11. Douglas–Rachford method and ADMM

- Douglas–Rachford splitting method
- examples
- alternating direction method of multipliers
- image deblurring example
- convergence

Douglas–Rachford splitting algorithm

$$\text{minimize} \quad f(x) + g(x)$$

f and g are closed convex functions

Douglas–Rachford iteration: start at any y_0 and repeat for $k = 0, 1, \ldots,$

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_f(y_k) \\ y_{k+1} &= y_k + \operatorname{prox}_g(2x_{k+1} - y_k) - x_{k+1} \end{aligned}$$

- useful when f and g have inexpensive prox-operators
- x_k converges to a solution of $0 \in \partial f(x) + \partial g(x)$ (if a solution exists)
- not symmetric in f and g

Douglas–Rachford iteration as fixed-point iteration

- iteration on page 11.2 can be written as fixed-point iteration

$$y_{k+1} = F(y_k)$$

  where

$$F(y) = y + \mathrm{prox}_g(2\mathrm{prox}_f(y) - y) - \mathrm{prox}_f(y)$$

- y is a fixed point of F if and only if $x = \mathrm{prox}_f(y)$ satisfies $0 \in \partial f(x) + \partial g(x)$:

$$y = F(y)$$

$$\Updownarrow$$

$$0 \in \partial f(\mathrm{prox}_f(y)) + \partial g(\mathrm{prox}_f(y))$$

  (proof on next page)

Proof.

$$x = \mathrm{prox}_f(y), \quad y = F(y)$$

$$\Updownarrow$$

$$x = \mathrm{prox}_f(y), \quad x = \mathrm{prox}_g(2x - y)$$

$$\Updownarrow$$

$$y - x \in \partial f(x), \quad x - y \in \partial g(x)$$

- therefore, if $y = F(y)$, then $x = \mathrm{prox}_f(y)$ satisfies

$$0 = (y - x) + (x - y) \in \partial f(x) + \partial g(x)$$

- conversely, if $-z \in \partial f(x)$ and $z \in \partial g(x)$, then $y = x - z$ is a fixed point of F

Equivalent form of Douglas–Rachford algorithm

- start iteration on page 11.2 at y-update and renumber iterates

$$\begin{aligned} y_{k+1} &= y_k + \operatorname{prox}_g(2x_k - y_k) - x_k \\ x_{k+1} &= \operatorname{prox}_f(y_{k+1}) \end{aligned}$$

- switch y- and x-updates

$$\begin{aligned} u_{k+1} &= \operatorname{prox}_g(2x_k - y_k) \\ x_{k+1} &= \operatorname{prox}_f(y_k + u_{k+1} - x_k) \\ y_{k+1} &= y_k + u_{k+1} - x_k \end{aligned}$$

- make change of variables $w_k = x_k - y_k$

$$\begin{aligned} u_{k+1} &= \operatorname{prox}_g(x_k + w_k) \\ x_{k+1} &= \operatorname{prox}_f(u_{k+1} - w_k) \\ w_{k+1} &= w_k + x_{k+1} - u_{k+1} \end{aligned}$$

Scaling

algorithm applied to cost function scaled by $t > 0$

$$\text{minimize} \quad tf(x) + tg(x)$$

- algorithm of page 11.2

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{tf}(y_k) \\ y_{k+1} &= y_k + \operatorname{prox}_{tg}(2x_{k+1} - y_k) - x_{k+1} \end{aligned}$$

- algorithm of page 11.5

$$\begin{aligned} u_{k+1} &= \operatorname{prox}_{tg}(x_k + w_k) \\ x_{k+1} &= \operatorname{prox}_{tf}(u_{k+1} - w_k) \\ w_{k+1} &= w_k + x_{k+1} - u_{k+1} \end{aligned}$$

- the algorithm is not invariant with respect to scaling
- in theory, t can be any positive constant; several heuristics exist for adapting t

Douglas–Rachford iteration with relaxation

- fixed-point iteration with relaxation

$$y_{k+1} = y_k + \rho_k(F(y_k) - y_k)$$

$1 < \rho_k < 2$ is overrelaxation, $0 < \rho_k < 1$ underrelaxation

- algorithm of page 11.2 with relaxation

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_f(y_k) \\ y_{k+1} &= y_k + \rho_k \left(\operatorname{prox}_g(2x_{k+1} - y_k) - x_{k+1}\right) \end{aligned}$$

- algorithm of page 11.5

$$\begin{aligned} u_{k+1} &= \operatorname{prox}_g(x_k + w_k) \\ x_{k+1} &= \operatorname{prox}_f(x_k + \rho_k(u_{k+1} - x_k) - w_k) \\ w_{k+1} &= w_k + x_{k+1} - x_k + \rho_k(x_k - u_{k+1}) \end{aligned}$$

Primal-dual formulation

$$\begin{array}{lll} \text{primal:} & \text{minimize} & f(x) + g(x) \\ \text{dual:} & \text{maximize} & -g^*(z) - f^*(-z) \end{array}$$

- use Moreau decomposition to simplify step 2 of DR iteration (page 11.2):

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_f(y_k) \\ y_{k+1} &= x_{k+1} - \operatorname{prox}_{g^*}(2x_{k+1} - y_k) \end{aligned}$$

- make change of variables $z_k = x_k - y_k$:

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_f(x_k - z_k) \\ z_{k+1} &= \operatorname{prox}_{g^*}(z_k + 2x_{k+1} - x_k) \end{aligned}$$

Outline

- Douglas–Rachford splitting method
- **examples**
- alternating direction method of multipliers
- image deblurring example
- convergence

Sparse inverse covariance selection

$$\text{minimize} \quad \operatorname{tr}(CX) - \log\det X + \gamma \sum_{i>j} |X_{ij}|$$

variable is $X \in \mathbf{S}^n$; parameters $C \in \mathbf{S}^n_+$ and $\gamma > 0$ are given

Douglas–Rachford splitting

$$f(X) = \operatorname{tr}(CX) - \log\det X, \qquad g(X) = \gamma \sum_{i>j} |X_{ij}|$$

- $X = \operatorname{prox}_{tf}(\hat{X})$ is positive solution of $C - X^{-1} + (1/t)(X - \hat{X}) = 0$

  easily solved via eigenvalue decomposition of $\hat{X} - tC$ (see homework)

- $X = \operatorname{prox}_{tg}(\hat{X})$ is soft-thresholding

Spingarn's method of partial inverses

Equality constrained convex problem (f closed and convex; V a subspace)

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in V \end{array}$$

Spingarn's method: Douglas–Rachford splitting with $g = \delta_V$ (indicator of V)

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{tf}(y_k) \\ y_{k+1} &= y_k + P_V(2x_{k+1} - y_k) - x_{k+1} \end{aligned}$$

Primal-dual form (algorithm of page 11.8):

$$\begin{aligned} x_{k+1} &= \operatorname{prox}_{tf}(x_k - z_k) \\ z_{k+1} &= P_{V^\perp}(z_k + 2x_{k+1} - x_k) \end{aligned}$$

Application to composite optimization problem

$$\text{minimize} \quad f_1(x) + f_2(Ax)$$

f_1 and f_2 have simple prox-operators

- problem is equivalent to minimizing $f(x_1, x_2)$ over subspace V where

$$f(x_1, x_2) = f_1(x_1) + f_2(x_2), \qquad V = \{(x_1, x_2) \mid x_2 = Ax_1\}$$

- prox_{tf} is separable:

$$\mathrm{prox}_{tf}(x_1, x_2) = \left(\mathrm{prox}_{tf_1}(x_1),\ \mathrm{prox}_{tf_2}(x_2)\right)$$

- projection of (x_1, x_2) on V reduces to linear equation:

$$\begin{aligned} P_V(x_1, x_2) &= \begin{bmatrix} I \\ A \end{bmatrix} (I + A^T A)^{-1}(x_1 + A^T x_2) \\ &= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} A^T \\ -I \end{bmatrix} (I + AA^T)^{-1}(x_2 - Ax_1) \end{aligned}$$

Decomposition of separable problems

$$\text{minimize} \quad \sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} g_i(A_{i1}x_1 + \cdots + A_{in}x_n)$$

- same problem as page 10.17, but without strong convexity assumption
- we assume the functions f_j and g_i have inexpensive prox-operators

Equivalent formulation

$$\begin{array}{ll} \text{minimize} & \sum\limits_{j=1}^{n} f_j(x_j) + \sum\limits_{i=1}^{m} g_i(y_{i1} + \cdots + y_{in}) \\ \text{subject to} & y_{ij} = A_{ij}x_j, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n \end{array}$$

- prox-operator of first term requires evaluations of prox_{tf_j} for $j = 1, \ldots, n$
- prox-operator of 2nd term requires prox_{ntg_i} for $i = 1, \ldots, m$ (see page 6.8)
- projection on constraint set reduces to n independent linear equations

Decomposition of separable problems

Second equivalent formulation: introduce extra splitting variables x_{ij}

$$\begin{array}{ll} \text{minimize} & \sum\limits_{j=1}^{n} f_j(x_j) + \sum\limits_{i=1}^{m} g_i(y_{i1} + \cdots + y_{in}) \\ \text{subject to} & x_{ij} = x_j, \quad\quad i = 1, \ldots, m, \quad j = 1, \ldots, n \\ & y_{ij} = A_{ij} x_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n \end{array}$$

- make first set of constraints part of domain of f_j:

$$\tilde{f}_j(x_j, x_{1j}, \ldots, x_{mj}) = \begin{cases} f_j(x_j) & x_{ij} = x_j, \quad i = 1, \ldots, m \\ +\infty & \text{otherwise} \end{cases}$$

  prox-operator of $\tilde{f}_j$ reduces to prox-operator of f_j

- projection on other constraints involves mn independent linear equations

Outline

- Douglas–Rachford splitting method
- examples
- **alternating direction method of multipliers**
- image deblurring example
- convergence

Dual application of Douglas–Rachford method

Separable convex problem

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) \\ \text{subject to} & A_1 x_1 + A_2 x_2 = b \end{array}$$

Dual problem

$$\text{maximize} \quad -b^T z - f_1^*(-A_1^T z) - f_2^*(-A_2^T z)$$

we apply the Douglas–Rachford method (page 11.5) to minimize

$$\underbrace{b^T z + f_1^*(-A_1^T z)}_{g(z)} + \underbrace{f_2^*(-A_2^T z)}_{f(z)}$$

Douglas–Rachford applied to the dual

$$u^+ = \mathrm{prox}_{tg}(z + w), \qquad z^+ = \mathrm{prox}_{tf}(u^+ - w), \qquad w^+ = w + z^+ - u^+$$

First line: use result on page 8.7 to compute $u^+ = \mathrm{prox}_{tg}(z + w)$

$$\begin{aligned} \hat{x}_1 &= \underset{x_1}{\mathrm{argmin}} \left(f_1(x_1) + z^T(A_1 x_1 - b) + \frac{t}{2}\|A_1 x_1 - b + w/t\|_2^2 \right) \\ u^+ &= z + w + t(A_1 \hat{x}_1 - b) \end{aligned}$$

Second line: similarly, compute $z^+ = \mathrm{prox}_{tf}(z + t(A_1\hat{x}_1 - b))$

$$\begin{aligned} \hat{x}_2 &= \underset{x_2}{\mathrm{argmin}} \left(f_2(x_2) + z^T A_2 x_2 + \frac{t}{2}\|A_1 \hat{x}_1 + A_2 x_2 - b\|_2^2 \right) \\ z^+ &= z + t(A_1 \hat{x}_1 + A_2 \hat{x}_2 - b) \end{aligned}$$

Third line reduces to $w^+ = tA_2\hat{x}_2$

Alternating direction method of multipliers (ADMM)

update $x_k = (x_{k,1}, x_{k,2})$ and z_k in three steps

1. minimize augmented Lagrangian over $\tilde{x}_1$

$$x_{k+1,1} = \operatorname*{argmin}_{\tilde{x}_1} \left(f_1(\tilde{x}_1) + z_k^T A_1 \tilde{x}_1 + \frac{t}{2} \|A_1 \tilde{x}_1 + A_2 x_{k,2} - b\|_2^2 \right)$$

2. minimize augmented Lagrangian over $\tilde{x}_2$

$$x_{k+1,2} = \operatorname*{argmin}_{\tilde{x}_2} \left(f_2(\tilde{x}_2) + z_k^T A_2 \tilde{x}_2 + \frac{t}{2} \|A_1 x_{k+1,1} + A_2 \tilde{x}_2 - b\|_2^2 \right)$$

3. dual update

$$z_{k+1} = z_k + t(A_1 x_{k+1,1} + A_2 x_{k+1,2} - b)$$

this the alternating direction method of multipliers

Comparison with other multiplier methods

Alternating minimization method (page 10.22) with $g(y) = \delta_{\{b\}}(y)$

- same dual update, same update for $x_{k+1,2}$
- x_1-update in alternating minimization method is simpler:

$$x_{k+1,1} = \underset{\tilde{x}_1}{\operatorname{argmin}} \left(f_1(\tilde{x}_1) + z_k^T A_1 \tilde{x}_1 \right)$$

- ADMM does not require strong convexity of f_1
- in theory, parameter t in ADMM can be any positive constant

Augmented Lagrangian method (page 10.23) with $g(y) = \delta_{\{b\}}(y)$

- same dual update
- update $x_{k+1,1}$, $x_{k+1,2}$ requires joint minimization of the augmented Lagrangian

$$f_1(\tilde{x}_1) + f_2(\tilde{x}_2) + z_k^T(A_1\tilde{x}_1 + A_2\tilde{x}_2) + \frac{t}{2} \|A_1\tilde{x}_1 + A_2\tilde{x}_2 - b\|_2^2$$

Application to composite optimization (method 1)

$$\text{minimize} \quad f_1(x) + f_2(Ax)$$

- apply ADMM to

$$\begin{array}{ll} \text{minimize} & f_1(x_1) + f_2(x_2) \\ \text{subject to} & Ax_1 = x_2 \end{array}$$

- augmented Lagrangian is

$$f_1(x_1) + f_2(x_2) + \frac{t}{2}\|Ax_1 - x_2 + z/t\|_2^2$$

- x_1-update requires (possibly nontrivial) minimization of

$$f_1(x_1) + \frac{t}{2}\|Ax_1 - x_2 + z/t\|_2^2$$

- x_2-update is evaluation of $\text{prox}_{t^{-1}f_2}$

Application to composite optimization (method 2)

introduce an extra "splitting" variable x_3

$$\begin{array}{ll} \text{minimize} & f_1(x_3) + f_2(x_2) \\ \text{subject to} & \begin{bmatrix} A \\ I \end{bmatrix} x_1 = \begin{bmatrix} x_2 \\ x_3 \end{bmatrix} \end{array}$$

- alternate minimization of augmented Lagrangian over x_1 and (x_2, x_3)

$$f_1(x_3) + f_2(x_2) + \frac{t}{2}\left(\|Ax_1 - x_2 + z_1/t\|_2^2 + \|x_1 - x_3 + z_2/t\|_2^2\right)$$

- x_1-update: linear equation with coefficient $I + A^T A$

- (x_2, x_3)-update: decoupled evaluations of $\operatorname{prox}_{t^{-1}f_1}$ and $\operatorname{prox}_{t^{-1}f_2}$

Outline

- Douglas–Rachford splitting method
- examples
- alternating direction method of multipliers
- **image deblurring example**
- convergence

Image blurring model

$$b = Kx_{\mathrm{t}} + w$$

- x_{t} is unknown image
- b is observed (blurred and noisy) image; w is noise
- $N \times N$-images are stored in column-major order as vectors of length N^2

Blurring matrix K

- represents 2D convolution with space-invariant point spread function
- with periodic boundary conditions, block-circulant with circulant blocks
- can be diagonalized by multiplication with unitary 2D DFT matrix W:

$$K = W^H \,\mathbf{diag}(\lambda) W$$

equations with coefficient $I + K^T K$ can be solved in $O(N^2 \log N)$ time

Total variation deblurring with 1-norm

$$\begin{array}{ll} \text{minimize} & \|Kx - b\|_1 + \gamma \|Dx\|_{\text{tv}} \\ \text{subject to} & 0 \preceq x \preceq \mathbf{1} \end{array}$$

second term in objective is *total variation* penalty

- Dx is discretized first derivative in vertical and horizontal direction

$$D = \begin{bmatrix} I \otimes D_1 \\ D_1 \otimes I \end{bmatrix}, \qquad D_1 = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 1 & -1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & -1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}$$

- $\|\cdot\|_{\text{tv}}$ is a sum of Euclidean norms: $\|(u, v)\|_{\text{tv}} = \sum_{i=1}^{n} \sqrt{u_i^2 + v_i^2}$

Solution via Douglas–Rachford method

an example of a composite optimization problem

$$\text{minimize} \quad f_1(x) + f_2(Ax)$$

with f_1 the indicator of $[0,1]^n$ and

$$A = \begin{bmatrix} K \\ D \end{bmatrix}, \qquad f_2(u,v) = \|u\|_1 + \gamma\|v\|_{\mathrm{tv}}$$

Primal DR method (page 11.11) and **ADMM** (page 11.19) require:

- decoupled prox-evaluations of $\|u\|_1$ and $\gamma\|v\|_{\mathrm{tv}}$, and projections on C
- solution of linear equations with coefficient matrix

$$I + K^T K + D^T D$$

solvable in $O(N^2 \log N)$ time

Example

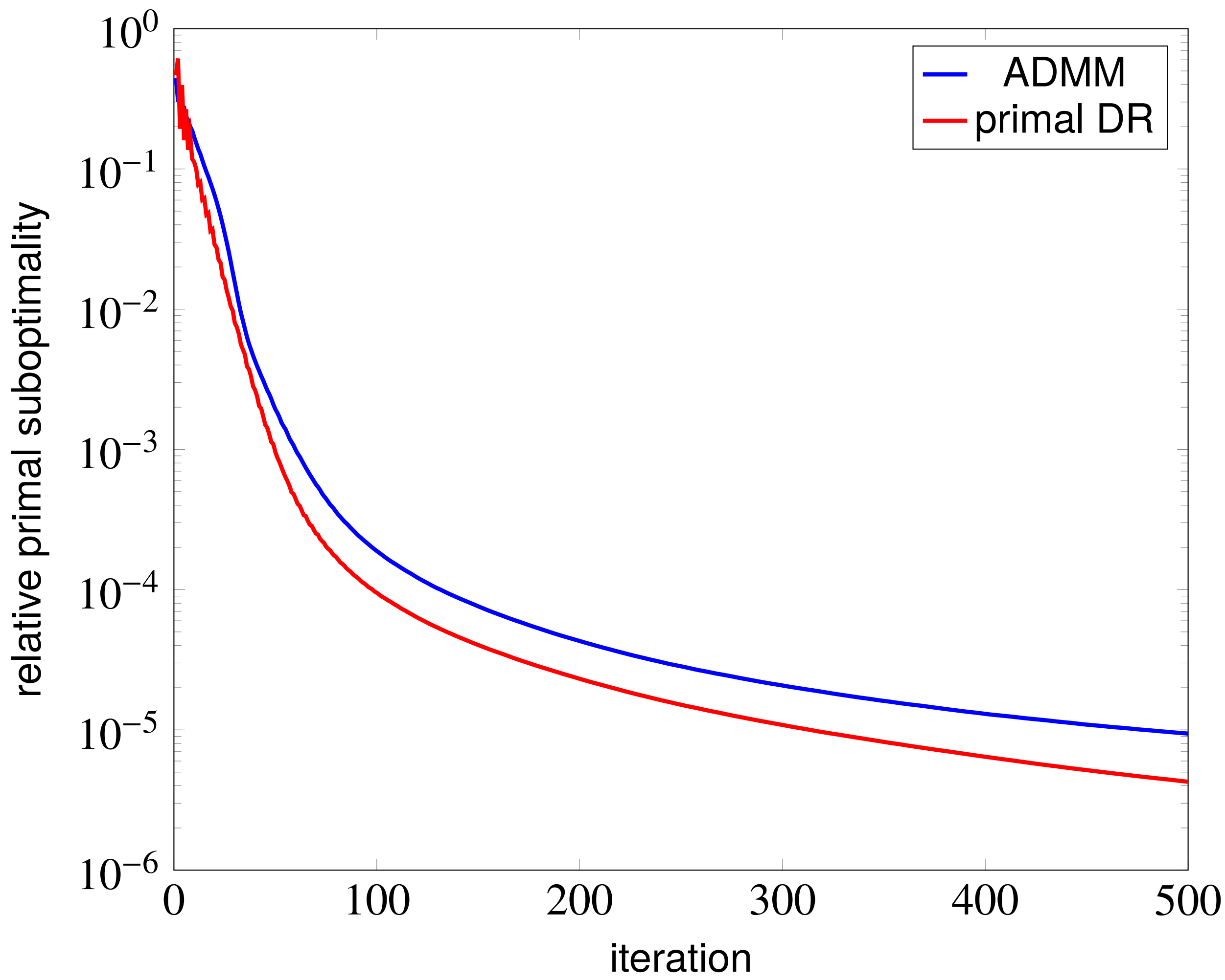
- 1024×1024 image, periodic boundary conditions
- Gaussian blur
- salt-and-pepper noise (50% pixels randomly changed to $0/1$)

original

noisy/blurred

restored

Convergence



cost per iteration is dominated by 2-D FFTs

Outline

- Douglas–Rachford splitting method
- examples
- alternating direction method of multipliers
- image deblurring example
- **convergence**

Douglas–Rachford iteration mappings

define iteration map F and negative step G (in notation of page 11.7)

$$
\begin{aligned}
F(y) &= y + \mathrm{prox}_g(2\mathrm{prox}_f(y) - y) - \mathrm{prox}_f(y) \\
G(y) &= y - F(y) \\
&= \mathrm{prox}_f(y) - \mathrm{prox}_g(2\mathrm{prox}_f(y) - y)
\end{aligned}
$$

- F is firmly nonexpansive (co-coercive with parameter 1)

$$(F(y) - F(\hat{y}))^T(y - \hat{y}) \geq \|F(y) - F(\hat{y})\|_2^2 \qquad \text{for all } y, \hat{y}$$

- this implies that G is firmly nonexpansive:

$$
\begin{aligned}
&(G(y) - G(\hat{y}))^T(y - \hat{y}) \\
&\quad = \|G(y) - G(\hat{y})\|_2^2 + (F(y) - F(\hat{y}))^T(y - \hat{y}) - \|F(y) - F(\hat{y})\|_2^2 \\
&\quad \geq \|G(y) - G(\hat{y})\|_2^2
\end{aligned}
$$

Proof (of firm nonexpansiveness of F).

- define $x = \mathrm{prox}_f(y)$, $\hat{x} = \mathrm{prox}_f(\hat{y})$, and

$$v = \mathrm{prox}_g(2x - y), \qquad \hat{v} = \mathrm{prox}_g(2\hat{x} - \hat{y})$$

- substitute expressions $F(y) = y + v - x$ and $F(\hat{y}) = \hat{y} + \hat{v} - \hat{x}$:

$$\begin{aligned}
&(F(y) - F(\hat{y}))^T(y - \hat{y}) \\
&\quad \geq \quad (y + v - x - \hat{y} - \hat{v} + \hat{x})^T(y - \hat{y}) - (x - \hat{x})^T(y - \hat{y}) + \|x - \hat{x}\|_2^2 \\
&\quad = \quad (v - \hat{v})^T(y - \hat{y}) + \|y - x - \hat{y} + \hat{x}\|_2^2 \\
&\quad = \quad (v - \hat{v})^T(2x - y - 2\hat{x} + \hat{y}) - \|v - \hat{v}\|_2^2 + \|F(y) - F(\hat{y})\|_2^2 \\
&\quad \geq \quad \|F(y) - F(\hat{y})\|_2^2
\end{aligned}$$

inequalities use firm nonexpansiveness of prox_f and prox_g (page 4.9):

$$(x - \hat{x})^T(y - \hat{y}) \geq \|x - \hat{x}\|_2^2, \qquad (2x - y - 2\hat{x} + \hat{y})^T(v - \hat{v}) \geq \|v - \hat{v}\|_2^2$$

Convergence result

$$
\begin{aligned}
y_{k+1} &= (1-\rho_k)y_k + \rho_k F(y_k) \\
&= y_k - \rho_k G(y_k)
\end{aligned}
$$

Assumptions

- F has fixed points (points x that satisfy $0 \in \partial f(x) + \partial g(x)$)
- $\rho_k \in [\rho_{\min}, \rho_{\max}]$ with $0 < \rho_{\min} < \rho_{\max} < 2$

Result

- y_k converges to a fixed point $y^\star$ of F
- $x_{k+1} = \mathrm{prox}_f(y_k)$ converges to a solution $x^\star = \mathrm{prox}_f(y^\star)$

  (follows from continuity of prox_f)

Proof: let $y^\star$ be any fixed point of $F(y)$ (zero of $G(y)$)

consider iteration k (with $y = y_k$, $\rho = \rho_k$, $y^+ = y_{k+1}$):

$$\begin{aligned}
\|y^+ - y^\star\|_2^2 - \|y - y^\star\|_2^2 &= 2(y^+ - y)^T(y - y^\star) + \|y^+ - y\|_2^2 \\
&= -2\rho G(y)^T(y - y^\star) + \rho^2\|G(y)\|_2^2 \\
&\leq -\rho(2-\rho)\|G(y))\|_2^2 \\
&\leq -M\|G(y))\|_2^2 \qquad (1)
\end{aligned}$$

where $M = \rho_{\min}(2 - \rho_{\max})$ (on line 3 we use firm nonexpansiveness of G)

- (1) implies that

$$M\sum_{k=0}^{\infty}\|G(y_k)\|_2^2 \leq \|y_0 - y^\star\|_2^2, \qquad \|G(y_k)\|_2 \to 0$$

- (1) implies that $\|y_k - y^\star\|_2$ is nonincreasing; hence y_k is bounded
- since $\|y_k - y^\star\|_2$ is nonincreasing, the limit $\lim_{k\to\infty}\|y_k - y^\star\|_2$ exists

Proof (continued)

- since the sequence y_k is bounded, it has a convergent subsequence
- let $\bar{y}_k$ be a convergent subsequence with limit $\bar{y}$; by continuity of G,

$$0 = \lim_{k\to\infty} G(\bar{y}_k) = G(\bar{y})$$

hence, $\bar{y}$ is a zero of G and the limit $\lim_{k\to\infty} \|y_k - \bar{y}\|_2$ exists

- let $\bar{u}$ and $\bar{v}$ be two limit points; the limits

$$\lim_{k\to\infty} \|y_k - \bar{u}\|_2, \qquad \lim_{k\to\infty} \|y_k - \bar{v}\|_2$$

exist, and subsequences of y_k converge to $\bar{u}$, resp. $\bar{v}$; therefore

$$\|\bar{u} - \bar{v}\|_2 = \lim_{k\to\infty} \|y_k - \bar{u}\|_2 = \lim_{k\to\infty} \|y_k - \bar{v}\|_2 = 0$$

References

- H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces* (2017).
- 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 (2010).
- P.L. Combettes and J.-C. Pesquet, *A Douglas–Rachford splitting approach to nonsmooth convex variational signal recovery,* IEEE Journal of Selected Topics in Signal Procesing (2007).
- J. Eckstein and D. Bertsekas, *On the Douglas–Rachford splitting method and the proximal algorithm for maximal monotone operators,* Mathematical Programming (1992).
- D. Gabay, *Applications of the method of multipliers to variational inequalities*. In *Studies in Mathematics and Its Applications* (1983).
- P. L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators,* SIAM Journal on Numerical Analysis (1979).
- J. E. Spingarn, *Applications of the method of partial inverses to convex programming: decomposition,* Mathematical Programming (1985).
- The image deblurring example is from D. O'Connor and L. Vandenberghe, *Primal-dual decomposition by operator splitting and applications to image deblurring,* SIAM J. Imaging Sciences (2014).