Douglas-Rachford method, ADMM and PDHG

Acknowledgement: this slides is based on Prof. Lieven Vandenberghes lecture notes

Outline

1. Douglas-Rachford splitting method
2. examples
3. alternating direction method of multipliers
4. image deblurring example
5. convergence

Douglas-Rachford splitting algorithm

Consider

$$\min \quad f(x) = g(x) + h(x)$$

g and h are closed convex functions

Douglas-Rachford iteration: starting at any $z^{(0)}$, repeat

$$\begin{aligned} x^{(k)} &= \operatorname{prox}_{th}(z^{(k-1)}) \\ y^{(k)} &= \operatorname{prox}_{tg}(2x^{(k)} - z^{(k-1)}) \\ z^{(k)} &= z^{(k-1)} + y^{(k)} - x^{(k)} \end{aligned}$$

- t is a positive constant (simply scales the objective)
- useful when g and h have inexpensive prox-operators
- under weak conditions (existence of a minimizer), $x^{(k)}$ converges

Equivalent form

- start iteration at y-update

$$y^+ = \text{prox}_{tg}(2x - z); \quad z^+ = z + y^+ - x; \quad x^+ = \text{prox}_{th}(z^+)$$

- switch z- and x-updates

$$y^+ = \text{prox}_{tg}(2x - z); \quad x^+ = \text{prox}_{th}(z + y^+ - x); \quad z^+ = z + y^+ - x$$

- make change of variables $w = z - x$

alternate form of DR iteration: start at $x^{(0)} \in \textbf{dom}\, h, w^{(0)} \in t\partial h(x^{(0)})$

$$\begin{aligned} y^+ &= \text{prox}_{tg}(x - w) \\ x^+ &= \text{prox}_{th}(y^+ + w) \\ w^+ &= w + y^+ - x^+ \end{aligned}$$

Interpretation as fixed-point iteration

Douglas-Rachford iteration can be written as

$$z^{(k)} = F(z^{(k-1)})$$

where $F(z) = z + \text{prox}_{tg}(2\text{prox}_{th}(z) - z) - \text{prox}_{th}(z)$

fixed points of F and minimizers of $g + h$

- if z is a fixed point, then $x = \text{prox}_{th}(z)$ is a minimizer:

$$\begin{aligned} z = F(z), \quad x = prox_{th}(z) &\Rightarrow \text{prox}_{tg}(2x - z) = x = \text{prox}_{th}(z) \\ &\Rightarrow x - z \in t\partial g(x); z - x \in t\partial h(x) \\ &\Rightarrow 0 \in t\partial g(x) + t\partial h(x) \end{aligned}$$

- if x is a minimizer and $u \in t\partial g(x) \cap -t\partial h(x)$, then $x - u = F(x - u)$

Douglas-Rachford iteration with relaxation

fixed-point iteration with relaxation

$$z^+ = z + \rho(F(z) - z)$$

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

first version of DR method

$$\begin{aligned} x^+ &= \mathrm{prox}_{th}(z) \\ y^+ &= \mathrm{prox}_{tg}(2x^+ - z) \\ z^+ &= z + \rho(y^+ - x^+) \end{aligned}$$

alternate version

$$\begin{aligned} y^+ &= \mathrm{prox}_{tg}(x - w) \\ x^+ &= \mathrm{prox}_{th}((1 - \rho)x + \rho y^+ + w) \\ w^+ &= w + \rho y^+ + (1 - \rho)x - x^+ \end{aligned}$$

Outline

1. Douglas-Rachford splitting method
2. examples
3. alternating direction method of multipliers
4. image deblurring example
5. convergence

Sparse inverse covariance selection

$$\min \quad \mathbf{tr}(CX) - \log\det X + \rho \sum_{i>j} |X_{ij}|$$

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

Douglas-Rachford splitting

$$g(X) = \mathbf{tr}(CX) - \log\det X, h(x) = \rho \sum_{i>j} |X_{ij}|$$

- $X = \mathrm{prox}_{tg}(\hat{X})$ is positive solution of $C - X^{-1} + (1/t)(X - \hat{X}) = 0$ easily solved via eigenvalue decomposition of $\hat{X} - tC$
- $X = \mathrm{prox}_{th}(\hat{X})$ is soft-thresholding

Spingarn's method of partial inverses

equality constrained convex problem

$$\begin{array}{ll} \min & h(x) \\ \text{s.t.} & x \in V \end{array}$$

h a closed convex function; V a subspace

Douglas-Rachford splitting: take $g = I_V$ (indicator of V)

$$\begin{aligned} x^+ &= \text{prox}_{th}(z) \\ y^+ &= P_V(2x^+ - z) \\ z^+ &= z + y^+ - x^+ \end{aligned}$$

Application to composite optimization problem

$$\min f_1(x) + f_2(Ax)$$

f_1 and f_2 have simple prox-operators

- equivalent to minimizing $h(x_1, x_2)$ over subspace V where

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

- prox_{th} is separable: $\mathrm{prox}_{th}(x_1, x_2) = (\mathrm{prox}_{tf_1}(x_1), \mathrm{prox}_{tf_2}(x_2))$
- projection of (x_1, x_2) on V reduces to linear equation:

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

Decomposition of separable problems

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

- same problem as the lecture on "dual proximal gradient method", but without strong convexity assumption
- we assume the functions f_j and g_i have inexpensive prox-operators

equivalent formulation

$$\begin{array}{ll} \min & \sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} g_i(y_{i1} + \cdots + y_{in}) \\ \text{s.t.} & y_{ij} = A_{ij}x_j, \quad i = 1, \cdots, m; \quad j = 1, \cdots, n \end{array}$$

- prox-operator of cost involves uncoupled prox-evaluations for f_j, g_i
- projection on constraint set reduces to n independent linear equations

Decomposition of separable problems

second equivalent formulation with extra splitting variables x_{ij}:

$$
\begin{array}{ll}
\min & \sum_{j=1}^{n} f_j(x_j) + \sum_{i=1}^{m} g_i(y_{i1} + \cdots + y_{in}) \\
\text{s.t.} & x_{ij} = x_j, \quad i = 1, \cdots, m; \quad j = 1, \cdots, n \\
 & y_{ij} = A_{ij} x_{ij}, \quad i = 1, \cdots, m; \quad j = 1, \cdots, n
\end{array}
$$

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

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

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

- projection on other constraints involves mn independent linear equations

Outline

Dual application of Douglas-Rachford method

separable convex problem

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) \\ \text{s.t.} & A_1x_1 + A_2x_2 = b \end{array}$$

dual problem

$$\max \quad -b^T z - f_1^*(-A_1^T z) - f_2^*(-A_2^T z)$$

we apply the Douglas-Rachford method (page 3) to minimize

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

Douglas Rachford on the dual

$$y^+ = \text{prox}_{tg}(z - w), \quad z^+ = \text{prox}_{th}(y^+ + w), \quad w^+ = w + y^+ - z^+$$

first line: use result in "lect-dualProxGrad.pdf" to compute $y^+ = \text{prox}_{tg}(z - w)$

$$\begin{aligned} \hat{x_1} &= \underset{x_1}{\text{argmin}}(f_1(x_1) + z^T(A_1x_1 - b) + \frac{t}{2}\|A_1x_1 - b - w/t\|_2^2) \\ y^+ &= z - w + t(A_1\hat{x_1} - b) \end{aligned}$$

second line: similarly, compute $z^+ = \text{prox}_{th}(z + t(A_1\hat{x_1} - b))$

$$\begin{aligned} \hat{x_2} &= \underset{x_1}{\text{argmin}}(f_1(x_2) + z^TA_2x_2 + \frac{t}{2}\|A_1\hat{x_1} + A_2x_2 - b\|_2^2) \\ z^+ &= z + t(A_1\hat{x_1} + A_2\hat{x_2} - b) \end{aligned}$$

third line reduces to $w^+ = -tA_2\hat{x_2}$

Alternating direction method of multipliers

1. minimize augmented Lagrangian over x_1

$$x_1^{(k)} = \underset{x_1}{\mathrm{argmin}} \left(f_1(x_1) + (z^{(k-1)})^T A_1 x_1 + \frac{t}{2} \|A_1 x_1 + A_2 x_2^{(k-1)} - b\|_2^2 \right)$$

2. minimize augmented Lagrangian over x_2

$$x_2^{(k)} = \underset{x_2}{\mathrm{argmin}} \left(f_2(x_2) + (z^{(k-1)})^T A_2 x_2 + \frac{t}{2} \|A_1 x_1^{(k)} + A_2 x_2 - b\|_2^2 \right)$$

3. dual update

$$z^{(k)} = z^{(k-1)} + t(A_1 x_1^{(k)} + A_2 x_2^{(k)} - b)$$

also known as split Bregman method

Comparison with other multiplier methods

alternating minimization method with $g(y) = I_{\{b\}}(y)$

- same dual update, same update for x_2
- x_1-update in alternating minimization method is simpler:

$$x_1^{(k)} = \underset{x_1}{\operatorname{argmin}} \left(f_1(x_1) + (z^{(k-1)})^T A_1 x_1 \right)$$

- ADMM does not require strong convexity of f_1

augmented Lagrangian method with $g(y) = I_{\{b\}}(y)$

- dual update is the same
- AL method requires joint minimization of the augmented Lagrangian

$$\min_{x_1, x_2} f_1(x_1) + f_2(x_2) + (z^{(k-1)})^T (A_1 x_1 + A_2 x_2) + \frac{t}{2} \|A_1 x_1 + A_2 x_2 - b\|_2^2$$

Application to composite optimization (method 1)

$$\min \quad f_1(x) + f_2(Ax)$$

apply ADMM to

$$\begin{array}{ll} \min & f_1(x_1) + f_2(x_2) \\ \text{s.t.} & Ax_1 = x_2 \end{array}$$

- 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 minimization of $f_1(x_1) + \frac{t}{2}\|Ax_1 - x_2 + z/t\|_2^2$
- x_2-update is evaluation of $\mathrm{prox}_{t^{-1}f_2}$

Application to composite optimization (method 2)

introduce extra 'splitting' or 'dummy' variable x_3

$$\begin{array}{ll} \min & f_1(x_3) + f_2(x_2) \\ \text{s.t.} & \begin{pmatrix} A \\ I \end{pmatrix} x_1 = \begin{pmatrix} x_2 \\ x_3 \end{pmatrix} \end{array}$$

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

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

- x_1-update: linear equation with coefficient $I + A^T A$
- (x_2, x_3)-update: decoupled evaluations of $\mathrm{prox}_{t^{-1}f_1}$ and $\mathrm{prox}_{t^{-1}f_2}$

Outline

Image blurring model

$$b = Kx_t + w$$

- 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

$$\begin{array}{ll} \min & \|Kx - b\|_1 + \gamma \|Dx\|_{tv} \\ \text{s.t.} & 0 \leq x \leq 1 \end{array}$$

second term in objective is **total variation penalty**

- Dx is discretized first derivative in vertical and horizontal direction

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

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

Solution via Douglas-Rachford method

an example of a composite optimization problem

$$\min \quad f_1(x) + f_2(Ax)$$

with f_1 the indicator of $[0,1]^n$ and $A = \begin{pmatrix} K \\ D \end{pmatrix}$,$f_2(u,v) = \|u\|_1 + \gamma\|v\|_{tv}$

$$\min \ \|u\|_1 + \gamma\|v\|_{tv}, \ \text{s.t. } u = Kx - b, \ v = Dx, y = x, 0 \leq y \leq 1$$

primal DR method and **ADMM** require:

- decoupled prox-evaluations of $\|u\|_1$ and $\|v\|_{tv}$, 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

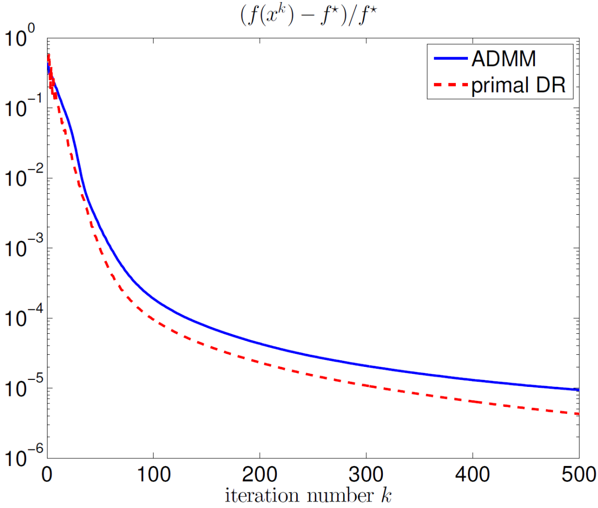
- 1024 $\times$ 1024 image, periodic boundary conditions
- Gaussian blur
- salt-and-pepper noise (50% pixels randomly changed to 0/1)


original


noisy/blurred


restored

Convergence



cost per iteration is dominated by 2D FFTs

Outline

1. Douglas-Rachford splitting method
2. examples
3. alternating direction method of multipliers
4. image deblurring example
5. convergence

Nonexpansiveness

if $u = \text{prox}_h(x), v = \text{prox}_h(y)$, then

$$(u - v)^\top (x - y) \geq \|u - v\|2$$

prox_h is *firmly nonexpansive*, or *co-coercive* with constant 1

- follows from characterization of proximal mapping and monotonicity

$$x - u \in \partial h(u), y - v \in \partial h(v) \quad \Rightarrow \quad (x - u - y + v)^\top (u - v) \geq 0$$

- implies (from Cauchy-Schwarz inequality)

$$\|\text{prox}_h(x) - prox_h(y)\|_2 \leq \|x - y\|_2$$

prox_h is *nonexpansive*, or *Lipschitz continuous* with constant 1

Douglas-Rachford iteration mappings

define iteration map F and negative step G

$$
\begin{aligned}
F(z) &= z + \mathrm{prox}_{tg}(2\mathrm{prox}_{th}(z) - z) - \mathrm{prox}_{th}(z) \\
G(z) &= z - F(z) \\
&= \mathrm{prox}_{th}(z) - \mathrm{prox}_{tg}(2\mathrm{prox}_{th}(z) - z)
\end{aligned}
$$

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

$$
(F(z) - F(\hat{z}))^T(z - \hat{z}) \geq \|F(z) - F(\hat{z})\|_2^2 \quad \forall z, \hat{z}
$$

- implies that G is firmly nonexpansive:

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

Proof.

firm nonexpansiveness of F

- define $x = \text{prox}_{th}(z), \hat{x} = \text{prox}_{th}(\hat{z})$, and

$$y = \text{prox}_{tg}(2x - z), \quad \hat{y} = \text{prox}_{tg}(2\hat{x} - \hat{z})$$

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

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

inequalities use firm nonexpansiveness of prox_{th} and prox_{tg}

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

□

Convergence result

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

assumptions

- optimal value $f^* = \inf_x (g(x) + h(x))$ is finite and attained
- $\rho_k \in [\rho_{\min}, \rho_{\max}]$ with $0 < \rho_{\min} < \rho_{\max} < 2$

result

- $z^{(k)}$ converges to a fixed point z^* of F
- $x^{(k)} = \text{prox}_{th}(z^{(k-1)})$ converges to a minimizer $x^* = \text{prox}_{th}(z^*)$ (follows from continuity of prox_{th})

Proof.

Let z^* be any fixed point of $F(z)$ (zero of $G(z)$). Consider iteration k (with $z = z^{(k-1)}, \rho = \rho_k, z^+ = z^{(k)}$):

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

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

- (1) implies that

$$
M \sum_{k=0}^{\infty} \|G(z^{(k)})\|_2^2 \leq \|z^{(0)} - z^*\|_2^2, \quad \|G(z^{(k)})\|_2 \to 0
$$

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

continued.

- since the sequence $z^{(k)}$ is bounded, it has a convergent subsequence
- let $\bar{z_k}$ be a convergent subsequence with limit $\bar{z}$; by continuity of G,

$$0 = \lim_{k\to\infty} G(\bar{z_k}) = G(\bar{z})$$

hence, $\bar{z}$ is a zero of G and the limit $\lim_{k\to\infty} \|z^{(k)} - \bar{z}\|_2$ exists

- let $\bar{z_1}$ and $\bar{z_2}$ be two limit points; the limits

$$\lim_{k\to\infty} \|z^{(k_{j_1})} - \bar{z_1}\|_2, \quad \lim_{k\to\infty} \|z^{(k_{j_2})} - \bar{z_2}\|_2$$

exist, and subsequences of $z^{(k)}$ converge to $\bar{z_1}$, resp. $\bar{z_2}$; therefore

$$\|\bar{z_2} - \bar{z_1}\|_2 = \lim_{k\to\infty} \|z^{(k)} - \bar{z_1}\|_2 = \lim_{k\to\infty} \|z^{(k)} - \bar{z_2}\|_2 = 0$$

□

References

Douglas-Rachford method, ADMM, Spingarn's method

- J. E. Spingarn, *Applications of the method of partial inverses to convex programming: decomposition*, Mathematical Programming (1985)
- J. Eckstein and D. Bertsekas, *On the Douglas-Rachford splitting method and the proximal algorithm for maximal monotone operators*, Mathematical Programming (1992)
- 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)
- S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, *Distributed optimization and statistical learning via the alternating direction method of multipliers* (2010)
- N. Parikh, S. Boyd, *Block splitting for distributed optimization* (2013)

image deblurring: the example is taken from
D. O'Connor and L. Vandenberghe, *Primal-dual decomposition by operator splitting and applications to image deblurring* (2014)

Primal-Dual Hybrid Gradient Methods

Consider

$$(P) \qquad \min_{x\in\mathbb{R}^n} \quad f_1(Bx) + f_2(x)$$

- Saddle point form by Fenchel duality

$$(PD) \qquad \min_{x\in\mathbb{R}^n} \sup_{y} \quad -f_1^*(y) + \langle y, Bx\rangle + f_2(x)$$

- The dual problem is

$$(D) \qquad \max_{y} -f_1^*(y) - f_2^*(-B^\top y)$$

- Saddle point form by introducing new variable $z = Bx$ in (P)

$$\max_{y} \min_{x,z\in\mathbb{R}^n} \quad f_1(z) + f_2(x) + \langle y, Bx - z\rangle$$

- Saddle point form by introducing new variable $p = B^\top y$ in (D).

$$\max_{y} \min_{y,p} -f_1^*(y) - f_2^*(p) + \left\langle x, -B^\top y - p\right\rangle$$

Primal-Dual method from (PD)

Saddle point form by Fenchel duality

$$(PD) \qquad \min_{x\in\mathbb{R}^n} \sup_{y} \quad -f_1^*(y) + \langle y, Bx\rangle + f_2(x)$$

- Primal-dual hybrid Gradient (PDHG) Method

$$\begin{aligned} x^{k+1} &= \arg\min_{x} \quad \left\langle y^k, Bx\right\rangle + f_2(x) + \frac{1}{2\alpha_k}\|x - x^k\|_2^2 \\ y^{k+1} &= \arg\max_{y} \quad -f_1^*(y) + \left\langle y, Bx^{k+1}\right\rangle - \frac{1}{2\beta_k}\|y - y^k\|_2^2 \end{aligned}$$

- Interpret PDHG as a primal-dual proximal point method for finding a saddle point of (PD)
- Convergence: Esser-Z-Chan, 2010; He-Yuan, 2010; Bonettini-Ruggiero

Modified PDHG

Saddle point form by Fenchel duality

$$(PD) \qquad \min_{x \in \mathbb{R}^n} \sup_{y} \quad -f_1^*(y) + \langle y, Bx \rangle + f_2(x)$$

- Modified Primal-dual hybrid Gradient (PDHG) Method

$$\begin{aligned} x^{k+1} &= \arg\min_{x} \quad \left\langle 2y^k - y^{k-1}, Bx \right\rangle + f_2(x) + \frac{1}{2\alpha_k}\|x - x^k\|_2^2 \\ y^{k+1} &= \arg\max_{y} \quad -f_1^*(y) + \left\langle y, Bx^k \right\rangle - \frac{1}{2\beta_k}\|y - y^k\|_2^2 \end{aligned}$$

- Esser-Z-Chan, 2010; A. Chambolle and T. Pock, 2010