Primal-dual algorithms for the sum of two and three functions[1]

Ming Yan
Michigan State University, CMSE/Mathematics

[1] This works is partially supported by NSF.

optimization problems for primal-dual algorithms

$$\underset{\mathbf{x}}{\text{minimize}}\ f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{A}\mathbf{x})$$

- f, g, and h are convex.
- $\mathcal{X}$ and $\mathcal{Y}$ are two Hilbert spaces (e.g., $\mathbf{R}^m$, $\mathbf{R}^n$).
- $f : \mathcal{X} \mapsto \mathbf{R}$ is differentiable with a $1/\beta$-Lipschitz continuous gradient for some $\beta \in (0, +\infty)$.
- $\mathbf{A} : \mathcal{X} \mapsto \mathcal{Y}$ is a bounded linear operator.

applications: statistics

Elastic net regularization (Zou-Hastie '05):

$$\underset{\mathbf{x}}{\text{minimize}} \;\; \mu_2\|\mathbf{x}\|_2^2 + \mu_1\|\mathbf{x}\|_1 + l(\mathbf{A}\mathbf{x}, \mathbf{b}),$$

where $\mathbf{x} \in \mathbf{R}^p$, $\mathbf{A} \in \mathbf{R}^{n \times p}$, $\mathbf{b} \in \mathbf{R}^n$, and l is the loss function, which may be nondifferentiable.

applications: statistics

Elastic net regularization (Zou-Hastie '05):

$$\underset{\mathbf{x}}{\text{minimize}} \;\; \mu_2\|\mathbf{x}\|_2^2 + \mu_1\|\mathbf{x}\|_1 + l(\mathbf{A}\mathbf{x}, \mathbf{b}),$$

where $\mathbf{x} \in \mathbf{R}^p$, $\mathbf{A} \in \mathbf{R}^{n\times p}$, $\mathbf{b} \in \mathbf{R}^n$, and l is the loss function, which may be nondifferentiable.

Fused lasso (Tibshirani et al. '05):

$$\underset{\mathbf{x}}{\text{minimize}} \;\; \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \mu_1\|\mathbf{x}\|_1 + \mu_2\|\mathbf{D}\mathbf{x}\|_1,$$

where $\mathbf{x} \in \mathbf{R}^p$, $\mathbf{A} \in \mathbf{R}^{n\times p}$, $\mathbf{b} \in \mathbf{R}^n$, and

$$\mathbf{D} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \cdots & \cdots & \\ & & & -1 & 1 \end{pmatrix}$$

is a matrix in $\mathbf{R}^{(p-1)\times p}$.

applications: decentralized optimization

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x) + g_i(x)$$

- f_i and g_i is known at node i only.
- Nodes $1, \cdots, n$ are connected in a undirected graph.
- f_i is differentiable with a Lipschitz continuous gradient.

applications: decentralized optimization

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x) + g_i(x)$$

- f_i and g_i is known at node i only.
- Nodes $1, \cdots, n$ are connected in a undirected graph.
- f_i is differentiable with a Lipschitz continuous gradient.

Introduce a copy x_i at node i:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{x}) := \sum_{i=1}^{n} f_i(x_i) + g_i(x_i) \quad \text{s.t. } \mathbf{W}\mathbf{x} = \mathbf{x}$$

- $x_i \in \mathbb{R}^p$, $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_n]^\top \in \mathbb{R}^{n \times p}$.
- $\mathbf{W}$ is a symmetric doubly stochastic mixing matrix.

applications: decentralized optimization

$$\underset{x}{\text{minimize}} \sum_{i=1}^{n} f_i(x) + g_i(x)$$

- f_i and g_i is known at node i only.
- Nodes $1, \cdots, n$ are connected in a undirected graph.
- f_i is differentiable with a Lipschitz continuous gradient.

Introduce a copy x_i at node i:

$$\underset{\mathbf{x}}{\text{minimize}} \; f(\mathbf{x}) + g(\mathbf{x}) := \sum_{i=1}^{n} f_i(x_i) + g_i(x_i) \quad \text{s.t. } \mathbf{W}\mathbf{x} = \mathbf{x}$$

- $x_i \in \mathbb{R}^p$, $\mathbf{x} = [x_1 \; x_2 \; \cdots \; x_n]^\top \in \mathbb{R}^{n \times p}$.
- $\mathbf{W}$ is a symmetric doubly stochastic mixing matrix.

The sum of three functions:

$$\underset{\mathbf{x}}{\text{minimize}} \; f(\mathbf{x}) + g(\mathbf{x}) + \iota_{\mathbf{0}}((\mathbf{I} - \mathbf{W})^{1/2}\mathbf{x})$$

applications: imaging

Image restoration with two regularizations:

$$\underset{\mathbf{x}}{\text{minimize}} \, \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \iota_C(\mathbf{x}) + \mu\|\mathbf{D}\mathbf{x}\|_1,$$

where $\mathbf{x} \in \mathbf{R}^n$ is the image to be reconstructed, $\mathbf{A} \in \mathbf{R}^{m\times n}$ is the forward projection matrix, $\mathbf{b} \in \mathbf{R}^m$ is the measured data with noise, $\mathbf{D}$ is a discrete gradient operator, and ι_C is the indicator function that returns zero if $\mathbf{x} \in C$ (here, C is the set of nonnegative vectors in $\mathbf{R}^n$) and $+\infty$ otherwise.

applications: imaging

Image restoration with two regularizations:

$$\underset{\mathbf{x}}{\text{minimize}}\ \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \iota_C(\mathbf{x}) + \mu\|\mathbf{D}\mathbf{x}\|_1,$$

where $\mathbf{x} \in \mathbf{R}^n$ is the image to be reconstructed, $\mathbf{A} \in \mathbf{R}^{m\times n}$ is the forward projection matrix, $\mathbf{b} \in \mathbf{R}^m$ is the measured data with noise, $\mathbf{D}$ is a discrete gradient operator, and ι_C is the indicator function that returns zero if $\mathbf{x} \in C$ (here, C is the set of nonnegative vectors in $\mathbf{R}^n$) and $+\infty$ otherwise.

Other problems:

- f: data fitting term (infimal convolution for mixed noise)
- $h \circ \mathbf{A}$: total variation; other transforms
- g: nonnegativity; box constraint

primal-dual formulation

Original problem:

$$\underset{\mathbf{x}}{\text{minimize}} \;\; f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{A}\mathbf{x})$$

primal-dual formulation

Original problem:

$$\underset{\mathbf{x}}{\text{minimize}} \;\; f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{A}\mathbf{x})$$

Introduce a dual variable $\mathbf{s}$:

$$\underset{\mathbf{x}}{\text{minimize}} \;\; \max_{\mathbf{s}} f(\mathbf{x}) + g(\mathbf{x}) + \langle \mathbf{A}\mathbf{x}, \mathbf{s} \rangle - h^*(\mathbf{s})$$

Here h^* is the conjugate function of h that is defined as

$$h^*(\mathbf{s}) = \max_{\mathbf{t}} \;\; \langle \mathbf{s}, \mathbf{t} \rangle - h(\mathbf{t}),$$

primal-dual formulation

Original problem:

$$\underset{\mathbf{x}}{\text{minimize}} \;\; f(\mathbf{x}) + g(\mathbf{x}) + h(\mathbf{A}\mathbf{x})$$

Introduce a dual variable $\mathbf{s}$:

$$\underset{\mathbf{x}}{\text{minimize}} \;\; \max_{\mathbf{s}} f(\mathbf{x}) + g(\mathbf{x}) + \langle \mathbf{A}\mathbf{x}, \mathbf{s} \rangle - h^*(\mathbf{s})$$

Here h^* is the conjugate function of h that is defined as

$$h^*(\mathbf{s}) = \max_{\mathbf{t}} \;\; \langle \mathbf{s}, \mathbf{t} \rangle - h(\mathbf{t}),$$

It is equivalent to ($\mathbf{s}^* \in \partial h(\mathbf{A}\mathbf{x}^*) \iff \mathbf{A}\mathbf{x}^* \in \partial h^*(\mathbf{s}^*)$):

$$\begin{cases} \mathbf{0} \in & \nabla f(\mathbf{x}^*) + \partial g(\mathbf{x}^*) + \mathbf{A}^\top \mathbf{s}^* \\ \mathbf{0} \in & \partial h^*(\mathbf{s}^*) - \mathbf{A}\mathbf{x}^* \end{cases}$$

All primal-dual algorithms try to find $(\mathbf{x}^*, \mathbf{s}^*)$.

existing algorithms: Condat-Vu, AFBA, and PDFP

Condat-Vu (Condat '13, Vu '13):

- Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\| + \gamma/(2\beta) \leq 1$
- Per-iteration computations: $\mathbf{A}$, $\mathbf{A}^\top$, ∇f, one $(\mathbf{I}+\gamma\partial g)^{-1}$, $(\mathbf{I}+\frac{\lambda}{\gamma}\partial h^*)^{-1}$ [2]

---

[2]

$$(\mathbf{I}+\gamma\partial g)^{-1}(\tilde{\mathbf{x}}) = \arg\min_{\mathbf{x}} \gamma g(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\tilde{\mathbf{x}}\|^2.$$

This is a backward step (or implicit step) because $(\mathbf{I}+\gamma\partial g)^{-1}(\tilde{\mathbf{x}}) \in \tilde{\mathbf{x}} - \gamma\partial g((\mathbf{I}+\gamma\partial g)^{-1}(\tilde{\mathbf{x}}))$

existing algorithms: Condat-Vu, AFBA, and PDFP

Condat-Vu (Condat '13, Vu '13):

- Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\| + \gamma/(2\beta) \leq 1$
- Per-iteration computations: $\mathbf{A}$, $\mathbf{A}^\top$, ∇f, one $(\mathbf{I} + \gamma\partial g)^{-1}$, $(\mathbf{I} + \frac{\lambda}{\gamma}\partial h^*)^{-1}$ [2]

AFBA (Latafat-Patrinos '16):

- Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\|/2 + \sqrt{\lambda\|\mathbf{A}\mathbf{A}^\top\|}/2 + \gamma/(2\beta) \leq 1$
- Per-iteration computations: $\mathbf{A}$, $\mathbf{A}^\top$, ∇f, one $(\mathbf{I} + \gamma\partial g)^{-1}$, $(\mathbf{I} + \frac{\lambda}{\gamma}\partial h^*)^{-1}$

2

$$(\mathbf{I} + \gamma\partial g)^{-1}(\tilde{\mathbf{x}}) = \arg\min_{\mathbf{x}} \gamma g(\mathbf{x}) + \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{x}}\|^2.$$

This is a backward step (or implicit step) because $(\mathbf{I} + \gamma\partial g)^{-1}(\tilde{\mathbf{x}}) \in \tilde{\mathbf{x}} - \gamma\partial g((\mathbf{I} + \gamma\partial g)^{-1}(\tilde{\mathbf{x}}))$

existing algorithms: Condat-Vu, AFBA, and PDFP

Condat-Vu (Condat '13, Vu '13):

- Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\| + \gamma/(2\beta) \leq 1$
- Per-iteration computations: $\mathbf{A}$, $\mathbf{A}^\top$, ∇f, one $(\mathbf{I}+\gamma\partial g)^{-1}$, $(\mathbf{I}+\frac{\lambda}{\gamma}\partial h^*)^{-1}$ [2]

AFBA (Latafat-Patrinos '16):

- Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\|/2 + \sqrt{\lambda\|\mathbf{A}\mathbf{A}^\top\|}/2 + \gamma/(2\beta) \leq 1$
- Per-iteration computations: $\mathbf{A}$, $\mathbf{A}^\top$, ∇f, one $(\mathbf{I}+\gamma\partial g)^{-1}$, $(\mathbf{I}+\frac{\lambda}{\gamma}\partial h^*)^{-1}$

PDFP (Chen-Huang-Zhang '16):

- Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\| < 1$; $\gamma/(2\beta) < 1$
- Per-iteration computations: $\mathbf{A}$, $\mathbf{A}^\top$, ∇f, two $(\mathbf{I}+\gamma\partial g)^{-1}$, $(\mathbf{I}+\frac{\lambda}{\gamma}\partial h^*)^{-1}$

2

$$(\mathbf{I}+\gamma\partial g)^{-1}(\tilde{\mathbf{x}}) = \arg\min_{\mathbf{x}} \gamma g(\mathbf{x}) + \frac{1}{2}\|\mathbf{x}-\tilde{\mathbf{x}}\|^2.$$

This is a backward step (or implicit step) because $(\mathbf{I}+\gamma\partial g)^{-1}(\tilde{\mathbf{x}}) \in \tilde{\mathbf{x}} - \gamma\partial g((\mathbf{I}+\gamma\partial g)^{-1}(\tilde{\mathbf{x}}))$

PDHG (Zhu-Chan '08)

When $f = 0$, we have

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \mathbf{0}$$

PDHG (Zhu-Chan '08)

When $f = 0$, we have

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \mathbf{0}$$

It is equivalent to

$$\begin{bmatrix} \partial g & \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \begin{bmatrix} \mathbf{0} & -\mathbf{A}^\top \\ & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix}$$

PDHG (Zhu-Chan '08)

When $f = 0$, we have

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \mathbf{0}$$

It is equivalent to

$$\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} + \partial g & \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} + \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \mathbf{A}^\top \\ & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix}$$

PDHG (Zhu-Chan '08)

When $f = 0$, we have

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \mathbf{0}$$

It is equivalent to

$$\begin{bmatrix} \frac{1}{\gamma}\mathbf{I}+\partial g & \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I}+\partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \mathbf{A}^\top \\ & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix}$$

Primal-dual hybrid gradient (PDHG)

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top \mathbf{s})$$

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1} \left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\mathbf{x}^+\right)$$

PDHG (Zhu-Chan '08)

When $f = 0$, we have

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \mathbf{0}$$

It is equivalent to

$$\begin{bmatrix} \frac{1}{\gamma}\mathbf{I}+\partial g & \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I}+\partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \mathbf{A}^\top \\ & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix}$$

Chambolle-Pock (Chambolle et.al '09, Esser-Zhang-Chan '10)

$$\begin{aligned} \mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s}) \\ \bar{\mathbf{x}}^+ &= \mathbf{x}^+ + \mathbf{x}^+ - \mathbf{x} \\ \mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}^+\right) \end{aligned}$$

Chambolle-Pock '11 as proximal point

Chambolle-Pock ($\mathbf{x} - \mathbf{s}$ order)

$$\mathbf{x}^+ = (\mathbf{I} + \gamma \partial g)^{-1}(\mathbf{x} - \gamma \mathbf{A}^\top \mathbf{s})$$

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{x})\right)$$

Chambolle-Pock '11 as proximal point

Chambolle-Pock ($\mathbf{x} - \mathbf{s}$ order)

$$\mathbf{x}^+ = (\mathbf{I} + \gamma \partial g)^{-1}(\mathbf{x} - \gamma \mathbf{A}^\top \mathbf{s})$$
$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma} \partial h^*\right)^{-1} \left(\mathbf{s} + \tfrac{\lambda}{\gamma} \mathbf{A}(2\mathbf{x}^+ - \mathbf{x})\right)$$

CP is equivalent to the backward operator applied on the KKT system.

Chambolle-Pock '11 as proximal point

Chambolle-Pock ($\mathbf{x} - \mathbf{s}$ order)

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s})$$
$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{x})\right)$$

CP is equivalent to the backward operator applied on the KKT system.

$$\left[\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} + \begin{bmatrix} \partial g & \\ -\mathbf{A} & \partial h^* \end{bmatrix}\right] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$$

- CP is 1/2-averaged under the metric induced by the matrix if λ satisfies the condition $\lambda\|\mathbf{A}\mathbf{A}^\top\| \leq 1$.

Chambolle-Pock '11 as proximal point

Chambolle-Pock ($\mathbf{x} - \mathbf{s}$ order)

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s})$$
$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{x})\right)$$

CP is equivalent to the backward operator applied on the KKT system.

$$\left[\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} + \begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix}\right]\begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix}$$

- CP is 1/2-averaged under the metric induced by the matrix if λ satisfies the condition $\lambda\|\mathbf{A}\mathbf{A}^\top\| \le 1$.

Condat-Vu (Condat '13, Vu '13)

The optimality condition:

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ 0 \end{bmatrix} \ni \mathbf{0}$$

CV is equivalent to the forward-backward applied on the KKT system.

$$\left[\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} + \begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix}\right] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} - \begin{bmatrix} \nabla f(\mathbf{x}) \\ 0 \end{bmatrix}$$

Condat-Vu (Condat '13, Vu '13)

The optimality condition:

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ 0 \end{bmatrix} \ni \mathbf{0}$$

CV is equivalent to the forward-backward applied on the KKT system.

$$\left[\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} + \begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix}\right] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} - \begin{bmatrix} \nabla f(\mathbf{x}) \\ 0 \end{bmatrix}$$

That is:

$$\mathbf{x}^+ = (\mathbf{I} + \gamma \partial g)^{-1}(\mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s})$$

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{x})\right)$$

It is equivalent to (by changing the update order)

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = (\mathbf{I} + \gamma \partial g)^{-1}(\mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s}^+)$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x}$$

Condat-Vu (Condat '13, Vu '13)

The optimality condition:

$$\begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ 0 \end{bmatrix} \ni \mathbf{0}$$

CV is equivalent to the forward-backward applied on the KKT system.

$$\left[\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} + \begin{bmatrix} \partial g & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix}\right] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & -\mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} - \begin{bmatrix} \nabla f(\mathbf{x}) \\ 0 \end{bmatrix}$$

That is:

$$\begin{aligned} \mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}) \\ \mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{x})\right) \end{aligned}$$

It is equivalent to (by changing the update order)

$$\begin{aligned} \mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right) \\ \mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+) \\ \bar{\mathbf{x}}^+ &= 2\mathbf{x}^+ - \mathbf{x} \end{aligned}$$

- CV is non-expansive (forward-backward) under the metric induced by the matrix if γ and λ satisfy the condition $\lambda\|\mathbf{A}\mathbf{A}^\top\| + \gamma/(2\beta) \leq 1$.

PDFP^{2}O/PAPC (Loris-Verhoeven '11, Chen-Huang-Zhang '13, Drori-Sabach-Teboulle '15)

When $g = 0$, the optimality condition becomes:

$$\begin{bmatrix} 0 & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ 0 \end{bmatrix} \ni \mathbf{0}$$

PDFP^{2}O/PAPC (Loris-Verhoeven '11, Chen-Huang-Zhang '13, Drori-Sabach-Teboulle '15)

When $g = 0$, the optimality condition becomes:

$$\begin{bmatrix} 0 & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ 0 \end{bmatrix} \ni \mathbf{0}$$

PAPC is equivalent to the forward-backward applied on the KKT system.

$$\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} - \gamma\mathbf{A}\mathbf{A}^\top + \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \\ & \frac{\gamma}{\lambda}\mathbf{I} - \gamma\mathbf{A}\mathbf{A}^\top \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} - \begin{bmatrix} \nabla f(\mathbf{x}) \\ 0 \end{bmatrix}$$

PDFP^{2}O/PAPC (Loris-Verhoeven '11, Chen-Huang-Zhang '13, Drori-Sabach-Teboulle '15)

When $g = 0$, the optimality condition becomes:

$$\begin{bmatrix} 0 & \mathbf{A}^\top \\ -\mathbf{A} & \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^* \\ \mathbf{s}^* \end{bmatrix} + \begin{bmatrix} \nabla f(\mathbf{x}^*) \\ 0 \end{bmatrix} \ni \mathbf{0}$$

PAPC is equivalent to the forward-backward applied on the KKT system.

$$\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \mathbf{A}^\top \\ -\mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} - \gamma\mathbf{A}\mathbf{A}^\top + \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \\ & \frac{\gamma}{\lambda}\mathbf{I} - \gamma\mathbf{A}\mathbf{A}^\top \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} - \begin{bmatrix} \nabla f(\mathbf{x}) \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \mathbf{A}^\top \\ & \frac{\gamma}{\lambda}\mathbf{I} + \partial h^* \end{bmatrix} \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{s}^+ \end{bmatrix} \ni \begin{bmatrix} \frac{1}{\gamma}\mathbf{I} & \\ \mathbf{A} & \frac{\gamma}{\lambda}\mathbf{I} - \gamma\mathbf{A}\mathbf{A}^\top \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{s} \end{bmatrix} - \begin{bmatrix} \nabla f(\mathbf{x}) \\ \gamma\mathbf{A}\nabla f(\mathbf{x}) \end{bmatrix}$$

- PAPC is non-expansive (forward-backward) under the metric induced by the matrix if γ and λ satisfy the conditions $\lambda\|\mathbf{A}\mathbf{A}^\top\| \le 1$ and $\gamma/(2\beta) \le 1$.

PAPC

PAPC can be expressed as

$$\begin{aligned}\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\left(\mathbf{x} - \gamma\nabla f(\mathbf{x})\right)\right)\\ \mathbf{x}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+\end{aligned}$$

PAPC

PAPC can be expressed as

$$\begin{aligned}\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\left(\mathbf{x} - \gamma\nabla f(\mathbf{x})\right)\right)\\ \mathbf{x}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+\end{aligned}$$

It is equivalent to

$$\begin{aligned}\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)\\ \mathbf{x}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+\\ \bar{\mathbf{x}}^+ &= \mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+\end{aligned}$$

PAPC

PAPC can be expressed as

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\left(\mathbf{x} - \gamma\nabla f(\mathbf{x})\right)\right)$$
$$\mathbf{x}^+ = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+$$

It is equivalent to

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$
$$\mathbf{x}^+ = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+$$
$$\bar{\mathbf{x}}^+ = \mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+$$

- PAPC is α-averaged under the metric induced by the matrix.
- PAPC converges if γ and λ satisfy the conditions $\lambda\|\mathbf{A}\mathbf{A}^\top\| < 4/3$ and $\gamma/(2\beta) < 1$ (Li-Yan '17).

PDFP (Chen-Huang-Zhang '16)

Rewrite PDFP^{2}O as

$$
\begin{aligned}
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right) \\
\mathbf{x}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+ \\
\bar{\mathbf{x}}^+ &= \mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+
\end{aligned}
$$

PDFP (Chen-Huang-Zhang '16)

Rewrite PDFP^{2}O as

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$
$$\mathbf{x}^+ = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+$$
$$\bar{\mathbf{x}}^+ = \mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+$$

PDFP, as a generalization of PDFP^{2}O, is

$$\mathbf{s}^+ = (\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$
$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+)$$
$$\bar{\mathbf{x}}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

- When g is the indicator function, PDFP reduces to Preconditioned Alternating Projection Algorithm (PAPA) (Krol-Li-Shen-Xu '12).

AFBA (Latafat-Patrinos '16)

Rewrite PAPC as

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = \bar{\mathbf{x}} - \gamma\mathbf{A}^\top(\mathbf{s}^+ - \mathbf{s})$$

$$\bar{\mathbf{x}}^+ = \mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+$$

AFBA (Latafat-Patrinos '16)

Rewrite PAPC as

$$\begin{aligned}
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)\\
\mathbf{x}^+ &= \bar{\mathbf{x}} - \gamma\mathbf{A}^\top(\mathbf{s}^+ - \mathbf{s})\\
\bar{\mathbf{x}}^+ &= \mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+
\end{aligned}$$

AFBA, as a generalization of PAPC, is

$$\begin{aligned}
\mathbf{s}^+ &= (\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)\\
\mathbf{x}^+ &= \bar{\mathbf{x}} - \gamma\mathbf{A}^\top(\mathbf{s}^+ - \mathbf{s})\\
\bar{\mathbf{x}}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+)
\end{aligned}$$

Convergence conditions: $\lambda\|\mathbf{A}\mathbf{A}^\top\|/2 + \sqrt{\lambda\|\mathbf{A}\mathbf{A}^\top\|}/2 + \gamma/(2\beta) \leq 1$

Chambolle-Pock and PAPC

Chambolle-Pock:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x}$$

Chambolle-Pock and PAPC

Chambolle-Pock:

$$\mathbf{s}^{+} = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^{*}\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^{+} = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^{\top}\mathbf{s}^{+})$$

$$\bar{\mathbf{x}}^{+} = 2\mathbf{x}^{+} - \mathbf{x}$$

PAPC:

$$\mathbf{s}^{+} = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^{*}\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^{+} = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^{\top}\mathbf{s}^{+}$$

$$\bar{\mathbf{x}}^{+} = \mathbf{x}^{+} - \gamma\nabla f(\mathbf{x}^{+}) - \gamma\mathbf{A}^{\top}\mathbf{s}^{+}$$

Chambolle-Pock and PAPC

Chambolle-Pock:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x}$$

PAPC:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x} - \gamma\nabla f(\mathbf{x}^+) + \gamma\nabla f(\mathbf{x})$$

Chambolle-Pock and PAPC

Chambolle-Pock:

$$\begin{aligned}
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)\\
\mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s}^+)\\
\bar{\mathbf{x}}^+ &= 2\mathbf{x}^+ - \mathbf{x}
\end{aligned}$$

PAPC:

$$\begin{aligned}
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)\\
\mathbf{x}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+\\
\bar{\mathbf{x}}^+ &= 2\mathbf{x}^+ - \mathbf{x} - \gamma\nabla f(\mathbf{x}^+) + \gamma\nabla f(\mathbf{x})
\end{aligned}$$

PD3O (Yan '16):

$$\begin{aligned}
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)\\
\mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+)\\
\bar{\mathbf{x}}^+ &= 2\mathbf{x}^+ - \mathbf{x} - \gamma\nabla f(\mathbf{x}^+) + \gamma\nabla f(\mathbf{x})
\end{aligned}$$

Chambolle-Pock

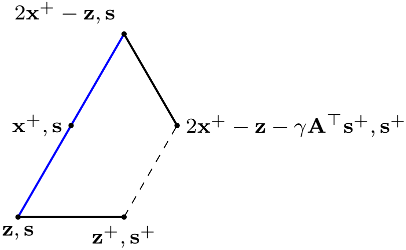
Chambolle-Pock ($\mathbf{x} - \mathbf{s}$ order):

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x}$$

Chambolle-Pock



Chambolle-Pock ($\mathbf{x} - \mathbf{s}$ order):

$$\begin{aligned}
\mathbf{z} &= \mathbf{x} - \gamma\mathbf{A}^\top\mathbf{s} \\
\mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{z}) \\
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{z})\right) \\
\mathbf{z}^+ &= \mathbf{z} + 2\mathbf{x}^+ - \mathbf{z} - \gamma\mathbf{A}^\top\mathbf{s}^+ - \mathbf{x}^+
\end{aligned}$$

C-P and PAPC

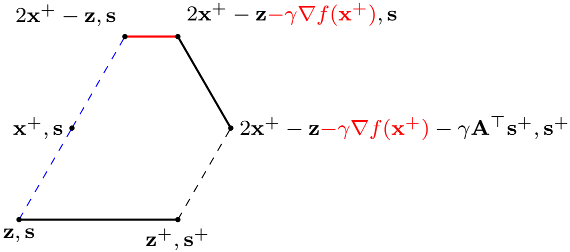
PAPC:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x} - \gamma\nabla f(\mathbf{x}^+) + \gamma\nabla f(\mathbf{x})$$

C-P and PAPC



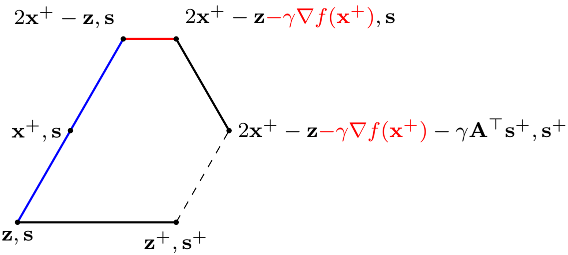
PAPC:

$$\mathbf{x}^+ = \mathbf{z} = \mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s}$$

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma} \partial h^*\right)^{-1} \left((\mathbf{I} - \lambda \mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma} \mathbf{A}(2\mathbf{x}^+ - \mathbf{z} - \gamma \nabla f(\mathbf{x}^+))\right)$$

$$\mathbf{z}^+ = \mathbf{z} + 2\mathbf{x}^+ - \mathbf{z} - \gamma \nabla f(\mathbf{x}^+) - \gamma \mathbf{A}^\top \mathbf{s}^+ - \mathbf{x}^+$$

PD3O



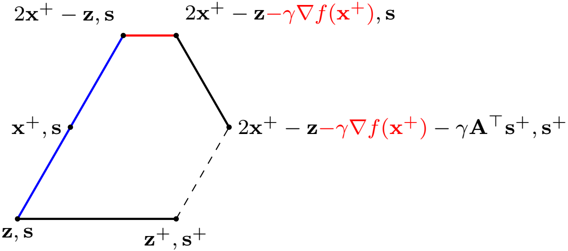
PD3O:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

$$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x} - \gamma\nabla f(\mathbf{x}^+) + \gamma\nabla f(\mathbf{x})$$

PD3O



PD3O:

$$\begin{aligned}
\mathbf{z} &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s} \\
\mathbf{x}^+ &= (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{z}) \\
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}(2\mathbf{x}^+ - \mathbf{z} - \gamma\nabla f(\mathbf{x}^+))\right) \\
\mathbf{z}^+ &= \mathbf{z} + 2\mathbf{x}^+ - \mathbf{z} - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+ - \mathbf{x}^+
\end{aligned}$$

PD3O vs Condat-Vu vs AFBA vs PDFP

Algorithms:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$

$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

PDFP	$\bar{\mathbf{x}}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+)$
Condat-Vu	$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x}$
PD3O	$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x} + \gamma\nabla f(\mathbf{x}) - \gamma\nabla f(\mathbf{x}^+)$

PD3O vs Condat-Vu vs AFBA vs PDFP

Algorithms:

$$\mathbf{s}^+ = \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left(\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\bar{\mathbf{x}}\right)$$
$$\mathbf{x}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+)$$

PDFP	$\bar{\mathbf{x}}^+ = (\mathbf{I} + \gamma\partial g)^{-1}(\mathbf{x}^+ - \gamma\nabla f(\mathbf{x}^+) - \gamma\mathbf{A}^\top\mathbf{s}^+)$
Condat-Vu	$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x}$
PD3O	$\bar{\mathbf{x}}^+ = 2\mathbf{x}^+ - \mathbf{x} + \gamma\nabla f(\mathbf{x}) - \gamma\nabla f(\mathbf{x}^+)$

Parameters:

	$f \neq 0,\ g \neq 0$	$f = 0$	$g = 0$
PDFP	$\lambda\|\mathbf{A}\mathbf{A}^\top\| < 1;\ \gamma/(2\beta) < 1$		PAPC
Condat-Vu	$\lambda\|\mathbf{A}\mathbf{A}^\top\| + \gamma/(2\beta) \leq 1$	C-P	
AFBA	$\lambda\|\mathbf{A}\mathbf{A}^\top\|/2 + \sqrt{\lambda\|\mathbf{A}\mathbf{A}^\top\|/2} + \gamma/(2\beta) \leq 1$		PAPC
PD3O	$\lambda\|\mathbf{A}\mathbf{A}^\top\| \leq 1;\ \gamma/(2\beta) < 1$	C-P	PAPC

convergence results: summary

Let $\mathbf{z} = \mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s}$ and $\mathbf{x}^+ \to \mathbf{x}$:

$$
\begin{aligned}
\mathbf{x} &= (\mathbf{I} + \gamma \partial g)^{-1} \mathbf{z} \\
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma} \partial h^*\right)^{-1} \left((\mathbf{I} - \lambda \mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma} \mathbf{A} \left(2\mathbf{x} - \mathbf{z} - \gamma \nabla f(\mathbf{x})\right)\right) \\
\mathbf{z}^+ &= \mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s}^+
\end{aligned}
$$

convergence results: summary

Let $\mathbf{z} = \mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s}$ and $\mathbf{x}^+ \to \mathbf{x}$:

$$\begin{aligned}
\mathbf{x} &= (\mathbf{I} + \gamma \partial g)^{-1} \mathbf{z} \\
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma} \partial h^*\right)^{-1} \left((\mathbf{I} - \lambda \mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma} \mathbf{A} \left(2\mathbf{x} - \mathbf{z} - \gamma \nabla f(\mathbf{x})\right)\right) \\
\mathbf{z}^+ &= \mathbf{x} - \gamma \nabla f(\mathbf{x}) - \gamma \mathbf{A}^\top \mathbf{s}^+
\end{aligned}$$

- $\|(\mathbf{z}^{k+1}, \mathbf{s}^{k+1}) - (\mathbf{z}^k, \mathbf{s}^k)\|_{\mathbf{M}}^2 = o\left(\frac{1}{k+1}\right)$, and $(\mathbf{z}^k, \mathbf{s}^k)$ weakly converges to a fixed point $(\mathbf{z}^*, \mathbf{s}^*)$

convergence results: summary

Let $\mathbf{z} = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}$ and $\mathbf{x}^+ \to \mathbf{x}$:

$$\begin{aligned}
\mathbf{x} &= (\mathbf{I} + \gamma\partial g)^{-1}\mathbf{z} \\
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\left(2\mathbf{x} - \mathbf{z} - \gamma\nabla f(\mathbf{x})\right)\right) \\
\mathbf{z}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+
\end{aligned}$$

- $\|(\mathbf{z}^{k+1}, \mathbf{s}^{k+1}) - (\mathbf{z}^k, \mathbf{s}^k)\|_{\mathbf{M}}^2 = o\left(\frac{1}{k+1}\right)$, and $(\mathbf{z}^k, \mathbf{s}^k)$ weakly converges to a fixed point $(\mathbf{z}^*, \mathbf{s}^*)$
- Let $\mathcal{L}(\mathbf{x}, \mathbf{s}) = f(\mathbf{x}) + g(\mathbf{x}) + \langle\mathbf{A}\mathbf{x}, \mathbf{s}\rangle - h^*(\mathbf{s})$, then

$$\mathcal{L}(\bar{\mathbf{x}}^k, \mathbf{s}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{s}}^{k+1}) \le \frac{\|(\mathbf{z}^1, \mathbf{s}^1) - (\mathbf{z}, \mathbf{s})\|^2}{k}$$

  where $(\bar{\mathbf{x}}^k, \bar{\mathbf{s}}^{k+1}) = \frac{1}{k}\sum_{i=1}^k(\mathbf{x}^i, \mathbf{s}^{i+1})$, and $\mathbf{z} = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}$.

convergence results: summary

Let $\mathbf{z} = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}$ and $\mathbf{x}^+ \to \mathbf{x}$:

$$\begin{aligned}
\mathbf{x} &= (\mathbf{I} + \gamma\partial g)^{-1}\mathbf{z} \\
\mathbf{s}^+ &= \left(\mathbf{I} + \tfrac{\lambda}{\gamma}\partial h^*\right)^{-1}\left((\mathbf{I} - \lambda\mathbf{A}\mathbf{A}^\top)\mathbf{s} + \tfrac{\lambda}{\gamma}\mathbf{A}\left(2\mathbf{x} - \mathbf{z} - \gamma\nabla f(\mathbf{x})\right)\right) \\
\mathbf{z}^+ &= \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}^+
\end{aligned}$$

- $\|(\mathbf{z}^{k+1}, \mathbf{s}^{k+1}) - (\mathbf{z}^k, \mathbf{s}^k)\|_{\mathbf{M}}^2 = o\left(\frac{1}{k+1}\right)$, and $(\mathbf{z}^k, \mathbf{s}^k)$ weakly converges to a fixed point $(\mathbf{z}^*, \mathbf{s}^*)$
- Let $\mathcal{L}(\mathbf{x}, \mathbf{s}) = f(\mathbf{x}) + g(\mathbf{x}) + \langle\mathbf{A}\mathbf{x}, \mathbf{s}\rangle - h^*(\mathbf{s})$, then

$$\mathcal{L}(\bar{\mathbf{x}}^k, \mathbf{s}) - \mathcal{L}(\mathbf{x}, \bar{\mathbf{s}}^{k+1}) \le \frac{\|(\mathbf{z}^1, \mathbf{s}^1) - (\mathbf{z}, \mathbf{s})\|^2}{k}$$

  where $(\bar{\mathbf{x}}^k, \bar{\mathbf{s}}^{k+1}) = \frac{1}{k}\sum_{i=1}^k(\mathbf{x}^i, \mathbf{s}^{i+1})$, and $\mathbf{z} = \mathbf{x} - \gamma\nabla f(\mathbf{x}) - \gamma\mathbf{A}^\top\mathbf{s}$.
- Linear convergence with additional assumptions on f, g, and h

convergence analysis: the general case

- Let $\mathbf{M} = \frac{\gamma^2}{\lambda}(\mathbf{I} - \lambda \mathbf{A}\mathbf{A}^\top)$ be positive definite. Define $\|\mathbf{s}\|_\mathbf{M} = \sqrt{\langle \mathbf{s}, \mathbf{s} \rangle_\mathbf{M}} = \sqrt{\langle \mathbf{s}, \mathbf{M}\mathbf{s} \rangle}$ and $\|(\mathbf{z}, \mathbf{s})\|_\mathbf{M} = \sqrt{\|\mathbf{z}\|^2 + \|\mathbf{s}\|_\mathbf{M}^2}$.

Lemma

The iteration $\mathbf{T}$ mapping $(\mathbf{z}, \mathbf{s})$ to $(\mathbf{z}^+, \mathbf{s}^+)$ is a nonexpansive operator under the metric defined by $\mathbf{M}$ if $\gamma \leq 2\beta$. Furthermore, it is α-averaged with $\alpha = \frac{2\beta}{4\beta - \gamma}$.

- Chambolle-Pock is firmly non-expansive under the new metric, which is different from the previous metric.

convergence analysis: the general case

Theorem

1) *Let* $(\mathbf{z}^*, \mathbf{s}^*)$ *be any fixed point of* $\mathbf{T}$*. Then* $(\|(\mathbf{z}^k, \mathbf{s}^k) - (\mathbf{z}^*, \mathbf{s}^*)\|_{\mathbf{M}})_{k\geq 0}$ *is monotonically nonincreasing.*
2) *The sequence* $(\|\mathbf{T}(\mathbf{z}^k, \mathbf{s}^k) - (\mathbf{z}^k, \mathbf{s}^k)\|_{\mathbf{M}})_{k\geq 0}$ *is monotonically nonincreasing and converges to 0.*
3) *We have the following convergence rate*

$$\|\mathbf{T}(\mathbf{z}^k, \mathbf{s}^k) - (\mathbf{z}^k, \mathbf{s}^k)\|_{\mathbf{M}}^2 = o\left(\frac{1}{k+1}\right)$$

4) $(\mathbf{z}^k, \mathbf{s}^k)$ *weakly converges to a fixed point of* $\mathbf{T}$*, and if* $\mathcal{X}$ *has finite dimension (e.g.,* $\mathbf{R}^m$*), then it is strongly convergent.*

convergence analysis: linear convergent

Denote:

$$\begin{aligned}
\mathbf{u}_h =& \tfrac{\gamma}{\lambda}(\mathbf{I} - \lambda \mathbf{A}\mathbf{A}^\top)\mathbf{s} + \mathbf{A}\tilde{\mathbf{z}} - \tfrac{\gamma}{\lambda}\mathbf{s}^+ \in \partial h^*(\mathbf{s}^+),\\
\mathbf{u}_g =& \tfrac{1}{\gamma}(\mathbf{z} - \mathbf{x}) \in \partial g(\mathbf{x}),\\
\mathbf{u}_h^* =& \mathbf{A}(\tilde{\mathbf{z}}^* - \gamma \mathbf{A}^\top \mathbf{s}^*) = \mathbf{A}\mathbf{x}^* \in \partial h^*(\mathbf{s}^*),\\
\mathbf{u}_g^* =& \tfrac{1}{\gamma}(\mathbf{z}^* - \mathbf{x}^*) \in \partial g(\mathbf{x}^*).
\end{aligned}$$

and

$$\begin{aligned}
\|\nabla g(\mathbf{x}) - \nabla g(\mathbf{y})\| \leq & L_g \|\mathbf{x} - \mathbf{y}\|,\\
\langle \mathbf{s}^+ - \mathbf{s}^*, \mathbf{u}_h - \mathbf{u}_h^* \rangle \geq & \tau_h \|\mathbf{s}^+ - \mathbf{s}^*\|_{\mathbf{M}}^2,\\
\langle \mathbf{x} - \mathbf{x}^*, \mathbf{u}_g - \mathbf{u}_g^* \rangle \geq & \tau_g \|\mathbf{x} - \mathbf{x}^*\|^2,\\
\langle \mathbf{x} - \mathbf{x}^*, \nabla f(\mathbf{x}) - \nabla f(\mathbf{x}^*) \rangle \geq & \tau_f \|\mathbf{x} - \mathbf{x}^*\|.
\end{aligned}$$

convergence analysis: linear convergent

Theorem

We have

$$\|\mathbf{z}^+ - \mathbf{z}^*\|^2 + (1 + 2\gamma\tau_h)\|\mathbf{s}^+ - \mathbf{s}^*\|_{\mathbf{M}}^2 \le \rho\left(\|\mathbf{z} - \mathbf{z}^*\|^2 + (1 + 2\gamma\tau_h)\|\mathbf{s} - \mathbf{s}^*\|_{\mathbf{M}}^2\right)$$

where

$$\rho = \max\left(\frac{1}{1+2\gamma\tau_h}, 1 - \frac{\left(\left(2\gamma - \frac{\gamma^2}{\beta}\right)\tau_f + 2\gamma\tau_g\right)}{1+\gamma L_g}\right). \tag{5}$$

When, in addition, $\gamma < 2\beta$, $\tau_h > 0$, *and* $\tau_f + \tau_g > 0$, *we have that* $\rho < 1$ *and the algorithm converges linearly.*

numerical experiment: fused lasso

$$\underset{\mathbf{x}}{\text{minimize}} \; \frac{1}{2}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2 + \mu_1\|\mathbf{x}\|_1 + \mu_2\sum_{i=1}^{p-1}|x_{i+1}-x_i|,$$

- $\mathbf{x} = (x_1, \cdots, x_p) \in \mathbf{R}^p$, $\mathbf{A} \in \mathbf{R}^{n\times p}$, $\mathbf{b} \in \mathbf{R}^n$

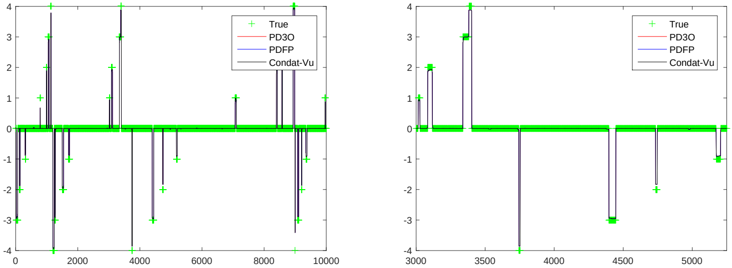


Figure: The true sparse signal and the reconstructed results using PD3O, PDFP, and Condat-Vu. The right figure is a zoom-in of the signal in [3000, 5500].

numerical experiment: fused lasso

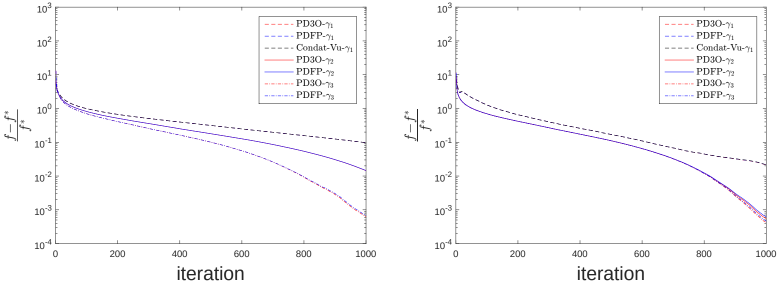


Figure: In the left figure, we fix $\lambda = 1/8$ and let $\gamma = \beta,\ 1.5\beta,\ 1.9\beta$. In the right figure, we fix $\gamma = 1.9\beta$ and let $\lambda = 1/80,\ 1/8,\ 1/4$.

applications: decentralized optimization

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x) + g_i(x)$$

- f_i and g_i is known at node i only.
- Nodes $1, \cdots, n$ are connected in a undirected graph.
- f_i is differentiable with a Lipschitz continuous gradient.

applications: decentralized optimization

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{n} f_i(x) + g_i(x)$$

- f_i and g_i is known at node i only.
- Nodes $1, \cdots, n$ are connected in a undirected graph.
- f_i is differentiable with a Lipschitz continuous gradient.

Introduce a copy x_i at node i:

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) + g(\mathbf{x}) := \sum_{i=1}^{n} f_i(x_i) + g_i(x_i) \quad \text{s.t. } \mathbf{W}\mathbf{x} = \mathbf{x}$$

- $x_i \in \mathbb{R}^p$, $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]^\top \in \mathbb{R}^{n \times p}$.
- $\mathbf{W}$ is a symmetric doubly stochastic mixing matrix.

applications: decentralized optimization

$$\underset{x}{\text{minimize}} \ \sum_{i=1}^{n} f_i(x) + g_i(x)$$

- f_i and g_i is known at node i only.
- Nodes $1, \cdots, n$ are connected in a undirected graph.
- f_i is differentiable with a Lipschitz continuous gradient.

Introduce a copy x_i at node i:

$$\underset{\mathbf{x}}{\text{minimize}} \ f(\mathbf{x}) + g(\mathbf{x}) := \sum_{i=1}^{n} f_i(x_i) + g_i(x_i) \quad \text{s.t. } \mathbf{W}\mathbf{x} = \mathbf{x}$$

- $x_i \in \mathbb{R}^p$, $\mathbf{x} = [x_1 \ x_2 \ \cdots \ x_n]^\top \in \mathbb{R}^{n \times p}$.
- $\mathbf{W}$ is a symmetric doubly stochastic mixing matrix.

The sum of three functions:

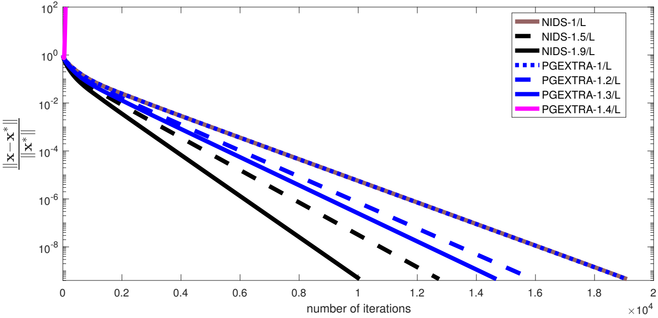
$$\underset{\mathbf{x}}{\text{minimize}} \ f(\mathbf{x}) + g(\mathbf{x}) + \iota_{\mathbf{0}}((\mathbf{I} - \mathbf{W})^{1/2}\mathbf{x})$$

comparing PG-EXTRA and NIDS (Li-Shi-Yan '17)

$$\underset{\mathbf{x}}{\text{minimize}} \ \frac{1}{2}\sum_{i=1}^{n} \|\mathbf{A}_i\mathbf{x}_i - \mathbf{b}_i\|^2 + \mu_1 \sum_{i=1}^{n} \|\mathbf{x}_i\|_1 + \iota_{\mathbf{0}}((\mathbf{I}-\mathbf{W})^{1/2}\mathbf{x})$$

comparing PG-EXTRA and NIDS (Li-Shi-Yan '17)

$$\underset{\mathbf{x}}{\text{minimize}} \;\; \frac{1}{2}\sum_{i=1}^{n}\|\mathbf{A}_i\mathbf{x}_i - \mathbf{b}_i\|^2 + \mu_1\sum_{i=1}^{n}\|\mathbf{x}_i\|_1 + \iota_{\mathbf{0}}((\mathbf{I}-\mathbf{W})^{1/2}\mathbf{x})$$



conclusion

- a new primal-dual algorithm for minimizing the sum of three functions.
- a new interpretation of Chambolle-Pock: Douglas-Rachford splitting on the KKT system under a new metric induced by a block diagonal matrix.
- PAPC is forward-backward splitting applied on the KKT system under the same metric; we proved the optimal bound for the parameters (dual stepsize).
- PD3O is a generalization of both Chambolle-Pock and PAPC, and it has the advantages of both Condat-Vu (a generalization of Chambolle-Pock), and AFBA and PDFP (two generalizations of PAPC).
- In decentralized consensus optimization, we derive a fast method whose stepsize does not depend on the network structure; we provide an optimal bound for the stepsize in PG-EXTRA (Shi et al. '15).

Thank You!

Paper 1 M. Yan, A new primal-dual method for minimizing the sum of three functions with a linear operator, Arxiv: arXiv:1611.09805

Code https://github.com/mingyan08/PD3O

Paper 2 Z. Li, W. Shi and M. Yan, A decentralized proximal-gradient method with network independent step-sizes and separated convergence rates, arXiv:1704.07807

Code https://github.com/mingyan08/NIDS

Paper 3 Z. Li and M. Yan, A primal-dual algorithm with optimal stepsizes and its application in decentralized consensus optimization, arXiv:1711.06785