follows from Theorem 16.13.

To prove the result on $\mathbf{z}_t$, let us define

$$\mathbf{u}_t := J^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t, \quad \mathbf{u}^*_t = {}^\eta\mathsf{B} R^\eta_{\mathsf{A}} \mathbf{w}_t = \tfrac{R^\eta_{\mathsf{A}}\mathbf{w}_t - J^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t}{\eta}.$$

Clearly, $(\mathbf{u}_t, \mathbf{u}^*_t) \in \operatorname{gph} \mathsf{B}$, $(\mathbf{z}_t, \mathbf{z}^*_t) \in \operatorname{gph} \mathsf{A}$, and

$$2(\mathbf{u}_t - \mathbf{z}_t) = 2J^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t - 2J^\eta_{\mathsf{A}} \mathbf{w}_t = R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t - \mathbf{w}_t \to \mathbf{0},$$

as shown in the proof of Theorem 16.13. Moreover,

$$\mathbf{u}^*_t + \mathbf{z}^*_t = \frac{R^\eta_{\mathsf{A}}\mathbf{w}_t - J^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t + \mathbf{w}_t - J^\eta_{\mathsf{A}}\mathbf{w}_t}{\eta} = \frac{J^\eta_{\mathsf{A}}\mathbf{w}_t - J^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t}{\eta} = \frac{\mathbf{z}_t - \mathbf{u}_t}{\eta} \to \mathbf{0}.$$

Clearly, $\{\mathbf{w}_t\}$ and hence $\{\mathbf{z}_t\}$ are bounded. Consider any subsequence $\mathbf{z}_{t_k} \rightharpoonup \mathbf{z}$ and $\mathbf{w}_{t_k} \rightharpoonup \mathbf{w}_\infty$. We have $\mathbf{z}^*_{t_k} = \frac{\mathbf{w}_{t_k} - \mathbf{z}_{t_k}}{\eta} \rightharpoonup \frac{\mathbf{w}_\infty - \mathbf{z}}{\eta} \in \mathsf{T}^{*-1}\mathbf{0}$, $\mathbf{z} \in \mathsf{T}^{-1}\mathbf{0}$, and $\frac{\mathbf{w}_\infty - \mathbf{z}}{\eta} \in \mathsf{A}\mathbf{z}$, thanks to Proposition 20.10. Rearranging we obtain $\mathbf{z} = J^\eta_{\mathsf{A}}\mathbf{w}_\infty$. Since $\mathbf{z}$ is arbitrary, we must have $\mathbf{z}_t \rightharpoonup J^\eta_{\mathsf{A}}\mathbf{w}_\infty \in \mathsf{T}^{-1}\mathbf{0}$ and hence $\mathbf{z}^*_t \rightharpoonup {}^\eta\mathsf{A}\mathbf{w}_\infty \in \mathsf{T}^{*-1}\mathbf{0}$. ■

We note that $\operatorname{gph}\mathsf{B} \ni (\mathbf{u}_t, \mathbf{u}^*_t) \rightharpoonup (J^\eta_{\mathsf{A}}\mathbf{w}_\infty, -{}^\eta\mathsf{A}\mathbf{w}_\infty)$.

Svaiter, B. F. (2011). "On Weak Convergence of the Douglas–Rachford Method". *SIAM Journal on Control and Optimization*, vol. 49, no. 1, pp. 280–287.

Theorem 20.14: Convergence of PR (Lions and Mercier 1979)

Let $\mathsf{A}, \mathsf{B} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ *be maximal monotone and* $\mathsf{T} = \mathsf{A} + \mathsf{B}$. *Consider Line 4 with* $\gamma_t \equiv 1$, $\eta_t \equiv \eta > 0$, *and* $\sum_t \|\boldsymbol{\epsilon}_t\|_2 < \infty$. *Assume* A *is strictly monotone.*

- *There exists at most one zero* $\mathbf{z}$ *of* T.
- *If* $\mathbf{z} = \mathsf{T}^{-1}\mathbf{0} = (\mathsf{A} + \mathsf{B})^{-1}\mathbf{0}$ *exists, then* $\mathrm{Fix} R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \neq \emptyset$, $\{\mathbf{w}_t\}$ *is quasi-Fejér monotone w.r.t.* $\mathrm{Fix} R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}}$ *and* $\mathbf{z}_t := J^\eta_{\mathsf{A}} \mathbf{w}_t \rightharpoonup \mathbf{z}$.
- *If* $\mathsf{T}^{-1}\mathbf{0} = (\mathsf{A} + \mathsf{B})^{-1}\mathbf{0} = \emptyset$, *then* $\{\mathbf{w}_t\}$ *is unbounded.*

Proof: The first claim follows from the strict monotonicity of A:

$$0 \leq \langle \mathbf{u} - \mathbf{v}; \mathbf{u}^* - \mathbf{v}^* \rangle = -\langle \mathbf{u} - \mathbf{v}; \mathbf{u}_* - \mathbf{v}_* \rangle \leq 0, \quad \text{where} \quad \mathbf{u}^* \in \mathsf{A}\mathbf{u}, \mathbf{v}^* \in \mathsf{A}\mathbf{v}, \mathbf{u}_* \in \mathsf{B}\mathbf{u}, \mathbf{v}_* \in \mathsf{B}\mathbf{v}.$$

Let $\mathbf{w} \in \mathrm{Fix} R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}}$. Using the definition of Line 4, we have

$$\begin{aligned} \|\mathbf{w}_{t+1} - \mathbf{w}\|_2 &\leq \epsilon_t + \|R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}_t - R^\eta_{\mathsf{B}} R^\eta_{\mathsf{A}} \mathbf{w}\|_2 \\ &\leq \epsilon_t + \|R^\eta_{\mathsf{A}} \mathbf{w}_t - R^\eta_{\mathsf{A}} \mathbf{w}\|_2 \\ (\ \mathbf{z}_t := J^\eta_{\mathsf{A}}\mathbf{w}_t,\ \mathbf{z} := J^\eta_{\mathsf{A}}\mathbf{w}\ ) &= \epsilon_t + \|2(\mathbf{z}_t - \mathbf{z}) - (\mathbf{w}_t - \mathbf{w})\|_2 \end{aligned}$$

$$= \epsilon_t + \sqrt{-4\eta \left\langle \mathbf{z}_t - \mathbf{z}; \tfrac{\mathbf{w}_t - \mathbf{z}_t}{\eta} - \tfrac{\mathbf{w}-\mathbf{z}}{\eta} \right\rangle + \|\mathbf{w}_t - \mathbf{w}\|_2^2}$$

$$(\ \mathbf{z}_t^* := \tfrac{\mathbf{w}_t - \mathbf{z}_t}{\eta} \in \mathsf{A}\mathbf{z}_t,\ \mathbf{z}^* := \tfrac{\mathbf{w}-\mathbf{z}}{\eta} \in \mathsf{A}\mathbf{z}\ ) = \epsilon_t + \sqrt{-4\eta \left\langle \mathbf{z}_t - \mathbf{z}; \mathbf{z}_t^* - \mathbf{z}^* \right\rangle + \|\mathbf{w}_t - \mathbf{w}\|_2^2},$$

whence $\{\mathbf{w}_t\}$ is quasi-Fejér monotone w.r.t. $\mathrm{Fix} R_{\mathsf{B}}^\eta R_{\mathsf{A}}^\eta$ (see Exercise 16.3). Moreover,

$$\langle \mathbf{z}_t - \mathbf{z}; \mathbf{z}_t^* - \mathbf{z}^* \rangle \to 0.$$

Since $\mathbf{z}_t$ and $\mathbf{z}_t^*$ are bounded and $\mathrm{gph}\,\mathsf{A}$ is closed, we may assume $(\mathbf{z}_t, \mathbf{z}_t^*) \rightharpoonup (\bar{\mathbf{z}}, \bar{\mathbf{z}}^*) \in \mathrm{gph}\,\mathsf{A}$. From monotonicity of A we have

$$\liminf_t \langle \mathbf{z}_t; \mathbf{z}_t^* \rangle \geq \lim_t \langle \bar{\mathbf{z}}; \mathbf{z}_t^* - \bar{\mathbf{z}}^* \rangle + \langle \mathbf{z}_t; \bar{\mathbf{z}}^* \rangle = \langle \bar{\mathbf{z}}; \bar{\mathbf{z}}^* \rangle, \quad \text{hence}$$
$$0 \leq \langle \bar{\mathbf{z}} - \mathbf{z}; \bar{\mathbf{z}}^* - \mathbf{z}^* \rangle \leq \liminf_t \langle \mathbf{z}_t - \mathbf{z}; \mathbf{z}_t^* - \mathbf{z}^* \rangle = 0.$$

Using strict monotonicity we have $\bar{\mathbf{z}} = \mathbf{z}$, hence follows $\mathbf{z}_t \rightharpoonup \mathbf{z}$, the unique zero of T.

Lastly, if $\{\mathbf{w}_t\}$ is bounded, we may restrict the nonexpansion $R_{\mathsf{B}}^\eta R_{\mathsf{A}}^\eta$ to a compact convex set and the existence of a fixed point would follow from Brouwer's celebrated fixed point theorem. ■

The proof above follows the classic idea of using uniqueness to force convergence. Needless to say a similar result holds if instead B satisfies the assumption.

Lions, P.-L. and B. Mercier (1979). "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979.

Alert 20.15: Non-convergence of PR

Consider the rotation in $\mathbb{R}^2$:

$$\mathsf{A} = \mathsf{B} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \text{where} \quad (\mathsf{A}+\mathsf{B})\mathbf{z} = \mathbf{0} \iff \mathbf{z} = \mathbf{0}.$$

Since $\langle \mathbf{w}, \mathsf{A}\mathbf{w} \rangle = 0$ for all $\mathbf{w}$, we know A is maximal monotone. We have

$$J_{\mathsf{A}}^\eta = \frac{1}{(1+\eta)^2} \begin{bmatrix} 1 & \eta \\ -\eta & 1 \end{bmatrix}, \quad R_{\mathsf{A}}^\eta = \frac{1}{(1+\eta)^2} \begin{bmatrix} 1-\eta^2 & 2\eta \\ -2\eta & 1-\eta^2 \end{bmatrix}.$$

Since both J_{A}^η and R_{A}^η are rotations (i.e. det = 1), with $\gamma_t \equiv 1$ and $\boldsymbol{\epsilon}_t \equiv \mathbf{0}$, $\|\mathbf{z}_t\|_2 \equiv \|\mathbf{w}_0\|_2$ hence may not converge to any point (hence also $\mathbf{0}$, the unique zero of $\mathsf{A}+\mathsf{B}$). Moreover, $\mathbf{w}_t$ may not converge (to any point) either.

We verify that A is not strictly monotone and convince yourself that $\mathbf{z}_t$ and $\mathbf{w}_t$ do converge if we choose say $\gamma_t \equiv \gamma \in (0,1)$.

Theorem 20.16: Strong convergence of DR and PR

Let $\mathsf{B} : \mathbb{R}^d \rightrightarrows \mathbb{R}^d$ be maximal monotone and $\mathsf{A} : \mathbb{R}^d \to \mathbb{R}^d$ be σ-strongly maximal monotone and L-Lipschitz continuous. Consider Line 4 with $\eta_t \equiv \eta > 0$, and $\sum_t \|\boldsymbol{\epsilon}_t\|_2 < \infty$. Then,

$$\|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2 \leq \|\mathbf{w}_{t+1} - \mathbf{w}_\infty\|_2 \leq \left(1 - \gamma_t + \gamma_t \sqrt{1 - \tfrac{4\eta\sigma}{(1+\eta\mathsf{L})^2}}\right) \cdot \|\mathbf{w}_t - \mathbf{w}_\infty\|_2 + \epsilon_t, \quad \textit{where} \quad \epsilon_t := \|\boldsymbol{\epsilon}_t\|_2,$$

and recall that $\mathbf{z}_t := J_{\mathsf{A}}^\eta \mathbf{w}_t$.

Proof: Immediate from Theorem 20.6. ■

To optimize the bound, we set $\eta = \frac{1}{\mathsf{L}}$ hence

$$\|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2 \le \|\mathbf{w}_{t+1} - \mathbf{w}_\infty\|_2 \le \left(1 - \gamma_t + \gamma_t\sqrt{1-\kappa}\right) \cdot \|\mathbf{w}_t - \mathbf{w}_\infty\|_2 + \epsilon_t, \quad \text{where} \quad \kappa = \sigma/\mathsf{L}.$$

Obviously, we should also set $\gamma_t \equiv 1$ so that PR converges fastest in this setting. In fact, we have

$$\begin{aligned}\|\mathbf{w}_{t+1} - \mathbf{w}_\infty\|_2^2 &= \|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2^2 + \eta^2\|\mathbf{z}^*_{t+1} - \mathbf{z}^*_\infty\|_2^2 + 2\eta\left\langle \mathbf{z}_{t+1} - \mathbf{z}_\infty, \mathbf{z}^*_{t+1} - \mathbf{z}^*_\infty\right\rangle \\ &\ge \|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2^2 + \eta^2\|\mathbf{z}^*_{t+1} - \mathbf{z}^*_\infty\|_2^2 + 2\eta\sigma\|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2^2 \\ &= (1 + 2\eta\sigma - \eta^2)\|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2^2 + \eta^2[\|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2^2 + \|\mathbf{z}^*_{t+1} - \mathbf{z}^*_\infty\|_2^2].\end{aligned}$$

Thus, for $\eta > 1$ we favor dual convergence while for $\eta < 1$ we favor primal convergence. When $1+2\eta\sigma-\eta^2 \ge 0$, we may apply the Lipschitz bound (20.2) to obtain

$$\|\mathbf{z}_{t+1} - \mathbf{z}_\infty\|_2^2 + \|\mathbf{z}^*_{t+1} - \mathbf{z}^*_\infty\|_2^2 \le \tfrac{1}{\eta^2}\left(1 - \tfrac{1+2\eta\sigma-\eta^2}{(1+\eta\mathsf{L})^2}\right)\|\mathbf{w}_{t+1} - \mathbf{w}_\infty\|_2^2,$$

so that both primal and dual converges at a linear rate for $\eta \in [1, \sigma + \sqrt{1+\sigma^2}]$.

Example 20.17: Unpacking Alternating direction method of multipliers (ADMM)

Consider the generic minimization problem

$$\inf_{\mathbf{a}}\ g(L\mathbf{a}) + h(\mathbf{a}), \text{ or equivalently } \inf_{\mathbf{a},\mathbf{b}}\ g(\mathbf{b}) + h(\mathbf{a}), \quad \text{s.t.} \quad L\mathbf{a} = \mathbf{b}, \tag{20.4}$$

and its Fenchel-Rockafellar dual

$$-\inf_{\boldsymbol{\mu}}\ h^*(-L^\top\boldsymbol{\mu}) + g^*(\boldsymbol{\mu}), \tag{20.5}$$

where $g : \mathbb{R}^p \to \mathbb{R} \cup \{\infty\}$ and $h : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ are closed proper convex and $L : \mathbb{R}^d \to \mathbb{R}^p$ is linear.

Introducing the Lagrangian multiplier $\boldsymbol{\mu}$ in (20.4) we obtain the Lagrangian:

$$\inf_{\mathbf{a},\mathbf{b}}\ \sup_{\boldsymbol{\mu}}\ g(\mathbf{b}) + h(\mathbf{a}) + \langle L\mathbf{a} - \mathbf{b}; \boldsymbol{\mu}\rangle \quad = \quad \sup_{\boldsymbol{\mu}}\ \underbrace{\inf_{\mathbf{a},\mathbf{b}}\ g(\mathbf{b}) + h(\mathbf{a}) + \langle L\mathbf{a} - \mathbf{b}; \boldsymbol{\mu}\rangle}_{\underline{L}(\boldsymbol{\mu})}.$$

One may then apply Uzawa's Algorithm 12.21 to maximize the dual function $\underline{L}(\boldsymbol{\mu})$, which however is often nonsmooth hence requires diminishing step sizes. Instead, we may consider the augmented Lagrangian, where the penalty parameter η need not increase to ∞:

$$\sup_{\boldsymbol{\mu}}\ \left[\inf_{\mathbf{a},\mathbf{b}}\ g(\mathbf{b}) + h(\mathbf{a}) \ + \langle L\mathbf{a} - \mathbf{b}; \boldsymbol{\mu}\rangle \ + \tfrac{\eta}{2}\|L\mathbf{a} - \mathbf{b}\|_2^2\right] \ = \ \sup_{\boldsymbol{\mu}}\left[\sup_{\boldsymbol{\nu}} \underline{L}(\boldsymbol{\nu}) - \tfrac{1}{2\eta}\|\boldsymbol{\nu} - \boldsymbol{\mu}\|_2^2\right], \tag{20.6}$$

whose inner function is now smooth hence we may apply Uzawa's Algorithm 12.21 with constant step size η:

$$\boldsymbol{\mu}_{t+1} \leftarrow \boldsymbol{\mu}_t + \eta(L\mathbf{a}_{t+1} - \mathbf{b}_{t+1}), \tag{20.7}$$

where $(\mathbf{a}_{t+1}, \mathbf{b}_{t+1})$ minimizes (20.6) (using any sensible algorithm) while fixing $\boldsymbol{\mu}_t$. As discussed in Example 18.42, this is exactly the proximal point algorithm applied to the dual $\underline{L}(\boldsymbol{\mu})$.

However, solving $\mathbf{a}$ and $\mathbf{b}$ simultaneously in the augmented Lagrangian (20.6) turns out to be challenging, after all they are coupled due to the quadratic penalty. Fortunately, we may apply just one step of alternating minimization to $\mathbf{a}$ and $\mathbf{b}$ sequentially (Fortin 1975; Gabay and Mercier 1976; Glowinski and Marroco 1975):

$$\begin{aligned}\mathbf{a}_{t+1} &\in \operatorname*{argmin}_{\mathbf{a}}\ h(\mathbf{a}) + \langle L\mathbf{a} - \mathbf{b}_t; \boldsymbol{\mu}_t\rangle + \tfrac{\eta}{2}\|L\mathbf{a} - \mathbf{b}_t\|_2^2 \quad \equiv \quad h(\mathbf{a}) + \tfrac{\eta}{2}\|L\mathbf{a} - \mathbf{b}_t + \boldsymbol{\mu}_t/\eta\|_2^2 \\ \mathbf{b}_{t+1} &= \operatorname*{argmin}_{\mathbf{b}}\ g(\mathbf{b}) + \langle L\mathbf{a}_{t+1} - \mathbf{b}; \boldsymbol{\mu}_t\rangle + \tfrac{\eta}{2}\|L\mathbf{a}_{t+1} - \mathbf{b}\|_2^2 \quad \equiv \quad g(\mathbf{b}) + \tfrac{\eta}{2}\|L\mathbf{a}_{t+1} - \mathbf{b} + \boldsymbol{\mu}_t/\eta\|_2^2.\end{aligned}$$

To understand the above updates, let us apply the Fenchel-Rockafellar duality again:

$$\mathbf{a}^*_{t+1} - \eta L\mathbf{a}_{t+1} = -\eta\mathbf{b}_t + \boldsymbol{\mu}_t, \quad \text{where} \quad \mathbf{a}^*_{t+1} = \operatorname*{argmin}_{\mathbf{a}^*} \tfrac{1}{2\eta}\|\mathbf{a}^* + \eta\mathbf{b}_t - \boldsymbol{\mu}_t\|_2^2 + h^*(-L^\top\mathbf{a}^*) \tag{20.8}$$

$$\mathbf{b}^*_{t+1} + \eta\mathbf{b}_{t+1} = \eta L\mathbf{a}_{t+1} + \boldsymbol{\mu}_t, \quad \text{where} \quad \mathbf{b}^*_{t+1} = \operatorname*{argmin}_{\mathbf{b}^*} \tfrac{1}{2\eta}\|\mathbf{b}^* - \eta L\mathbf{a}_{t+1} - \boldsymbol{\mu}_t\|_2^2 + g^*(\mathbf{b}^*). \tag{20.9}$$

(These relations also follow from Moreau's identity, where we remind that the conjugate function of $\mathbf{a}^* \mapsto h^*(-L^\top\mathbf{a}^*)$ is $\mathbf{b} \mapsto \inf\{h(-\mathbf{a}) : \mathbf{b} = L\mathbf{a}\}$.) It follows from (20.7) and (20.9) that $\boldsymbol{\mu}_t = \mathbf{b}^*_t$.

From the optimality conditions of $\mathbf{b}^*_{t+1}$ and $\mathbf{a}_{t+1}$ we verify that (see Exercise 17.8):

$$(\mathbf{b}^*_{t+1}, \mathbf{b}_{t+1}) \in \operatorname{gph}\partial g^* \implies \mathbf{b}^*_{t+1} = J^\eta_{\partial g^*}\mathbf{w}_{t+1}, \quad \text{where} \quad \mathbf{w}_{t+1} := \eta\mathbf{b}_{t+1} + \mathbf{b}^*_{t+1},$$

$$(\mathbf{a}^*_{t+1}, \mathbf{a}_{t+1}) \in \operatorname{gph}[\partial h^* \circ (-L^\top)] \implies \mathbf{a}^*_{t+1} = J^\eta_{-L\circ\partial h^*\circ(-L^\top)}(-\eta L\mathbf{a}_{t+1} + \mathbf{a}^*_{t+1})$$

$$\overset{(20.8)}{=} J^\eta_{-L\circ\partial h^*\circ(-L^\top)}(-\eta\mathbf{b}_t + \mathbf{b}^*_t), \quad \text{since } \boldsymbol{\mu}_t = \mathbf{b}^*_t.$$

Therefore, we deduce that

$$\mathbf{w}_{t+1} := \eta\mathbf{b}_{t+1} + \mathbf{b}^*_{t+1} \overset{(20.9)}{=} \eta L\mathbf{a}_{t+1} + \boldsymbol{\mu}_t \overset{(20.8)}{=} \mathbf{a}^*_{t+1} + \eta\mathbf{b}_t = J^\eta_{-L\circ\partial h^*\circ(-L^\top)}(-\eta\mathbf{b}_t + \mathbf{b}^*_t) + \eta\mathbf{b}_t$$

$$= J^\eta_{-L\circ\partial h^*\circ(-L^\top)}(2\mathbf{b}^*_t - \mathbf{w}_t) + \mathbf{w}_t - \mathbf{b}^*_t = \frac{\mathrm{Id} + R^\eta_{-L\circ\partial h^*\circ(-L^\top)}R^\eta_{\partial g^*}}{2}\mathbf{w}_t,$$

which, as recognized by Gabay (1983), is exactly the Douglas-Rachford algorithm applied to the dual problem (20.5), with the maximal monotone maps ∂h^* and $-L\circ\partial g^*\circ(-L^\top)$ (under mild conditions so that the chain rule holds for $g^*\circ -L^\top$)! It follows immediately from Theorem 20.13 that the dual variable $\boldsymbol{\mu}_t = \mathbf{b}^*_t = J^\eta_{\partial g^*}\mathbf{w}_t$ in (20.7) converges to a solution of the dual problem (20.5) (whose existence we assume).

Fortin, M. (1975). "Minimization of some non-differentiable functionals by the Augmented Lagrangian Method of Hestenes and Powell". *Applied Mathematics and Optimization*, vol. 2, pp. 236–250.

Gabay, D. and B. Mercier (1976). "A dual algorithm for the solution of nonlinear variational problems via finite element approximation". *Computers & Mathematics with Applications*, vol. 2, no. 1, pp. 17–40.

Glowinski, R. and A. Marroco (1975). "Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une class de problèmes de Dirichlet non linéaires". *ESAIM: Mathematical Modelling and Numerical Analysis*, vol. 9, no. R2, pp. 41–76.

Gabay, D. (1983). "Applications of the Method of Multipliers to Variational Inequalities". In: *Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems.* Vol. 15. 9, pp. 299–331.

Remark 20.18: More refinements

We further note in Example 20.17 that

$$(-L^\top\mathbf{a}^*_t, \mathbf{a}_t) \in \operatorname{gph}\partial h^*, \quad (\mathbf{b}^*_t, \mathbf{b}) \in \operatorname{gph}\partial g^*, \quad \mathbf{b}_t - L\mathbf{a}_t = \frac{\mathbf{b}^*_t - \mathbf{b}^*_{t-1}}{\eta} \to \mathbf{0}, \quad \mathbf{b}^*_t \rightharpoonup \mathbf{b}^*_\infty,$$

$\mathbf{a}^*_t - \mathbf{b}^*_t \rightharpoonup \mathbf{0}$ and $L\mathbf{a}_t \rightharpoonup L\mathbf{a}_\infty$. Thus, in our finite dimensional setting, if L has full column rank, $\mathbf{a}_t \rightharpoonup \mathbf{a}_\infty$ and applying Exercise 20.12 we know $\mathbf{a}_\infty$ is a solution of the primal problem (20.4) (whose existence we assume). Obviously, we may also apply PR (Gabay 1983) or the general splitting Line 4 (Eckstein and Bertsekas 1992) to the dual problem (20.5) and obtain similar convergence results.

In a nutshell, the ADMM algorithm is Douglas-Rachford applied to the dual, where we evaluate the resolvants through Moreau's identity:

$$\mathrm{P}^\eta_{h^*\circ -L^\top} \to \mathrm{Id} - \mathrm{P}^{1/\eta}_{-Lh}\eta, \qquad \mathrm{P}^\eta_{g^*} \to \mathrm{Id} - \mathrm{P}^{1/\eta}_{g}\eta.$$

It is clear that computing the resolvants on the right-hand sides up to summable error leads to the resolvants on the left-hand sides with summable error (Eckstein and Bertsekas 1992). Thus, Theorem 20.13 and Theorem 20.14 still apply.

(Lions and Mercier 1979)

Gabay, D. (1983). "Applications of the Method of Multipliers to Variational Inequalities". In: *Augmented Lagrangian methods: Applications to the numerical solution of boundary-value problems.* Vol. 15. 9, pp. 299–331.

Eckstein, J. and D. P. Bertsekas (1992). "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators". *Mathematical Programming*, vol. 55, pp. 293–318.
Lions, P.-L. and B. Mercier (1979). "Splitting Algorithms for the Sum of Two Nonlinear Operators". *SIAM Journal on Numerical Analysis*, vol. 16, no. 6, pp. 964–979.

Exercise 20.19: Reverse-engineering ADMM

Show that the general splitting algorithm

$$\mathbf{w}_{t+1} = (1-\gamma_t)\mathbf{w}_t + \gamma_t R^{\eta}_{-L\circ\partial h^*\circ(-L^\top)} R^{\eta}_{\partial g^*}\mathbf{w}_t$$

can be unwrapped into the following "generalized ADMM" updates:

$$\begin{aligned}
\mathbf{a}_{t+1} &\in \operatorname*{argmin}_{\mathbf{a}} \ h(\mathbf{a}) + \langle L\mathbf{a} - \mathbf{b}_t; \boldsymbol{\mu}_t\rangle + \tfrac{\eta}{2}\|L\mathbf{a} - \mathbf{b}_t\|_2^2 \\
\mathbf{c}_{t+1} &:= 2\gamma_t L\mathbf{a}_{t+1} + (1-2\gamma_t)\mathbf{b}_t \\
\mathbf{b}_{t+1} &= \operatorname*{argmin}_{\mathbf{b}} \ g(\mathbf{b}) + \langle L\mathbf{a}_{t+1} - \mathbf{b}; \boldsymbol{\mu}_t\rangle + \tfrac{\eta}{2}\|\mathbf{c}_{t+1} - \mathbf{b}\|_2^2 \\
\boldsymbol{\mu}_{t+1} &= \boldsymbol{\mu}_t + \eta(\mathbf{c}_{t+1} - \mathbf{b}_{t+1}).
\end{aligned}$$

[Hint: Define as before $\mathbf{w}_{t+1} = \eta\mathbf{b}_{t+1} + \mathbf{b}^*_{t+1}$ so that $\boldsymbol{\mu}_{t+1} = \mathbf{b}^*_{t+1} = J^{\eta}_{\partial g^*}\mathbf{w}_t$.]