Faster PET Reconstruction with a Stochastic Primal-Dual Hybrid Gradient Method

Peter Richtárik

KAUST, The University of Edinburgh

Joint work with A. Chambolle (École Polytechnique), M. J. Ehrhardt (Cambridge) and C.-B. Schönlieb (Cambridge)

April 10, 2017

King Abdullah University of Science and Technology

peter.richtarik@kaust.edu.sa

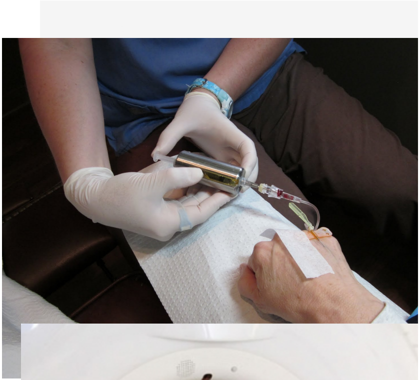
Motivation

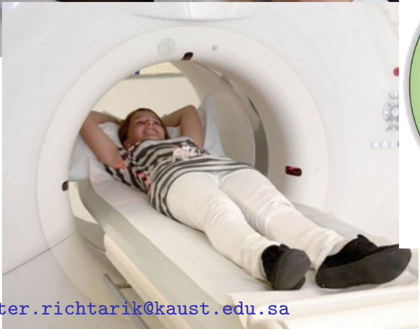
Positron Emission Tomography (PET)

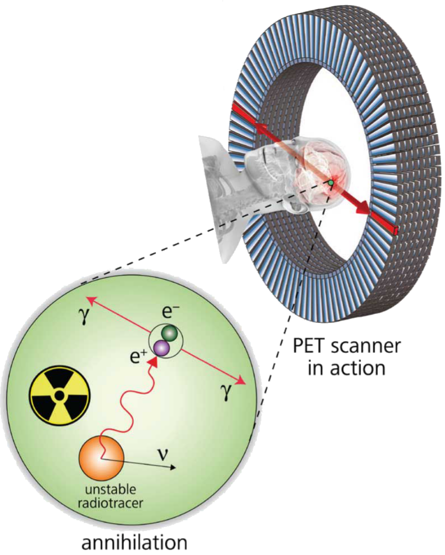
Positron Emission Tomography (PET)

Positron Emission Tomography (PET)

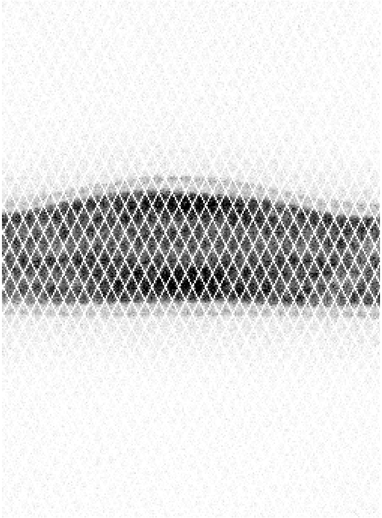
Positron Emission Tomography (PET)

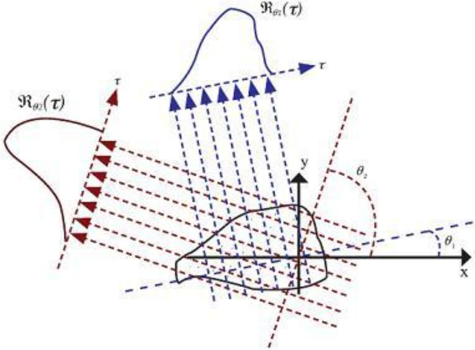






Simple Model for PET Operator





$$(\mathbf{X}x)(s, s^{\perp}) = \int_{\mathbb{R}} x(s + ts^{\perp})dt \qquad \text{(X-ray transform)}$$

Advanced Model for PET Operator

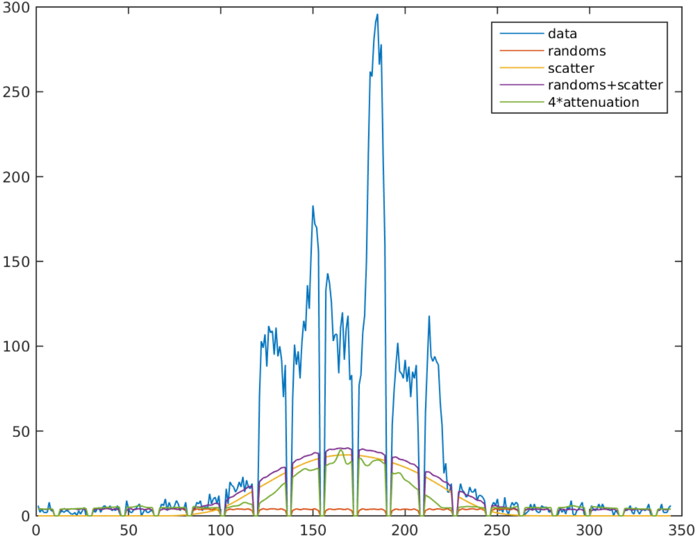
$$\mathbf{A} \; : \; x \mapsto \mathbf{MNX}(x * k)$$

- resolution modelling
- X-ray transform $\mathbf{X}$
- multiplicative correction $\mathbf{N}$ (attenuation and normalization)
- dead detector modelling (subsampling) $\mathbf{M}$
- $\mathbf{A} \geq 0$ ($\mathbf{A}_{ij} \geq 0$). Therefore: $x \geq 0 \Rightarrow \mathbf{A}x \geq 0$

$$d \sim \text{Poisson}(\mathbf{A}x + r)$$

background $r \geq 0$: scatter, randoms

Correction Factors



Maximum A-Posteriori Estimation

Image Reconstruction by MAP Estimation

Data $\{d_i\}_{i=1}^N$ is independent Poisson with mean $\lambda_i = (\mathbf{A}x + r)_i$:

$$d_i \sim \mathrm{Poisson}(\lambda_i)$$

- Want to estimate the unknown parameters λ_i
- This is "equivalent to" estimating the unknown image x

MAP: Maximum A-Posteriori Estimation of Image x

$$x_{MAP} \in \arg\max_x \underbrace{p(d_1, \ldots, d_N \mid x)}_{\text{likelihood}} \times \underbrace{\psi(x)}_{\text{prior}}$$

PET Data Fidelity: Poisson likelihood

Poisson likelihood

$$
\begin{aligned}
p(d_1, \ldots, d_N \mid x) &:= \prod_{i=1}^{N} \lambda_i^{d_i} \exp(-\lambda_i)/d_i! \\
&= \prod_{i=1}^{N} \exp\left\{-KL(d_i, \lambda_i) - \xi(d_i)\right\},
\end{aligned}
$$

- $KL(d_i, \lambda_i) := d_i \log(d_i/\lambda_i) + \lambda_i - d_i$
  generalized relative entropy; generalized KL divergence
- $\xi(d_i) := d_i \log d_i - d_i - \log(d_i!)$

MAP Estimation via Minimization of KL Divergence

$$
\begin{aligned}
&\arg\max_x p(d_1, \dots, d_N \mid x) \times \psi(x) \\
&= \arg\max_x \log p(d_1, \dots, d_N \mid x) + \log(\psi(x)) \\
&= \arg\min_x \left(\sum_{i=1}^N KL(d_i, \lambda_i(x)) + \xi(d_i) \right) \underbrace{- \log(\psi(x))}_{g(x)}
\end{aligned}
$$

Partition the Sum into $n \ll N$ Blocks

$$\min_x \left(\sum_{i=1}^{n} \underbrace{\sum_{j \in B_i} KL(d_j, \lambda_j(x)) + \xi(d_j)}_{f_i(\mathbf{A}_i x)} \right) + g(x)$$

MAP Reconstruction

$$\min_x \left\{ \sum_{i=1}^{n} f_i(\mathbf{A}_i x) + g(x) \right\}$$

Assumptions on f_i and g

Block Loss/Fidelity Functions

$$f_i \ : \ \mathbb{Y}_i \mapsto \mathbb{R} \cup \{+\infty\}, \quad i = 1, 2, \ldots, n$$

Regularizer

$$g \ : \ \mathbb{X} \mapsto \mathbb{R} \cup \{+\infty\}$$

Examples: total variation Rudin, Osher, Fatemi 1992, total generalized variation Bredies, Kunisch, Pock 2010

Assumption

- Functions $f_1, \ldots, f_n$ and g are proper, convex, closed

Consequence:

$$f_i(y_i) = f_i^{**}(y_i) := \max_{z_i \in \mathbb{Y}_i} \langle y_i, z_i \rangle - f_i^*(z_i)$$

Reformulation into a Saddle Point Optimization Problem

MAP Reconstruction

$$\text{Find} \quad x^* \in \arg\min_{x \in \mathbb{X}} \left\{ \sum_{i=1}^{n} f_i(\mathbf{A}_i x) + g(x) \right\}$$

Saddle Point Problem

$$\min_{x \in \mathbb{X}} \max_{y \in \mathbb{Y}} \left\{ \sum_{i=1}^{n} \langle \mathbf{A}_i x, y_i \rangle - f_i^*(y_i) + g(x) \right\}$$

- regularizer can be dualized as well, i.e., part of f

Algorithm

Primal-Dual Hybrid Gradient (PDHG) Algorithm*

PDHG (aka Chambolle-Pock) Algorithm

- initial iterates: $x^0 \in \mathbb{X}$, $y^0 \in \mathbb{Y}$, $\overline{y}^0 = y^0$
- step sizes: $\mathbf{T} \in \mathbb{R}^{|\mathbb{X}| \times |\mathbb{X}|}$, $\mathbf{S}_i \in \mathbb{R}^{|\mathbb{Y}_i| \times |\mathbb{Y}_i|}$, $\theta > 0$

Iterate:

- $x^{k+1} = \mathrm{prox}_g^{\mathbf{T}}(x^k - \mathbf{T}\mathbf{A}^* \overline{y}^k)$
- $y_i^{k+1} = \mathrm{prox}_{f_i^*}^{\mathbf{S}_i}(y_i^k + \mathbf{S}_i \mathbf{A}_i x^{k+1}), \quad i = 1, \ldots, n$
- $\overline{y}^{k+1} = y^{k+1} + \theta(y^{k+1} - y^k)$

- $\mathrm{prox}_g^{\mathbf{M}}(z) := \arg\min_x \left\{ \frac{1}{2} \|x - z\|^2_{\mathbf{M}^{-1}} + g(x) \right\}$
- $\|x\|^2_{\mathbf{M}^{-1}} := \langle \mathbf{M}^{-1} x, x \rangle$
- Evaluation of $\mathbf{A}_i$ and $\mathbf{A}_i^*$ for all $i = 1, \ldots, n$.

*Pock, Cremers, Bischof, Chambolle 2009, Chambolle and Pock 2011, Pock and Chambolle 2011

Stochastic PDHG Algorithm*

SPDHG Algorithm

- initial iterates: $x^0 \in \mathbb{X}, \quad y^0 \in \mathbb{Y}, \quad \overline{y}^0 = y^0$
- step sizes: $\mathbf{T} \in \mathbb{R}^{|\mathbb{X}| \times |\mathbb{X}|}, \quad \mathbf{S}_i \in \mathbb{R}^{|\mathbb{Y}_i| \times |\mathbb{Y}_i|}, \quad \theta > 0$

Iterate:

- $x^{k+1} = \mathrm{prox}_g^{\mathbf{T}}(x^k - \mathbf{T}\mathbf{A}^* \overline{y}^k)$
- Select randomly $\hat{S} \subseteq \{1, \ldots, n\}$
- $y_i^{k+1} = \begin{cases} \mathrm{prox}_{f_i^*}^{\mathbf{S}_i}(y_i^k + \mathbf{S}_i \mathbf{A}_i x^{k+1}) & \text{if } i \in \hat{S} \\ y_i^k & \text{otherwise} \end{cases}$
- $\overline{y}^{k+1} = y^{k+1} + \theta \mathbf{P}^{-1}(y^{k+1} - y^k)$

- matrix of probabilities $\mathbf{P} := \mathrm{Diag}(p_1, \cdots, p_n)$, $p_i := \mathbb{P}(i \in \hat{S})$
- Evaluation of $\mathbf{A}_i$ and $\mathbf{A}_i^*$ only for $i \in \hat{S}$.

*generalizes Pock and Chambolle 2011 and Zhang and Xiao 2015

Convergence

ESO Parameters and Inequality*

Definition (Expected Separable Overapproximation (ESO))

Let $\hat{S} \subseteq \{1, \ldots, n\}$ be any sampling, with $p_i := \mathbb{P}(i \in \hat{S})$. Let $\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_n \in \mathbb{R}^{|\mathbb{Y}_i| \times |\mathbb{X}|}$. We say that scalars $v_1, \ldots, v_n$ **(ESO parameters)** fulfil the **ESO inequality** if

$$\mathbb{E}_{\hat{S}} \left\| \sum_{i \in \hat{S}} \mathbf{C}_i^* y_i \right\|^2 \leq \sum_{i=1}^n p_i v_i \|y_i\|^2, \quad \text{for all} \quad y_1 \in \mathbb{Y}_1, \ldots, y_n \in \mathbb{Y}_n.$$

Example (Full Sampling: $\hat{S} = \{1, \ldots, n\}$ with probability 1)

$1 \| \sum_{i=1}^n \mathbf{C}_i^* y_i \|^2 \leq \sum_{i=1}^n 1 v_i \|y_i\|^2$. Can choose: $v_i = \|\mathbf{C}^*\|^2$

Example (Serial Sampling: $\hat{S} = \{i\}$ with probability p_i)

$\sum_{i=1}^n p_i \|\mathbf{C}_i^* y_i\|^2 \leq \sum_{i=1}^n p_i v_i \|y_i\|^2$. Can choose: $v_i = \|\mathbf{C}_i^*\|^2$

*Richtárik, Takáč 2011, Qu, Richtárik, Zhang 2014

Inequality with ESO Parameters

Lemma (Estimating Inner Products)

Let y^k be generated by SPDHG and $\gamma^2 \geq \max_i v_i$, where $v_1, \ldots, v_n$ are ESO parameters and $\mathbf{C}_i = p_i^{-1/2} \mathbf{S}_i^{1/2} \mathbf{A}_i \mathbf{T}^{1/2}$. Then for any $x \in \mathbb{X}$

$$\mathbb{E}^k \langle \mathbf{P}^{-1} \mathbf{A} x, y^k - y^{k-1} \rangle \geq -\frac{\gamma}{2} \mathbb{E}^k \left\{ \|x\|^2_{\mathbf{T}^{-1}} + \|y^k - y^{k-1}\|^2_{(\mathbf{S}\mathbf{P})^{-1}} \right\}.$$

Example (Full Sampling: $\hat{S} = \{1, \ldots, n\}$)

$$\|\mathbf{S}^{1/2} \mathbf{A} \mathbf{T}^{1/2}\|^2 \leq \gamma^2, \qquad \sigma \tau \|\mathbf{A}\|^2 \leq \gamma^2$$

Example (Serial Sampling: $\hat{S} = \{i\}$)

$$\frac{\|\mathbf{S}_i^{1/2} \mathbf{A}_i \mathbf{T}^{1/2}\|^2}{p_i} \leq \gamma^2, \qquad \frac{\sigma_i \tau \|\mathbf{A}_i\|^2}{p_i} \leq \gamma^2, \qquad i = 1, \ldots, n$$

Convergence: General Theorem

$$\mathcal{E}(x,y) := \frac{1}{2}\|x - x^*\|^2_{\mathbf{T}^{-1}} + \frac{1}{2}\|y - y^*\|^2_{(\mathbf{SP})^{-1}} + \sum_{i=1}^{n}(p_i^{-1} - 1)D_{f_i^*}^{q_i^*}(y_i, y_i^*)$$

Theorem (Convergence of SPDHG)

Assume a saddle point exists. Let (x^*, y^*) be any saddle point, $p^* := -\mathbf{A}^* y^* \in \partial g(x^*)$, $q^* := \mathbf{A}x^* \in \partial f(y^*)$. Choose $\mathbf{S}, \mathbf{T}$ such that $0 < \gamma^2 < 1$ upper bounds ESO parameters, $\theta = 1$. Then

- (x^k, y^k) *is bounded in the sense that* $\mathbb{E}\mathcal{E}(x^k, y^k) \leq \frac{\mathcal{E}(x^0,y^0)}{1-\gamma}$.
- $\|x^{k+1} - x^k\| \to 0, \quad \|y^{k+1} - y^k\| \to 0$ *almost surely*
- $D_g^{p^*}(x^k, x^*) \to 0, \quad D_{f^*}^{q^*}(y^k, y^*) \to 0$ *almost surely*
- *ergodic sequence* $(x_K, y_K) := \frac{1}{K}\sum_{k=1}^{K}(x^k, y^k)$.

$$\mathbb{E}D_g^{p^*}(x_K, x^*) + \mathbb{E}D_{f^*}^{q^*}(y_K, y^*) \leq \frac{\mathcal{E}(x^0, y^0)}{K}$$

- Bregman distance: $D_g^{p^*}(x, x^*) := g(x) - g(x^*) - \langle p^*, x - x^* \rangle$
- Deterministic setting: convergence in norm to a saddle point

Dual Accelerated SPDHG

DA-SPDHG Algorithm

- initial iterates: $x^0 \in \mathbb{X}, \quad y^0 \in \mathbb{Y}, \quad \overline{y}^0 = y^0$
- step sizes: $\mathbf{T}_0 \in \mathbb{R}^{|\mathbb{X}| \times |\mathbb{X}|}, \quad \tilde{\sigma}_0 > 0$

Iterate:

- $x^{k+1} = \mathrm{prox}_g^{\mathbf{T}_k}(x^k - \mathbf{T}_k \mathbf{A}^* \overline{y}^k)$
- Select a random subset $\hat{S} \subseteq \{1, \dots, n\}$
- $\sigma_i^k = \frac{\tilde{\sigma}_k}{\mu_i [p_i - 2(1-p_i)\tilde{\sigma}_k]}, \quad i \in \hat{S}$
- $y_i^{k+1} = \begin{cases} \mathrm{prox}_{f_i^*}^{\sigma_i^k}(y_i^k + \sigma_i^k \mathbf{A}_i x^{k+1}) & \text{if } i \in \hat{S} \\ y_i^k & \text{otherwise} \end{cases}$
- $\theta_k = (1 + 2\tilde{\sigma}_k)^{-1/2}, \ \mathbf{T}_{k+1} = \mathbf{T}_k / \theta_k, \quad \tilde{\sigma}_{k+1} = \theta_k \tilde{\sigma}_k$
- $\overline{y}^{k+1} = y^{k+1} + \theta_k \mathbf{P}^{-1}(y^{k+1} - y^k)$

Dual Accelerated SPDHG

Theorem (Convergence of DA-SPDHG)

Let (x^*, y^*) be a saddle point and assume f_i are $\mu_i > 0$ strongly convex for $i = 1, \dots, n$. Choose $\tilde{\sigma}_0, \mathbf{T}_0$ such that $0 < \gamma^2 \le 1$ upper bounds ESO parameters and $\tilde{\sigma}_0 < \min_i \frac{p_i}{2(1-p_i)}$. Let $\mathbf{Y}_k := (\mathbf{S}_k \mathbf{P})^{-1} + 2\mathbf{M}_f(\mathbf{P}^{-1} - I)$, $\mathbf{M}_f = \mathrm{Diag}(\mu_1, \dots, \mu_n)$. Then there exists $K_0 \in \mathbb{N}$, $C > 0$ such that for all $K \ge K_0$

$$\mathbb{E}\|y^K - y^*\|^2_{\mathbf{Y}_0} \le \frac{C}{K^2}\left\{\|x^0 - x^*\|^2_{\mathbf{T}_0^{-1}} + \|y^0 - y^*\|^2_{\mathbf{Y}_0}\right\}.$$

- For serial sampling: $\tilde{\sigma}_0 \le \min_i \frac{\gamma^2 \mu_i p_i^2}{\|\mathbf{A}_i \mathbf{T}_0^{1/2}\|^2 + 2\gamma^2 \mu_i p_i (1-p_i)}$

Linear Convergence

Theorem (Linear Convergence in the Strongly Convex Case)

Let (x^*, y^*) be a saddle point and g, f_i are $\mu_g, \mu_i > 0$ strongly convex for $i = 1, \dots, n$. Choose $\mathbf{S}, \mathbf{T}, \theta \in (0,1)$ such that $\gamma^2 \leq 1$ upper bounds ESO parameters and

$$\theta(\mathbf{I} + 2\mu_g \mathbf{T}) \geq \mathbf{I}$$
$$\theta(\mathbf{I} + 2\mu_i \mathbf{S}_i) \geq \mathbf{I} + 2(1 - p_i)\mu_i \mathbf{S}_i, \quad i = 1, \dots, n$$

in a positive semidefinite sense for matrices. Let

$$\mathbf{X} := \mathbf{T}^{-1} + 2\mu_g \mathbf{I}, \ \mathbf{Y} := (\mathbf{S}^{-1} + 2\mathbf{M}_f)\mathbf{P}^{-1}, \ \mathbf{M}_f = \mathrm{Diag}(\mu_1, \dots, \mu_n).$$

Then the iterates of SPDHG satisfy

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

where the constant is $C := \|x^0 - x^*\|^2_{\mathbf{X}} + \|y^0 - y^*\|^2_{\mathbf{Y}}$.

Parameters for Serial Sampling $\hat{S} = \{i\}$

- scalar parameters: $\mathbf{T} = \tau \mathbf{I}$, $\mathbf{S}_i = \sigma_i \mathbf{I}$
- condition number: $\kappa_i := \frac{\|\mathbf{A}_i\|^2}{\mu_i \mu_g}$

Example (Uniform Sampling: $p_i = 1/n$)

$$\sigma_i = \frac{\gamma}{\kappa_i^{1/2} \mu_i}, \tau = \frac{\gamma}{n \max_j \kappa_j^{1/2} \mu_g}, \quad \theta = 1 - \left(n + \frac{n \max_j \kappa_j^{1/2}}{2\gamma} \right)^{-1}$$

Example (Importance Sampling: $p_i = \frac{\kappa_i^{1/2}}{\sum_{j=1}^n \kappa_j^{1/2}}$)

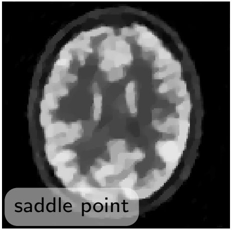
$$\sigma_i = \frac{\gamma}{\kappa_i^{1/2} \mu_i}, \tau = \frac{\gamma}{\sum_{j=1}^n \kappa_j^{1/2} \mu_g}, \quad \theta = 1 - \left(\frac{\sum_j \kappa_j^{1/2}}{\max_j \kappa_j^{1/2}} + \frac{\sum_j \kappa_j^{1/2}}{2\gamma} \right)^{-1}$$

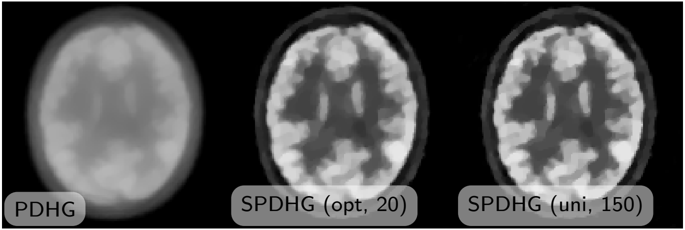
Numerical Results

PET reconstruction, linear rate

$$x^* \in \arg\min_{x \geq 0} \left\{ \sum_{i=1}^{n} \tilde{\mathrm{KL}}(d_i, \mathbf{A}_i x + r_i) + \alpha \|\nabla x\|_{1,2} + \frac{\mu_g}{2}\|x\|^2 \right\}$$

- 5 epochs





PET reconstruction, linear rate

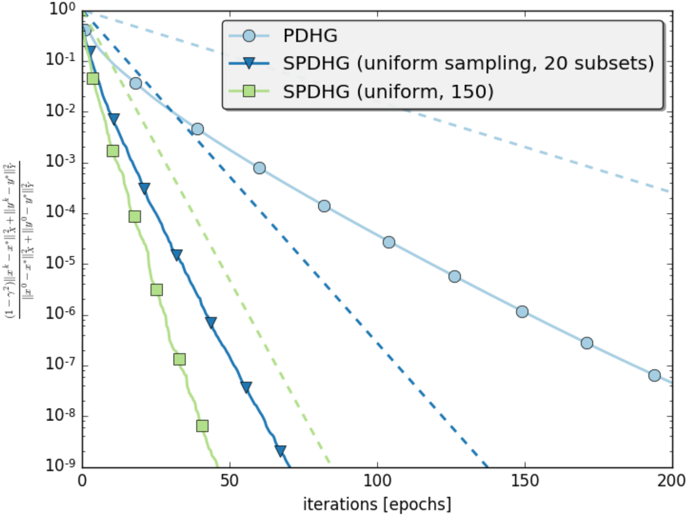


Figure: Distance to the saddle point

PET reconstruction, linear rate

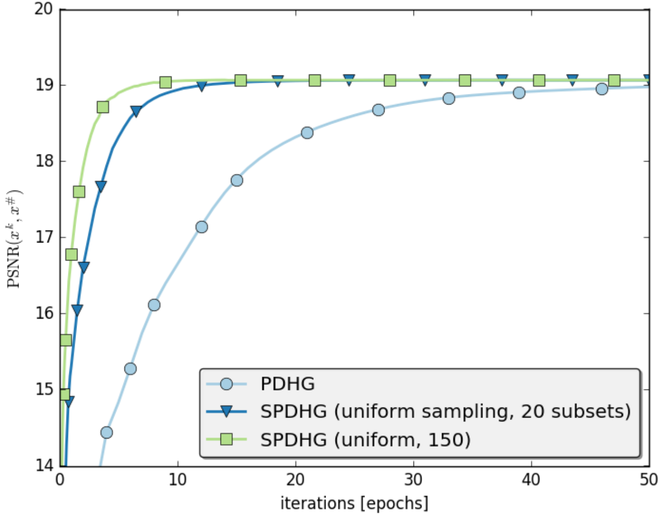


Figure: Peak signal-to-noise ratio

PET reconstruction, linear rate

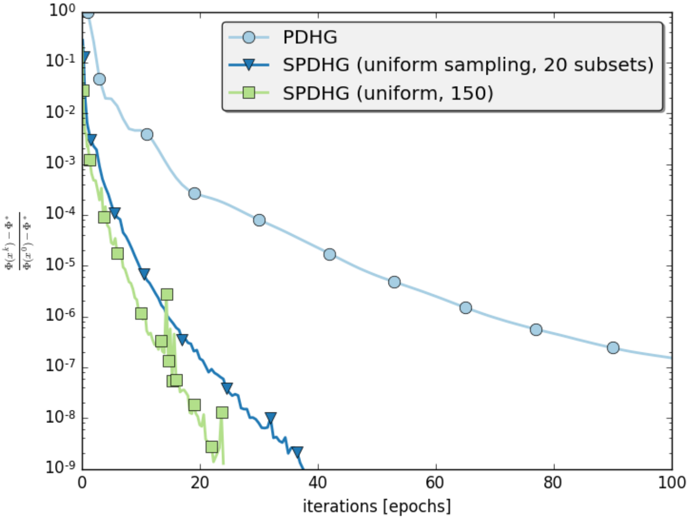


Figure: Objective function value

Conclusions and Outlook

Conclusions:

- Stochastic optimization for separable cost functionals
- Stochastic generalization of PDHG of Chambolle and Pock: non-smooth, acceleration, linear convergence
- Application to PET: incredible speed-up!

Outlook:

- Theory: Other $1/k^2$ acceleration techniques
- Application: real 3D PET data, block selection, non-uniform sampling
- Other applications: CT, MRI

Extra Material: More Experimental Results

PET reconstruction with TV, $1/k$ rate

$$x^* \in \arg\min_{x \geq 0} \left\{ \sum_{i=1}^{n} \mathrm{KL}(d_i, \mathbf{A}_i x + r_i) + \alpha \|\nabla x\|_{1,2} \right\}$$

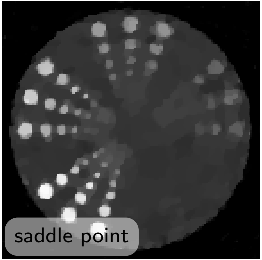
- Proximal operator for TV with non-negativity constraint approximated with 5 iterations of warm started FGP Beck & Teboulle 2009.
- $\gamma = 0.95, \theta = 1$, uniform sampling $p_i = 1/n$

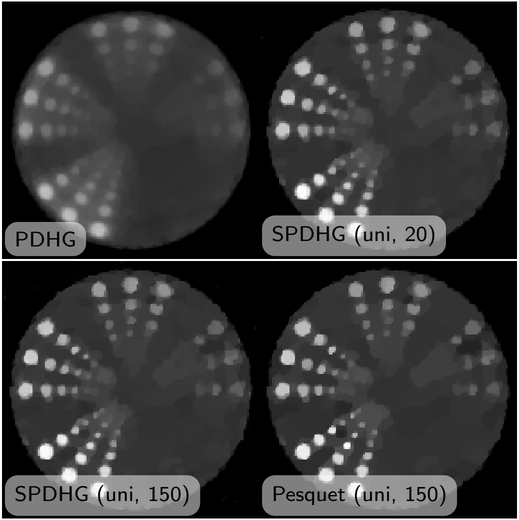
Compare methods:

- PDHG:
  $\sigma = \frac{\gamma}{\|\mathbf{A}\|} \approx$ 1.8e-03, $\tau = \frac{\gamma}{\|\mathbf{A}\|} \approx$ 1.8e-03
- SPDHG ($n = 20$):
  $\sigma_i = \frac{\gamma}{\|\mathbf{A}_i\|} \approx$ 8.0e-03 (mean), $\tau = \frac{\gamma}{n \max_i \|\mathbf{A}_i\|} \approx$ 3.8e-04
- SPDHG ($n = 150$):
  $\sigma_i = \frac{\gamma}{\|\mathbf{A}_i\|} \approx$ 1.6e-02 (mean), $\tau = \frac{\gamma}{n \max_i \|\mathbf{A}_i\|} \approx$ 7.7e-05
- Pesquet and Repetti 2015 ($n = 150$):
  $\sigma = \frac{\gamma}{\|\mathbf{A}\|} \approx$ 1.8e-03, $\tau = \frac{\gamma}{\|\mathbf{A}\|} \approx$ 1.8e-03

PET reconstruction with TV, $1/k$ rate

- 10 epochs





PET reconstruction with TV, $1/k$ rate

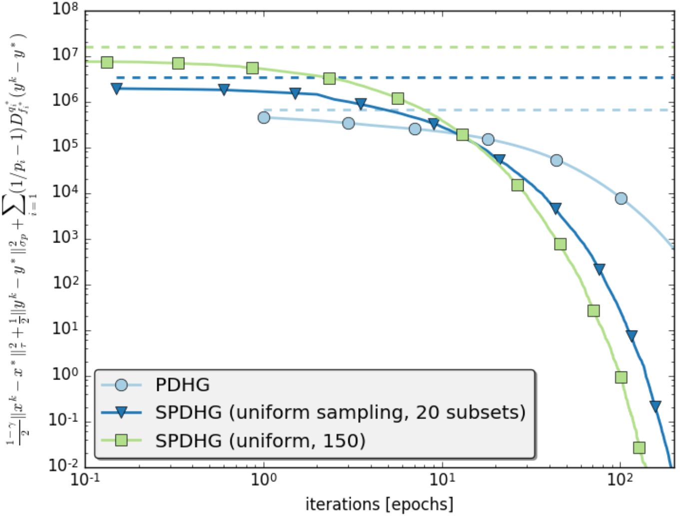


Figure: distance to a saddle point

PET reconstruction with TV, $1/k$ rate

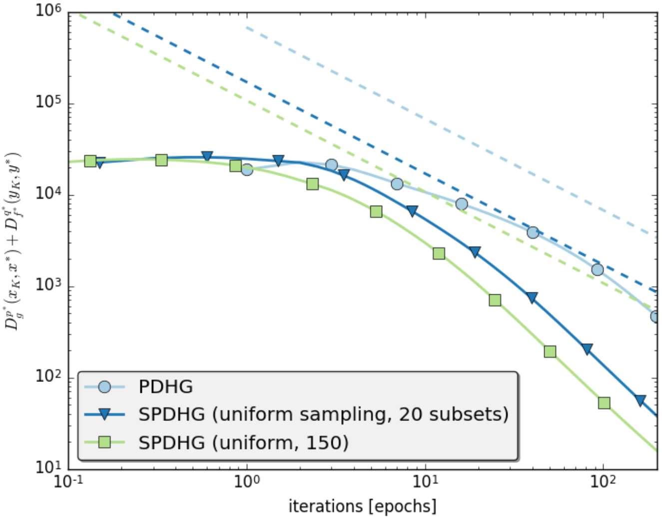


Figure: Bregman distance of ergodic sequence

PET reconstruction with TV, $1/k$ rate

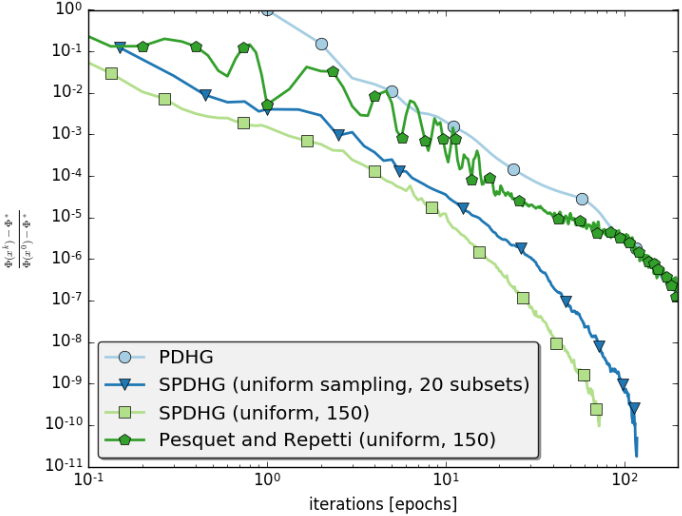


Figure: Objective function value

PET reconstruction with TV, $1/k$ rate

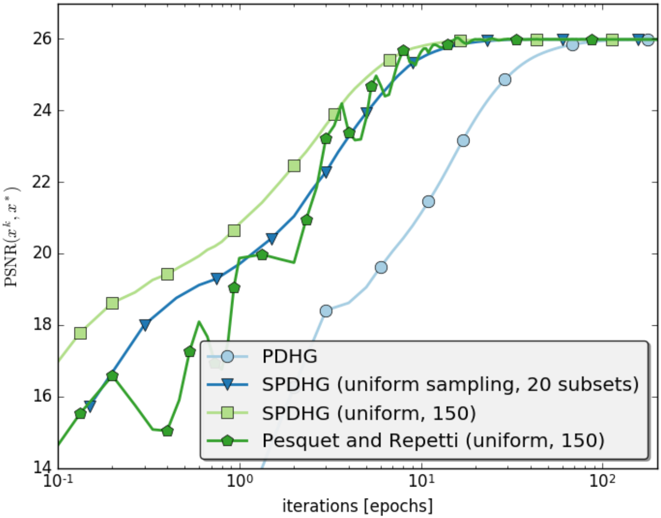


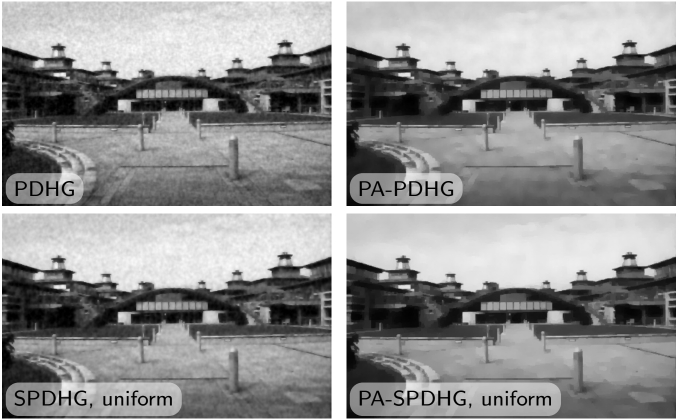
Figure: Peak signal-to-noise ratio

Extra Material: Other Imaging Tasks

TV Denoising

$$x^* \in \arg\min_x \left\{ \frac{1}{2} \|x - d\|^2 + \alpha \sum_{i=1}^{2} \|\nabla_i x\|_1 \right\}$$

- primal acceleration $1/k^2$, 20 epochs
- implemented using ODL Adler, Kohr, Öktem, 2017



TV Denoising

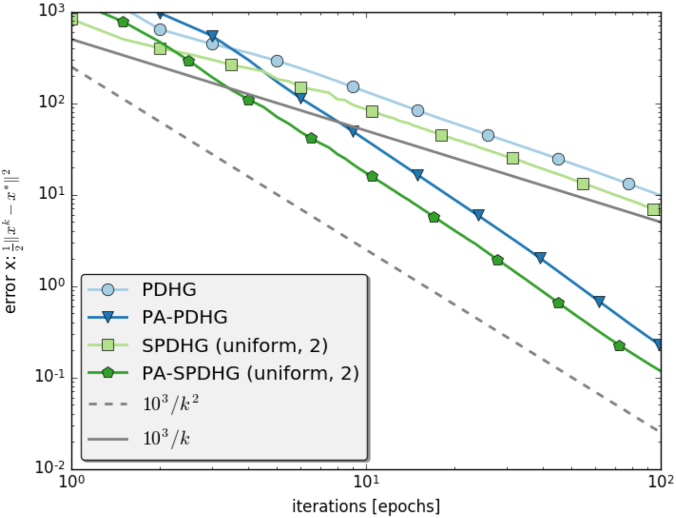


Figure: Primal distance to saddle point.

TV Denoising

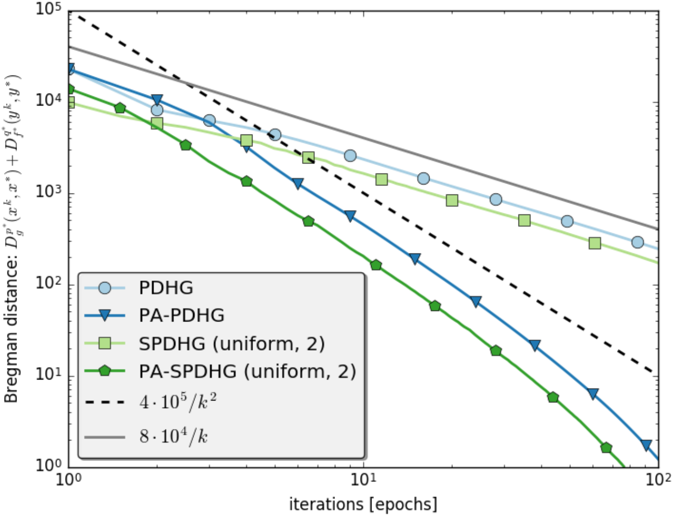


Figure: Bregman distance between iterates and saddle point.

TV Denoising

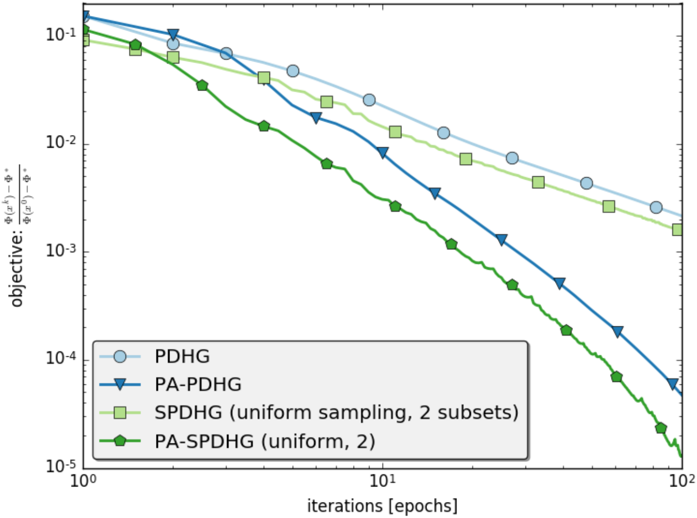


Figure: Relative objective function values.

TV Denoising

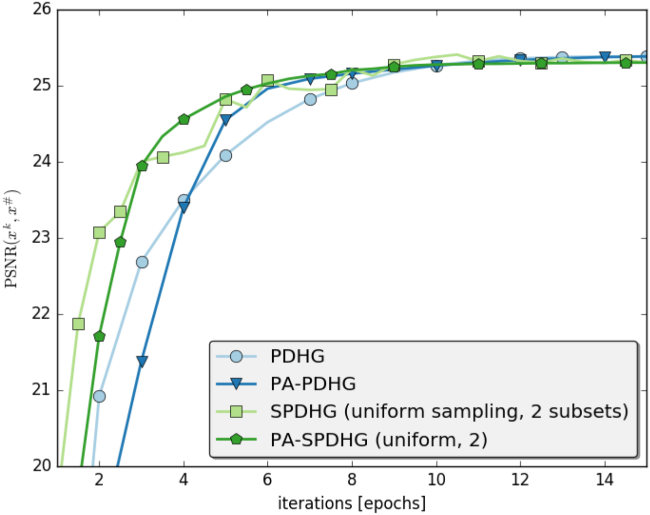
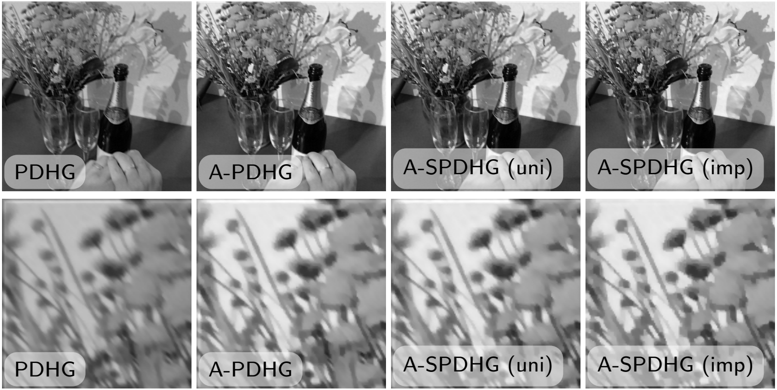


Figure: PSNR between iterates and ground truth solution.

Poisson TV deblurring with unknown boundary*

$$x^* \in \arg\min_{a \le x \le b} \left\{ \tilde{\mathrm{KL}}(d, M(x * k) + r) + \alpha \sum_{i=1}^{2} \mathrm{Huber}_\beta(\nabla_i x) \right\}$$

- dual acceleration $1/k^2$, 100 epochs



*Almeida, Figueiredo 2013

Poisson TV deblurring with unknown boundary

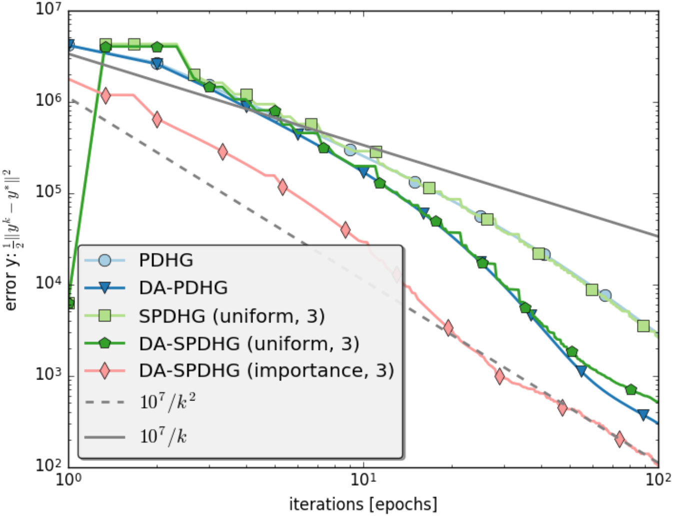


Figure: Distance to dual part of the saddle point.

Poisson TV deblurring with unknown boundary

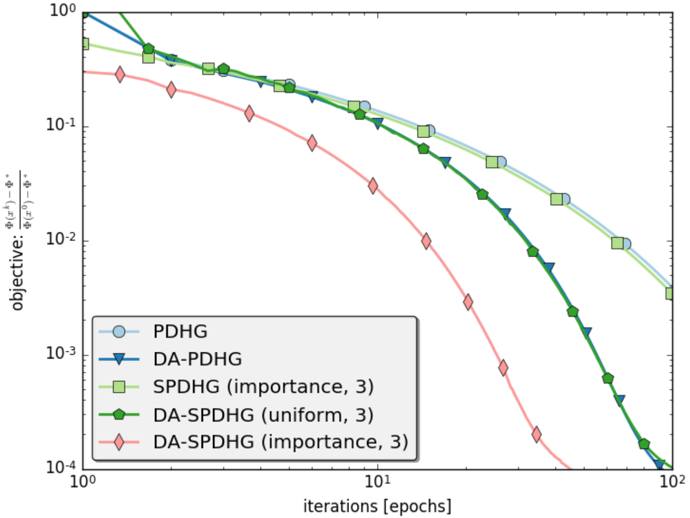


Figure: Relative objective function value.

Poisson TV deblurring with unknown boundary

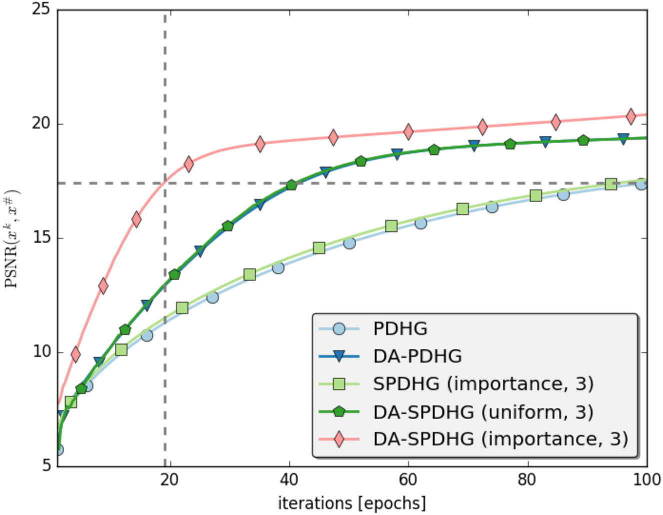


Figure: Peak signal-to-noise ratio (PSNR).

Extra Material: Further Details on Mathematical Abstraction

Mathematical Abstraction: $n+2$ Hilbert Spaces

Primal Space

- **space:** $\mathbb{X}$ **element:** $x \in \mathbb{X}$
- **inner product:** $\langle x, x' \rangle$ for $x, x' \in \mathbb{X}$
- **norm:** $\|x\| := \sqrt{\langle x, x \rangle}$

n Dual Block Spaces

- **space:** $\mathbb{Y}_i$, $i = 1, 2, \dots, n$ **element:** $y_i \in \mathbb{Y}_i$
- **inner product:** $\langle y_i, y_i' \rangle$ for $y_i, y_i' \in \mathbb{Y}_i$
- **norm:** $\|y_i\| := \sqrt{\langle y_i, y_i \rangle}$

Dual Product Space

- **space:** $\mathbb{Y} := \prod_{i=1}^n \mathbb{Y}_i$ **element:** $y = (y_1, \dots, y_n) \in \mathbb{Y}$
- **inner product:** $\langle y, y' \rangle := \sum_{i=1}^n \langle y_i, y_i' \rangle$
- **norm:** $\|y\|^2 := \sum_{i=1}^n \|y_i\|^2$

Mathematical Abstraction: Linear Operators

Block Operators

- $\mathbf{A}_i : \mathbb{X} \mapsto \mathbb{Y}_i$ for $i = 1, 2, \ldots, n$ (we write $\mathbf{A}_i x = y_i$)
- adjoint $\mathbf{A}_i^* : \mathbb{Y}_i \mapsto \mathbb{X}$

Aggregated Operator

- $\mathbf{A} : \mathbb{X} \mapsto \mathbb{Y}$ defined by

$$\mathbf{A}x := (\mathbf{A}_1 x, \mathbf{A}_2 x, \cdots, \mathbf{A}_n x)$$

- adjoint $\mathbf{A}^* : \mathbb{Y} \mapsto \mathbb{X}$ given by

$$\mathbf{A}^* y = \sum_{i=1}^{n} \mathbf{A}_i^* y_i$$