# Stochastic Primal-Dual Hybrid Gradient Algorithm with Arbitrary Sampling and Imaging Applications

Antonin Chambolle, Matthias J. Ehrhardt, Peter Richtárik, and Carola-Bibiane Schönlieb

Presented by Phalguni Nanda,
Department of Statistics, Purdue University

December 4, 2024

# Problem Statement

PURDUE UNIVERSITY

- Set-up:
  - $\mathbb{X}, \mathbb{Y}_i,\ i = 1, \ldots, n$: real Hilbert spaces.
  - $\mathbb{Y} = \prod_{i=1}^{n} \mathbb{Y}_i := \{y = (y_1, \ldots, y_n) : y_i \in \mathbb{Y}_i,\ i = 1, \ldots, n\}$
  - $\mathrm{A}_i : \mathbb{X} \longrightarrow \mathbb{Y}_i,\ i = 1, \ldots, n$: bounded linear operators
  - $\mathrm{A} : \mathbb{X} \longrightarrow \mathbb{Y}$: $\mathrm{A}x = (\mathrm{A}_1 x, \ldots, \mathrm{A}_n x)$
  - $f : \mathbb{Y} \longrightarrow \mathbb{R}_\infty$ and $g : \mathbb{X} \longrightarrow \mathbb{R}_\infty$ are convex.
  - $f$ is separable: $f(y) = \sum_{i=1}^{n} f_i(y_i)$, $f_i : \mathbb{Y}_i \longrightarrow \mathbb{R}_\infty$
- Objective function:

$$\min_{x \in \mathbb{X}} \left\{ \Phi(x) := \sum_{i=1}^{n} f_i(\mathrm{A}_i x) + g(x) \right\}$$

# Problem Statement

PURDUE UNIVERSITY

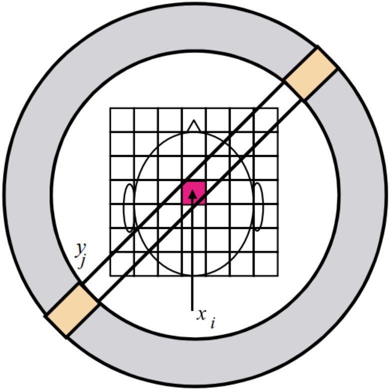
- Set-up:
  - $\mathbb{X}, \mathbb{Y}_i,\ i = 1, \ldots, n$: real Hilbert spaces.
  - $\mathbb{Y} = \prod_{i=1}^{n} \mathbb{Y}_i := \{y = (y_1, \ldots, y_n) : y_i \in \mathbb{Y}_i,\ i = 1, \ldots, n\}$
  - $\mathrm{A}_i : \mathbb{X} \longrightarrow \mathbb{Y}_i,\ i = 1, \ldots, n$: bounded linear operators
  - $\mathrm{A} : \mathbb{X} \longrightarrow \mathbb{Y}$: $\mathrm{A}x = (\mathrm{A}_1 x, \ldots, \mathrm{A}_n x)$
  - $f : \mathbb{Y} \longrightarrow \mathbb{R}_\infty$ and $g : \mathbb{X} \longrightarrow \mathbb{R}_\infty$ are convex.
  - $f$ is separable: $f(y) = \sum_{i=1}^{n} f_i(y_i)$, $f_i : \mathbb{Y}_i \longrightarrow \mathbb{R}_\infty$
- Objective function:

$$\min_{x \in \mathbb{X}} \left\{ \Phi(x) := \sum_{i=1}^{n} f_i(\mathrm{A}_i x) + g(x) \right\}$$

# Positron Emission Tomography

PURDUE UNIVERSITY

Reconstruct image from intensity measurements!




Schematic of PET scanner

- $n$ voxels or grid of boxes imposed over the emitting objects
- $\xi_i$: Number of emissions from the voxel-i.

  $\xi_i \sim \text{Poisson}(x_i)$, are independent.
- $y_j$: Intensity measurement at the $j$-th detector pairs/conincidence line, $j = 1, \ldots, N$.

# Positron Emission Tomography

PURDUE UNIVERSITY

- $C_{i,j} \propto$ probability of the emission from voxel-$i$ being detected by the $j$-th detector pair or hitting the $j$-th coincidence line
- $\Xi_{i,j} = C_{i,j}\xi_i$ number of emissions from voxel-$i$ hitting the $j$-th coincidence line
- $y_j = \sum_{i=1}^{n} \Xi_{i,j}$: measurement at the $j$-th detector pair

To estimate the true emission intensities $x_i$ using the measurements $y_j$

- Maximize the likelihood function:

$$l(x \mid y) \equiv p(y \mid x) = \prod_{j=1}^{N} \frac{1}{y_j!} e^{-\sum_{i=1}^{n} C_{i,j}\xi_i} \left( \sum_{i=1}^{n} C_{i,j}\xi_i \right)^{y_j}$$

# Positron Emission Tomography

PURDUE UNIVERSITY

- $C_{i,j} \propto$ probability of the emission from voxel-$i$ being detected by the $j$-th detector pair or hitting the $j$-th coincidence line
- $\Xi_{i,j} = C_{i,j}\xi_i$ number of emissions from voxel-$i$ hitting the $j$-th coincidence line
- $y_j = \sum_{i=1}^{n} \Xi_{i,j}$: measurement at the $j$-th detector pair

To estimate the true emission intensities $x_i$ using the measurements $y_j$

- Maximize the likelihood function:

$$l(x \mid y) \equiv p(y \mid x) = \prod_{j=1}^{N} \frac{1}{y_j!} e^{-\sum_{i=1}^{n} C_{i,j}\xi_i} \left( \sum_{i=1}^{n} C_{i,j}\xi_i \right)^{y_j}$$

# Positron Emission Tomography

PURDUE UNIVERSITY

- $C_{i,j} \propto$ probability of the emission from voxel-$i$ being detected by the $j$-th detector pair or hitting the $j$-th coincidence line
- $\Xi_{i,j} = C_{i,j}\xi_i$ number of emissions from voxel-$i$ hitting the $j$-th coincidence line
- $y_j = \sum_{i=1}^{n} \Xi_{i,j}$: measurement at the $j$-th detector pair

To estimate the true emission intensities $x_i$ using the measurements $y_j$

- Maximize the likelihood function:

$$l(x \mid y) \equiv p(y \mid x) = \prod_{j=1}^{N} \frac{1}{y_j!} e^{-\sum_{i=1}^{n} C_{i,j}\xi_i} \left( \sum_{i=1}^{n} C_{i,j}\xi_i \right)^{y_j}$$

# An optimization problem

PURDUE UNIVERSITY

- Maximize the log-likelihood function or equivalently the function:

$$\Phi(x) := -Ce_N x + \sum_{j=1}^{N} y_j \log(C^T x)_j$$

- Define

$$\begin{aligned} g : \mathbb{R}^n &\longrightarrow \mathbb{R} \quad \text{as} \quad g(x) := -Ce_N x \\ \mathrm{A}_j : \mathbb{R}^n &\longrightarrow \mathbb{R} \quad \text{as} \quad A_j := C_{.j}^T x = (C^T x)_j \\ f_j : \mathbb{R} &\longrightarrow \mathbb{R} \quad \text{as} \quad f_j(x) := y_j \log\big((C^T x)_j\big) \\ f : \mathbb{R}^N = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{N \text{ times}} &\longrightarrow \mathbb{R} \quad \text{as} \quad f(z) := \sum_{j=1}^{N} y_j \log(z_j) \end{aligned}$$

- So maximizing the likelihood is of the form

$$\min_{x \in \mathbb{X}} \left\{ \Phi(x) := \sum_{i=1}^{n} f_i(\mathrm{A}_i x) + g(x) \right\}$$

# Primal-Dual Set-up

PURDUE UNIVERSITY

- Saddle point problem:

$$\min_{x \in \mathbb{S}} \sup_{y \in \mathbb{Y}} \left\{ \Psi(x,y) := \sum_{i=1}^{n} \langle \mathsf{A}_i x, y_i \rangle - f_i^*(y_i) + g(x) \right\}$$

- Optimality conditions: A saddle point $(x^\sharp, y^\sharp) \in \mathbb{X} \times \mathbb{Y}$ satisfies

$$\mathsf{A}_i x^\sharp \in \partial f_i^*(y^\sharp), \quad i = 1, \ldots, n$$
$$-\mathsf{A}^* y^\sharp \in \partial g(x^\sharp)$$

- Primal-Dual update:

$$x^{(k+1)} = \mathrm{Prox}_g^{\mathsf{T}} \left( x^{(k)} - \mathsf{T}\mathsf{A}^* \bar{y}^{(k)} \right)$$
$$y_i^{(k+1)} = \mathrm{Prox}_{f_i^*}^{\mathsf{S}_i} \left( y_i^{(k)} + \mathsf{S}_i \mathsf{A}_i x^{(k+1)} \right), \ i = 1, \ldots, n,$$
$$\bar{y}^{(k+1)} = y^{(k+1)} + \theta \left( y^{(k+1)} - y^{(k)} \right)$$

- $\mathsf{T}$ and $\mathsf{S} = \mathrm{diag}(\mathsf{S}_1, \ldots, \mathsf{S}_2)$ are symmetric and positive definite.
- Guaranteed convergence under $\|\mathsf{S}^{1/2}\mathsf{A}\mathsf{T}^{1/2}\| < 1$ and $\theta = 1$

# The Stochastic Case

Update a *random* subset of the dual variables at every iteration

- To sample a random subset $\mathbb{S}$ of indices $\{1, \dots, n\} =: [n]$
- $\mathbb{S}$ is a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $2^{[n]}$
- $\mathbb{S}$ induces a probability distribution $P_{\mathbb{S}}$ on $2^{[n]}$ given by

$$P_{\mathbb{S}}(S) := \mathbb{P}\{\mathbb{S} = S\}, \quad S \subseteq [n].$$

- The probability of an integer/index $i$ being included in the random set $\mathbb{S}$ is

$$p_i := \mathbb{P}\{i \in \mathbb{S}\} = \sum_S \mathbb{1}_{\{i \in S\}} \mathbb{P}\{\mathbb{S} = S\}, \quad i = 1, \dots, n.$$

Update a *random* subset of the dual variables at every iteration

- To sample a random subset $\mathbb{S}$ of indices $\{1, \ldots, n\} =: [n]$
- $\mathbb{S}$ is a random variable defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and taking values in $2^{[n]}$
- $\mathbb{S}$ induces a probability distribution $P_{\mathbb{S}}$ on $2^{[n]}$ given by

$$P_{\mathbb{S}}(S) := \mathbb{P}\{\mathbb{S} = S\}, \quad S \subseteq [n].$$

- The probability of an integer/index $i$ being included in the random set $\mathbb{S}$ is

$$p_i := \mathbb{P}\{i \in \mathbb{S}\} = \sum_S \mathbb{1}_{\{i \in S\}} \mathbb{P}\{\mathbb{S} = S\}, \quad i = 1, \ldots, n.$$

# Stochastic PDHG Algorithm

PURDUE UNIVERSITY

**Algorithm 1 Input**: $x^{(0)}, y^{(0)}, \mathsf{S} = \operatorname{diag}(\mathsf{S}_1, \ldots, \mathsf{S}_n), \mathsf{T}, \theta, \mathbb{S}^{(k)}, K$.
**Initialize**: $\bar{y}^{(0)} = y^{(0)}$

1: **for** $k = 0, \ldots, K-1$ **do**
2: $\quad x^{(k+1)} = \operatorname{Prox}_g^{\mathsf{T}}\left(x^{(k)} - \mathsf{T}A^*\bar{y}^{(k)}\right)$
3: $\quad$ Select $\mathbb{S}^{(k+1)} \subset \{1, \ldots, n\}$
4: $\quad y_i^{(k+1)} = \begin{cases} \operatorname{Prox}_{f_i^*}^{\mathsf{S}_i}\left(y_i^{(k)} + \mathsf{S}_i A_i x^{(k+1)}\right) & \text{if } i \in \mathbb{S}^{(k+1)} \\ y_i^{(k)} & \text{else} \end{cases}$
5: $\quad \bar{y}^{(k+1)} = y^{(k+1)} + \theta \mathsf{Q}\left(y^{(k+1)} - y^{(k)}\right)$
6: **end for**

- $\mathsf{Q} := \operatorname{diag}\left(\frac{1}{p_1}\mathsf{I}_1, \ldots, \frac{1}{p_n}\mathsf{I}_n\right)$

# Discrepancy Measures

PURDUE UNIVERSITY

**Bregman distance $D_f^q(x, y)$:**

For any function $f : \mathbb{X} \longrightarrow \mathbb{R}_\infty$, $x, y \in \mathbb{X}$, and a subgradient $q \in \partial f(y)$ of $f$ at $y$,

$$D_f^q(x, y) := f(x) - f(y) - \langle q, x - y \rangle$$

**Distance functions**

$$\mathcal{F}_i(y_i|\tilde{x}, \tilde{y}_i) := f_i^*(y_i) - f_i^*(\tilde{y}_i) - \langle \mathrm{A}_i \tilde{x}, y_i - \tilde{y}_i \rangle, \quad i = 1, \ldots, n.$$

$$\mathcal{F}(y|\tilde{x}, \tilde{y}) := \sum_{i=1}^n \mathcal{F}_i(y_i|\tilde{x}, \tilde{y}_i).$$

$$\mathcal{F}^p(y|\tilde{x}, \tilde{y}) := \sum_{i=1}^n \left( \frac{1}{p_i} - 1 \right) \mathcal{F}_i(y_i|\tilde{x}, \tilde{y}_i).$$

$$\mathcal{G}(x|\tilde{x}, \tilde{y}) := g(x) - g(\tilde{x}) - \langle -\mathrm{A}^* \tilde{y}, x - \tilde{x} \rangle.$$

# Discrepancy Measures

PURDUE UNIVERSITY

## Bregman distance $D_f^q(x,y)$:

For any function $f : \mathbb{X} \longrightarrow \mathbb{R}_\infty$, $x, y \in \mathbb{X}$, and a subgradient $q \in \partial f(y)$ of $f$ at $y$,

$$D_f^q(x,y) := f(x) - f(y) - \langle q, x - y \rangle$$

## Distance functions

$$\mathcal{F}_i(y_i|\tilde{x},\tilde{y}_i) := f_i^*(y_i) - f_i^*(\tilde{y}_i) - \langle \mathrm{A}_i\tilde{x}, y_i - \tilde{y}_i \rangle, \quad i = 1, \ldots, n.$$

$$\mathcal{F}(y|\tilde{x},\tilde{y}) := \sum_{i=1}^n \mathcal{F}_i(y_i|\tilde{x},\tilde{y}_i).$$

$$\mathcal{F}^p(y|\tilde{x},\tilde{y}) := \sum_{i=1}^n \left(\frac{1}{p_i} - 1\right) \mathcal{F}_i(y_i|\tilde{x},\tilde{y}_i).$$

$$\mathcal{G}(x|\tilde{x},\tilde{y}) := g(x) - g(\tilde{x}) - \langle -\mathrm{A}^*\tilde{y}, x - \tilde{x} \rangle.$$

# Partial Primal-Dual Gap

PURDUE UNIVERSITY

**Partial primal-dual gap:**

$$G_{\mathbb{B}_1 \times \mathbb{B}_2}(x, y) := \sup_{\tilde{y} \in \mathbb{B}_2} \Psi(x, \tilde{y}) - \inf_{\tilde{x} \in \mathbb{B}_1} \Psi(\tilde{x}, y), \ \mathbb{B}_1 \times \mathbb{B}_2 \subset \mathbb{X} \times \mathbb{Y}$$

**Relation to the distance functions**

$$G_{\mathbb{B}_1 \times \mathbb{B}_2}(x, y) = \sup_{(\tilde{x}, \tilde{y}) \in \mathbb{B}_1 \times \mathbb{B}_2} [\mathcal{F}(y|\tilde{x}, \tilde{y}) + \mathcal{G}(x|\tilde{x}, \tilde{y})]$$

# The General Convex Case

PURDUE UNIVERSITY

## Theorem (4.3)

*Then the Bregman distance between iterates of Algorithm 1* $(x^{(k)}, y^{(k)}) \in \mathbb{X} \times \mathbb{Y}$ *and any saddle point* $(x^\sharp, y^\sharp) \in \mathbb{X} \times \mathbb{Y}$ *converges to zero almost surely,*

$$D^q_{f^*+g}\left((x^{(k)}, y^{(k)}), (x^\sharp, y^\sharp)\right) \to 0 \quad a.s.$$

*Moreover, the ergodic sequence* $(x_{(K)}, y_{(K)}) := \frac{1}{K}\sum_{k=1}^{K}(x^{(k)}, y^{(k)})$ *converges with rate* $1/K$ *in an expected partial primal-dual gap sense; i.e., for any set* $\mathbb{B} := \mathbb{B}_1 \times \mathbb{B}_2 \subset \mathbb{W}$ *it holds that*

$$\mathbb{E} G_{\mathbb{B}}\left(x_{(K)}, y_{(K)}\right) \leq \frac{C_{\mathbb{B}}}{K},$$

*where the constant is given by*

$$C_{\mathbb{B}} := \sup_{x \in \mathbb{B}_1} \frac{1}{2}\left\|x^{(0)} - x\right\|^2_{\mathrm{T}^{-1}} + \sup_{y \in \mathbb{B}_2} \frac{1}{2}\left\|y^{(0)} - y\right\|^2_{\mathrm{QS}^{-1}} + \sup_{w \in \mathbb{B}} \mathcal{F}^p\left(y^{(0)} | w\right).$$

*The same rate holds for the expected Bregman distance,* $\mathbb{E} D^q_{f^*+g} \leq C_{\{(x^\sharp, y^\sharp)\}}/K$.

# A Deterministic Update

PURDUE UNIVERSITY

## Lemma (4.4)

*Consider the deterministic updates*

$$x^{(k+1)} = \mathrm{Prox}_g^{\mathsf{T}_{(k)}}\left(x^{(k)} - \mathsf{T}_{(k)}\mathsf{A}^*\bar{y}^{(k)}\right),$$

$$\hat{y}_i^{(k+1)} = \mathrm{Prox}_{f_i^*}^{\mathsf{S}_{(k)}^i}\left(y_i^{(k)} + \mathsf{S}_{(k)}^i\mathsf{A}_i x^{(k+1)}\right), \quad i = 1, \dots, n,$$

*with iteration varying step sizes* $\mathsf{T}_{(k)}$ *and* $\mathsf{S}_{(k)} = \mathrm{diag}\left(\mathsf{S}_{(k)}^1, \dots, \mathsf{S}_{(k)}^n\right)$. *Then for any* $(x, y) \in \mathbb{W}$ *it holds that*

$$\left\|x^{(k)} - x\right\|_{\mathsf{T}_{(k)}^{-1}}^2 + \left\|y^{(k)} - y\right\|_{\mathsf{S}_{(k)}^{-1}}^2 \geq \left\|x^{(k+1)} - x\right\|_{\mathsf{T}_{(k)}^{-1}+\mu_g\mathsf{I}}^2 + \left\|\hat{y}^{(k+1)} - y\right\|_{\mathsf{S}_{(k)}^{-1}+\mathsf{M}}^2$$
$$+ 2\left(\mathcal{G}\left(x^{(k+1)}|w\right) + \mathcal{F}\left(\hat{y}^{(k+1)}|w\right)\right) - 2\left\langle \mathsf{A}\left(x^{(k+1)} - x\right), \hat{y}^{(k+1)} - \bar{y}^{(k)}\right\rangle$$
$$+ \left\|x^{(k+1)} - x^{(k)}\right\|_{\mathsf{T}_{(k)}^{-1}}^2 + \left\|\hat{y}^{(k+1)} - y^{(k)}\right\|_{\mathsf{S}_{(k)}^{-1}}^2$$

# The Stochastic Set-up

PURDUE UNIVERSITY

- $y_i^{(k)}$ at step-$k$ is updated only if $i$ gets picked in the random subset $\mathbb{S}$, i.e.,

$$y_i^{(k+1)} = \begin{cases} \hat{y}_i^{(k+1)}, & \text{if } i \in \mathbb{S} \\ y_i^{(k)}, & \text{if } i \notin \mathbb{S} \end{cases} .$$

- A key initial step:

$$\begin{aligned} \mathbb{E}^{(k+1)}\left[\varphi(y_i^{(k+1)})\right] &= \mathbb{E}_{\mathbb{S}^{(k)}}\left[\varphi(y_i^{(k+1)})\right] \\ &= \varphi(\hat{y}_i^{(k+1)})\mathbb{P}\{i \in \mathbb{S}\} + \varphi(y_i^{(k)})\mathbb{P}\{i \notin \mathbb{S}\} \\ \Rightarrow \varphi(\hat{y}_i^{(k+1)}) &= \left(\frac{1}{p_i} - 1\right)\mathbb{E}^{(k+1)}\left[\varphi(y_i^{(k+1)})\right] \\ &\quad - \left(\frac{1}{p_i} - 1\right)\varphi(y_i^{(k)}) + \mathbb{E}^{(k+1)}\left[\varphi(y_i^{(k+1)})\right]. \end{aligned}$$

- Will need to bound terms involving inner products

# Technical Requirements

PURDUE UNIVERSITY

## Expected Separable Overapproximation (ESO)

For a bounded linear operator $\mathrm{C} : \mathbb{X} \longrightarrow \mathbb{Y}$, a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ satisfies the ESO inequality if

$$\mathbb{E}_{\mathbb{S}} \left\| \sum_{i \in \mathbb{S}} z_i \mathrm{C}_i^* \right\|^2 \leq \sum_{i=1}^{n} p_i v_i \|z_i\|^2, \quad \forall z = (z_1, \ldots, z_n) \in \mathbb{Y}.,$$

The parameters $v_i$, $i = 1, \ldots, n$ are called as ESO parameters of C and $\mathbb{S}$.

## Lemma (4.2)

*Let* $y_i^+ = \begin{cases} \hat{y}_i, & \text{if } i \in \mathbb{S} \\ y_i, & \text{otherwise} \end{cases}$ *and let* $\{v_i\}_{i=1}^n$ *be some ESO parameters of* $\mathrm{S}^{1/2}\mathrm{A}\mathrm{T}^{1/2}$. *Then for any* $x \in \mathbb{X}$ *and* $c > 0$,

$$2\mathbb{E}_{\mathbb{S}} \langle \mathrm{QA}x, y^+ - y \rangle \geq -\mathbb{E}_{\mathbb{S}} \left\{ \frac{1}{c} \|x\|^2_{\mathrm{T}^{-1}} + c \max_{1 \leq i \leq n} \frac{v_i}{p_i} \|y^+ - y\|^2_{\mathrm{QS}^{-1}} \right\}$$

# Technical Requirements

PURDUE UNIVERSITY

## Expected Separable Overapproximation (ESO)

For a bounded linear operator $C : \mathbb{X} \longrightarrow \mathbb{Y}$, a vector $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ satisfies the ESO inequality if

$$\mathbb{E}_{\mathbb{S}} \left\| \sum_{i \in \mathbb{S}} z_i C_i^* \right\|^2 \leq \sum_{i=1}^{n} p_i v_i \|z_i\|^2, \quad \forall z = (z_1, \ldots, z_n) \in \mathbb{Y}.,$$

The parameters $v_i$, $i = 1, \ldots, n$ are called as ESO parameters of C and $\mathbb{S}$.

## Lemma (4.2)

*Let* $y_i^+ = \begin{cases} \hat{y}_i, & \text{if } i \in \mathbb{S} \\ y_i, & \text{otherwise} \end{cases}$ *and let* $\{v_i\}_{i=1}^n$ *be some ESO parameters of* $S^{1/2}AT^{1/2}$. *Then for any* $x \in \mathbb{X}$ *and* $c > 0$,

$$2\mathbb{E}_{\mathbb{S}} \langle QAx, y^+ - y \rangle \geq -\mathbb{E}_{\mathbb{S}} \left\{ \frac{1}{c} \|x\|^2_{T^{-1}} + c \max_{1 \leq i \leq n} \frac{v_i}{p_i} \|y^+ - y\|^2_{QS^{-1}} \right\}$$

# Semi-strongly Convex Case

PURDUE UNIVERSITY

**Algorithm 2** SPDHG with acceleration on the dual variable (DA-SPDHG)

**Input**: $x^{(0)}, y^{(0)}, \tau_{(0)} \in \mathbb{R}, \tilde{\sigma}_{(0)} \in \mathbb{R}, \mathbb{S}^{(k)}, K$.

**Initialize**: $\bar{y}^{(0)} = y^{(0)}$.

1: **for** $k = 0, \ldots, K-1$ **do**

2: $x^{(k+1)} = \mathrm{Prox}_g^{\tau_{(k)}} \left( x^k - \tau_{(k)} \mathrm{A}^* \bar{y}^{(k)} \right)$

3: Select $\mathbb{S}^{(k+1)} \subset \{1, \ldots, n\}$

4: $\sigma_i^{(k)} = \dfrac{\tilde{\sigma}_{(k)}}{\mu_i \left[ p_i - 2(1 - p_i) \tilde{\sigma}_{(k)} \right]}, \quad i \in \mathbb{S}^{(k+1)}$

5: $y_i^{(k+1)} = \begin{cases} \mathrm{Prox}_{f_i^*}^{\sigma_i^{(k)}} \left( y_i^{(k)} + \sigma_i^{(k)} \mathrm{A}_i x^{(k+1)} \right) & \text{if } i \in \mathbb{S}^{(k+1)} \\ y_i^{(k)} & \text{else} \end{cases}$

6: $\theta_{(k)} = \left(1 + 2\tilde{\sigma}_{(k)}\right)^{-1/2}, \quad \tau_{(k+1)} = \tau_{(k)}/\theta_{(k)}, \quad \tilde{\sigma}_{(k+1)} = \theta_{(k)} \tilde{\sigma}_{(k)}$

7: $\bar{y}^{(k+1)} = y^{(k+1)} + \theta_{(k)} \mathrm{Q} \left( y^{(k+1)} - y^{(k)} \right)$

8: **end for**

# Semi-strongly Convex Case

PURDUE UNIVERSITY

## Theorem (5.1)

*Let* $f_i^*$ *be strongly convex with constants* $\mu_i > 0, i = 1, \ldots, n$. *Let the initial step sizes* $\tilde{\sigma}_{(0)}, \tau_{(0)}$ *be chosen such that*

$$\tilde{\sigma}_{(0)} < \min_i \tfrac{p_i}{2(1-p_i)},$$

*and for the ESO parameters* $\{v_i\}$ *of* $\mathrm{S}_{(0)}^{1/2}\mathrm{A}\tau_{(0)}^{1/2}$ *it holds that*

$$v_i \leq p_i, \quad i = 1, \ldots, n$$

*with* $\left[\mathrm{S}_{(k)}\right]_i = \sigma_i^{(k)}\mathrm{I}$ *and*

$$\sigma_i^{(k)} = \frac{\tilde{\sigma}_{(k)}}{\mu_i\left[p_i - 2\left(1 - p_i\right)\tilde{\sigma}_{(k)}\right]}.$$

*Then there exists* $\tilde{K} \in \mathbb{N}$ *such that for all* $K \geq \tilde{K}$ *it holds that*

$$\mathbb{E}\left\|y^{(K)} - y^\sharp\right\|_{\mathrm{Y}_{(0)}}^2 \leq \frac{2}{K^2}\left[\left\|x^{(0)} - x^\sharp\right\|_{\tau_{(0)}^{-1}}^2 + \left\|y^{(0)} - y^\sharp\right\|_{\mathrm{Y}_{(0)}}^2\right],$$

*where the metric on* $\mathbb{Y}$ *is defined by* $\mathrm{Y}_{(k)} := \mathrm{Q}\mathrm{S}_{(k)}^{-1} + 2\mathrm{M}(\mathrm{Q} - \mathrm{I})$.

# Strongly Convex Case

PURDUE UNIVERSITY

## Theorem (6.1)

*Let the step sizes* $\tau, \sigma_1, \ldots, \sigma_n, 0 < \theta < 1$ *be chosen such that the ESO parameters* $\{v_i\}$ *of* $\mathrm{S}^{1/2}\mathrm{A}\tau^{1/2}$ *can be estimated as*

$$v_i < p_i/\theta, \quad i = 1, \ldots, n$$

*and the extrapolation* $\theta$ *satisfies the lower bounds*

$$\theta \geq \frac{1}{1 + 2\mu_g \tau}, \quad \theta \geq \frac{1 + 2\left(1 - p_i\right)\mu_i \sigma_i}{1 + 2\mu_i \sigma_i}, \quad i = 1, \ldots, n$$

*Then the iterates of Algorithm 1 converge linearly to the saddle point; in particular*

$$\mathbb{E}\left\{\left(1 - \gamma^2\theta\right)\left\|x^{(K)} - x^\sharp\right\|_{\mathrm{X}}^2 + \left\|y^{(K)} - y^\sharp\right\|_{\mathrm{Y}}^2\right\} \leq \theta^K\left\{\left\|x^{(0)} - x^\sharp\right\|_{\mathrm{X}}^2 + \left\|y^{(0)} - y^\sharp\right\|_{\mathrm{Y}}^2\right\}$$

*holds where the metrics are given by* $\mathrm{X} := \left(\tau^{-1} + 2\mu_g\right)\mathrm{I}, \mathrm{Y} := \left(\mathrm{S}^{-1} + 2\mathrm{M}\right)\mathrm{Q}$, *and* $\gamma^2 = \max_i v_i/p_i$.

# Back to PET Reconstruction

PURDUE UNIVERSITY

- Space of tracer distributions or images: $\mathbb{X} = \mathbb{R}^{d_1 \times d_2}$, $d_1 = d_2 = 250$.
- Data spaces: $\mathbb{Y}_i = \mathbb{R}^{|\mathbb{B}_i|}$
  - $\mathbb{B}_i \subset [200 \times 250]$, $\mathbb{B}_i \cap \mathbb{B}_j = \emptyset$, $\bigcup_{i=1}^{n} \mathbb{B}_i = [200 \times 250]$
- $$f_i(y) = \begin{cases} \sum_{j \in \mathbb{B}_i} \left[ y_j + r_j - b_j + b_j \log\left(\frac{b_j}{y_j + r_j}\right)\right], & \text{if } y_j + r_j > 0, \\ +\infty, & \text{else.} \end{cases}$$
- Operator A is a scaled X-ray transform where in each of 200 directions 250 line integrals are computed.
- TV prior: $g(x) = \alpha \|\nabla x\|_{2,1} + \iota_{\geq 0}(x)$
  - $\alpha = 0.2$, $\nabla x = (\nabla_1 x, \nabla_2 x) \in \mathbb{R}^{d_1 d_2 \times 2}$.

# PET Reconstruction

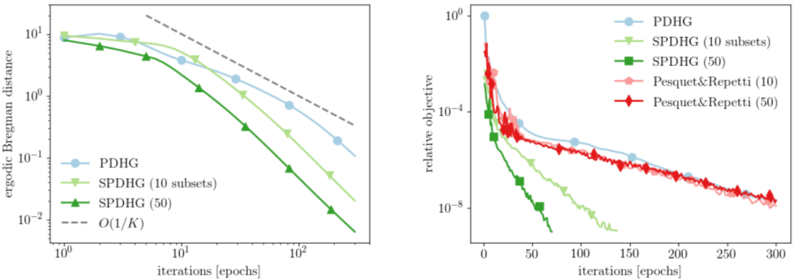



FIG. 1. *PET reconstruction with TV solved as a nonstrongly convex problem.* Left: *As proven in Theorem* 4.3, *the ergodic Bregman distances converge indeed with rate* $O(1/K)$. Right: *Speed comparison measured in terms of relative objective* $[\Phi(x^{(K)}) - \Phi(x^\sharp)]/[\Phi(x^{(0)}) - \Phi(x^\sharp)]$. *The proposed algorithm SPDHG converges faster than the algorithm of Pesquet and Repetti* [35] *and the deterministic PDHG.*

# PET Reconstruction

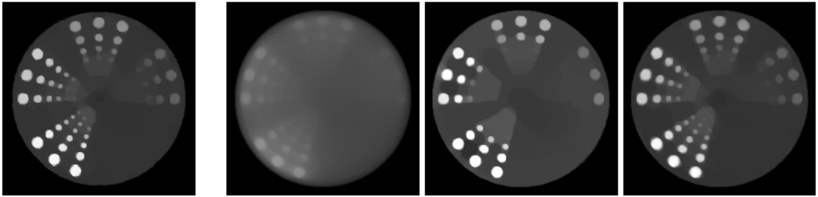


FIG. 2. *PET reconstruction results after* 5 *epochs with uniform sampling of* 50 *subsets.* From left to right: *Approximate primal part of saddle point, PDHG, Pesquet and Repetti* [35], *and SPDHG. With the same number of operator evaluations, both stochastic algorithms make much more progress towards the saddle point.*