Random Primal-Dual Method for Parallel MRI Reconstruction

Eric Baruch Gutierrez Castillo

6 July 2022

Joint work with Claire Delplancke and Matthias J. Ehrhardt
University of Bath, UK

- **Examples of Convex Minimization**
- **Stochastic Primal-Dual Hybrid Gradient (SPDHG)**
- **Parallel MRI Reconstruction**

Convex Minimization

$$\hat{x} \in \operatorname*{argmin}_{x \in X} \sum_{i=1}^{n} f_i(A_i x) + g(x)$$

- $f_i : Y_i \to \mathbf{R}_\infty, \;\; g : X \to \mathbf{R}_\infty$ **convex, proper, lower-semicontinuous**
- $A_i : X \to Y_i$ **linear**

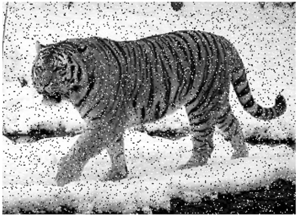
- f_i, g can be **non-smooth**
- $\hat{x}$ can be **not unique**

Examples

$$\hat{x} \in \operatorname*{argmin}_{x \in X} \sum_{i=1}^{n} f_i(A_i x) + g(x)$$

Image Denoising:

$$\hat{x} \in \operatorname*{argmin}_{x} \|\nabla x\|_1 + \lambda \|u^0 - x\|_1$$



Noisy image u^0

Solution $\hat{x}$

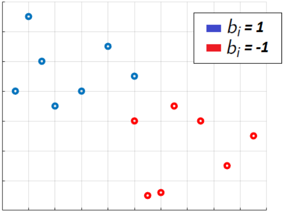
$$n = 1$$
$$A = \nabla$$
$$f = \|\cdot\|_1$$
$$g = \lambda\|u^0 - \cdot\|_1$$

Examples

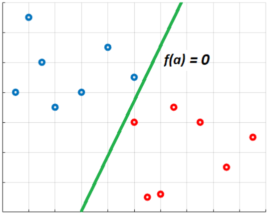
$$\hat{x} \in \operatorname*{argmin}_{x \in X} \sum_{i=1}^{n} f_i(A_i x) + g(x)$$

Binary Classification:

$$\hat{x} \in \operatorname*{argmin}_{x} \sum_{i=1}^{n} \max\{0, 1 - b_i \langle a_i, x \rangle\} + \frac{\lambda}{2}\|x\|^2$$



$b_i = \text{class}(a_i)$



$\langle a, \hat{x} \rangle = 0$

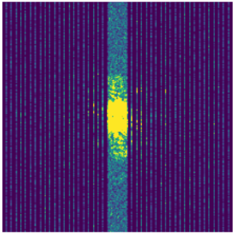
$$A_i = \langle a_i, \cdot \rangle$$
$$f_i = \max(0, 1 - b_i \cdot)$$
$$g = \frac{\lambda}{2}\| \cdot \|^2$$

Examples

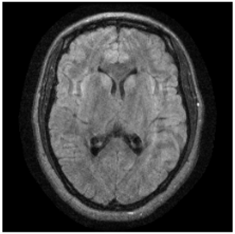
$$\hat{x} \in \operatorname*{argmin}_{x \in X} \sum_{i=1}^{n} f_i(A_i x) + g(x)$$

Parallel MRI Reconstruction:

$$\hat{x} \in \operatorname*{argmin}_{x} \sum_{i=1}^{n} \|(S \circ F \circ C_i)x - b_i\|^2 + \frac{\lambda}{2}\|x\|^2$$



Data b_i in Fourier space



Reconstructed image $\hat{x}$

$$A_i = S \circ F \circ C_i$$
$$f_i = \| \cdot - b_i\|^2$$
$$g = \frac{\lambda}{2}\| \cdot \|^2$$

Why Primal-Dual Methods?

$$\hat{x} \in \operatorname*{argmin}_{x \in X} \left\{ \Phi(x) := \sum_{i=1}^{n} f_i(A_i x) + g(x) \right\}$$

- Gradient Descent

$$x^{k+1} = x^k - \alpha \nabla \Phi(x^k)$$

requires smooth functionals f_i, g.

- Proximal Point Algorithm

$$x^{k+1} = \operatorname{prox}_{\sigma\Phi}(x^k)$$

proximity operator $\operatorname{prox}_{\sigma\Phi} = \operatorname{prox}_{\sigma(f \circ A + g)}$ **is generally not explicit**:

$$\operatorname{prox}_{\sigma\Phi}(x) := \operatorname*{argmin}_{u} \frac{\|u - x\|^2}{2\sigma} + \Phi(u)$$

$$\hat{x} \ \in \ \underset{x \in X}{\operatorname{argmin}} \left\{ \Phi(x) := \sum_{i=1}^{n} f_i(A_i x) + g(x) \right\}$$

$y = (y_1, ..., y_n)$:

$$\iff$$

$$(\hat{x}, \hat{y}) \ \in \ \arg\min_{x} \max_{y} \sum_{i=1}^{n} \langle A_i x, y_i \rangle - f_i^*(y_i) + g(x)$$

where f^* is the *fenchel dual*:

$$f^*(y) \ := \ \sup_{w} \langle w, y \rangle - f(w)$$

Primal-Dual Formulation

$$\hat{x} \;\in\; \underset{x\in X}{\operatorname{argmin}} \left\{ \Phi(x) := \sum_{i=1}^{n} f_i(A_i x) + g(x) \right\}$$

$y = (y_1, ..., y_n):$ $\iff$

$$(\hat{x}, \hat{y}) \;\in\; \arg\min_{x}\max_{y} \sum_{i=1}^{n} \langle A_i x, y_i\rangle - f_i^*(y_i) + g(x)$$

A **primal-dual algorithm** with step-size parameters $\tau, \sigma > 0$ reads:

$$\begin{aligned} x^{k+1} &= \operatorname{prox}_{\tau g}(x^k - \tau A^* y^k) \\ y_i^{k+1} &= \operatorname{prox}_{\sigma_i f_i^*}(y_i^k + \sigma_i A_i x^k) \qquad i \in \{1, ..., n\} \end{aligned}$$

In general, $\operatorname{prox}_{\tau g}$ and $\operatorname{prox}_{\sigma_i f_i^*}$ are explicit, while $\operatorname{prox}_{(f\circ A+g)}$ is not

Primal-Dual Hybrid Gradient

$$(\hat{x}, \hat{y}) \in \arg\min_{x} \max_{y} \sum_{i=1}^{n} \langle A_i x, y_i \rangle - f_i^*(y_i) + g(x)$$

PDHG :

$$\begin{aligned} x^{k+1} &= \text{prox}_{\tau g}(x^k - \tau A^* \bar{y}^k) \\ y_i^{k+1} &= \text{prox}_{\sigma f_i^*}(y_i^k + \sigma_i A_i x^{k+1}) \qquad i \in \{1, \dots, n\} \\ \bar{y}^k &= 2y^{k+1} - y^k \end{aligned}$$

Theorem (Chambolle & Pock 2011)

Let f^*, g be proper, convex and lower-semicontinuous, and let τ, σ satisfy

$$\|S^{1/2} A \tau^{1/2}\|^2 < 1$$

where $S = \text{diag}(\sigma_1, \dots, \sigma_n)$. Then PDHG converges to a saddle point.

Primal-Dual Hybrid Gradient

$$(\hat{x}, \hat{y}) \in \arg\min_{x} \max_{y} \sum_{i=1}^{n} \langle A_i x, y_i \rangle - f_i^*(y_i) + g(x)$$

PDHG :

$$\begin{aligned} x^{k+1} &= \mathrm{prox}_{\tau g}(x^k - \tau A^* \bar{y}^k) \\ y_i^{k+1} &= \mathrm{prox}_{\sigma f_i^*}(y_i^k + \sigma_i A_i x^{k+1}) \qquad i \in \{1, ..., n\} \\ \bar{y}^k &= 2y^{k+1} - y^k \end{aligned}$$

Idea: Update only one dual variable y_i at random.

Stochastic Primal-Dual Hybrid Gradient

$$(\hat{x}, \hat{y}) \in \arg\min_{x} \max_{y} \sum_{i=1}^{n} \langle A_i x, y_i \rangle - f_i^*(y_i) + g(x)$$

SPDHG: (serial sampling)

select $j \in \{1, \dots, n\}$ with probabilities $p_i = \mathbb{P}(i = j) > 0$

$$x^{k+1} = \operatorname{prox}_{\tau g}(x^k - \tau A^* \bar{y}^k)$$

$$y_i^{k+1} = \begin{cases} \operatorname{prox}_{\sigma_i f_i^*}(y_i^k + \sigma_i A_i x^{k+1}) & \text{if } i = j \\ y_i^k & \text{else} \end{cases}$$

$$\bar{y}^k = y^{k+1} + \theta {p_j}^{-1}(y^{k+1} - y^k)$$

1 iteration of PDHG $\approx$ n iterations of SPDHG

Stochastic Primal-Dual Hybrid Gradient

UNIVERSITY OF BATH

$$(\hat{x}, \hat{y}) \in \arg\min_{x} \max_{y} \sum_{i=1}^{n} \langle A_i x, y_i \rangle - f_i^*(y_i) + g(x)$$

SPDHG (arbitrary sampling $\mathbb{S}$):

select $\mathbb{S}^k \in \{1, \dots, n\}$ with probabilities $p_i = \mathbb{P}(i \in \mathbb{S}^k) > 0$

$$\begin{aligned}
x^{k+1} &= \operatorname{prox}_{\tau g}(x^k - \tau \bar{z}^k) \\
y_i^{k+1} &= \begin{cases} \operatorname{prox}_{\sigma_i f_i^*}(y_i^k + \sigma_i A_i x^{k+1}) & \text{if } i \in \mathbb{S}^k \\ y_i^k & \text{else} \end{cases} \\
\delta_i^k &= A_i^*(y_i^{k+1} - y_i^k) \text{ for all } i \in \mathbb{S}^k \\
z^{k+1} &= z^k + \textstyle\sum_{i \in \mathbb{S}^k} \delta_i^k \\
\bar{z}^{k+1} &= z^{k+1} + \theta \textstyle\sum_{i \in \mathbb{S}^k} p_i^{-1} \delta_i^k
\end{aligned}$$

only A_i and A_i^* such that $i \in \mathbb{S}^k$ are evaluated

Stochastic Primal-Dual Hybrid Gradient

Previous result:
Convergence in Bregman distance

Theorem (Chambolle et al. 2018):

Let f_i^*, g be convex, proper and lower-semicontinuous, and

$$\tau\sigma_i\|A_i\|^2 < p_i, \quad i \in \{1, ..., n\}$$

then the **Bregman distance** of the iterates (x^k, y^k) to a solution $(\hat{x}, \hat{y})$ converges almost surely to zero, i.e.

$$\left[\langle A^*\hat{y}, x^k\rangle - f^*(\hat{y}) + g(x^k)\right] - \left[\langle A\hat{x}, y^k\rangle - f^*(y^k) + g(\hat{x})\right] \overset{a.s.}{\to} 0$$

Convergence in Bregman distance **does not** imply convergence in the norm

Stochastic Primal-Dual Hybrid Gradient

New result:
Convergence for serial sampling and finite dimension

Theorem (Gutiérrez, Delplancke, Ehrhardt 2021)

Let f_i^*, g be convex, proper and lower-semicontinuous in **finite dimensional** separable Hilbert spaces, and

$$\tau \sigma_i \|A_i\|^2 < p_i, \quad i \in \{1, ..., n\}$$

then SPDHG with **serial sampling** converges a.s. to a saddle point.

New result:
Convergence for serial sampling and finite dimension

Theorem (Gutiérrez, Delplancke, Ehrhardt 2021)

Let f_i^*, g be convex, proper and lower-semicontinuous in **finite dimensional** separable Hilbert spaces, and

$$\tau \sigma_i \|A_i\|^2 < p_i, \quad i \in \{1, ..., n\}$$

then SPDHG with **serial sampling** converges a.s. to a saddle point.

Theorem (Alacaoglu et al. 2020)

Even newer result:
Convergence for any sampling and infinite dimensions

Theorem

Let f_i^*, g be convex, proper and lower-semicontinuous in separable Hilbert spaces of **arbitrary dimension** and let $\mathbb{S}$ denote **any random sampling** of $\{1, ..., n\}$. Let $\{v_1, ..., v_n\} > 0$ be such that

$$\mathbb{E}_{\mathbb{S}} \left\| \sum_{i \in \mathbb{S}} \tau^{1/2} \sigma_i^{1/2} A_i^* z_i \right\|^2 \leq \sum_{i=1}^{n} p_i v_i \|z_i\|^2 \quad \text{and} \quad v_i < p_i \quad i \in \{1, ..., n\}$$

Then SPDHG converges a.s. to a saddle point.

Furthermore, for different samplings we propose step size parameters for optimal convergence speed.

Parallel MRI Reconstruction

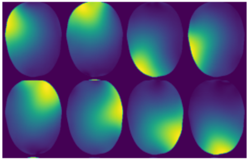
MRI Scanner

Original:

$$x^0$$

Encoded Data:

$$b_i = (S \circ F \circ C_i)x^0 + \eta_i$$



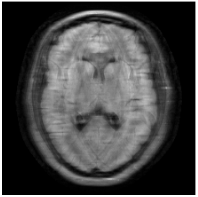
Sensitivity of each coil

Reconstructed Solution:

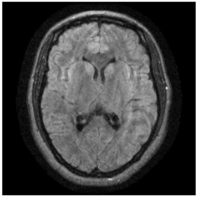
$$\hat{x} \in \underset{x}{\operatorname{argmin}} \sum_{i=1}^{n} \|(S \circ F \circ C_i)x - b_i\|^2 + \frac{\lambda}{2}\|x\|^2$$

Parallel MRI Reconstruction

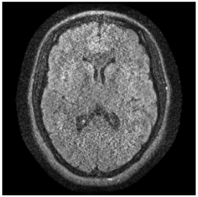
Dataset 1:



$\lambda = 1$

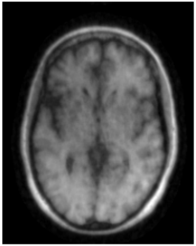


$\lambda = 10^{-2}$

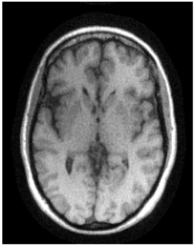


$\lambda = 10^{-4}$

Dataset 2:



$\lambda = 1$

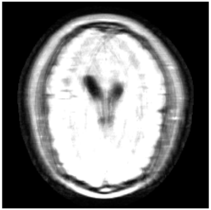


$\lambda = 10^{-2}$

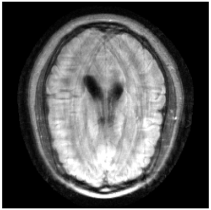
$\lambda = 10^{-4}$

Parallel MRI Reconstruction

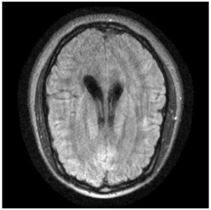
After 1 epoch:



PDHG



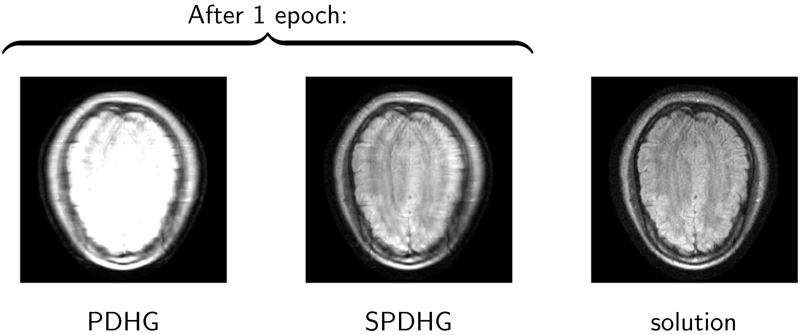
SPDHG



solution

1 iteration of PDHG $\approx n$ iterations of SPDHG $\approx$ 1 epoch

Parallel MRI Reconstruction



1 iteration of PDHG $\approx n$ iterations of SPDHG $\approx$ 1 epoch

Parallel MRI Reconstruction

full sampling: 1 block of size n $\implies$ 1 epoch $\approx$ 1 iteration

$$\Big\{\,(1, ..., n)\,\Big\}$$

serial sampling: n blocks of size 1 $\implies$ 1 epoch $\approx$ n iterations

$$\Big\{\,(1), (2), ..., (n)\,\Big\}$$

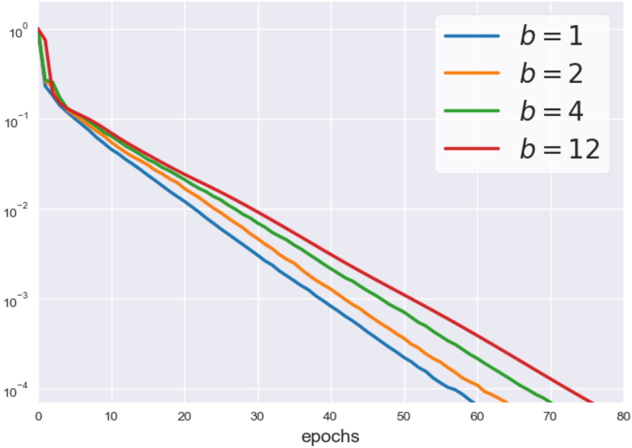
b**-serial sampling:** n/b blocks of size b $\implies$ 1 epoch $\approx$ n/b iterations

$$\Big\{\,(1, ..., b), (b+1, ..., 2b),\ \dots,(n-b+1, ..., n)\,\Big\}$$

Parallel MRI Reconstruction

Reconstruction error for $n = 12, \lambda = 10^{-2}$

$$\frac{\|x^k - \hat{x}\|}{\|\hat{x}\|}$$



The converge speed depends on the batch size

Parallel MRI Reconstruction

Theorem (Chambolle et al. 2018)

Let f_i^*, g be **strongly convex**, proper and lower-semicontinuous, and

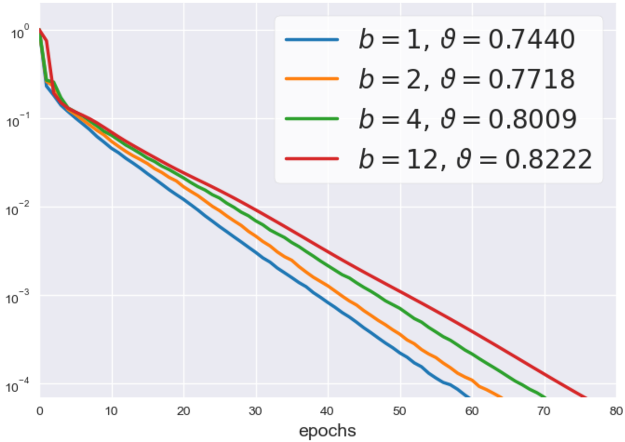
$$\tau\sigma_i\|A_i\|^2 < p_i, \quad i \in \{1, ..., n\}$$

then $(x^k, y^k) \overset{a.s.}{\to} (\hat{x}, \hat{y})$ with **linear rate** $\mathcal{O}(\theta^k)$.

Furthermore, for each sampling (full, serial, b-serial, etc...) we can compute the **optimal** convergence rate $\theta \in (0, 1)$.

Parallel MRI Reconstruction

$$\frac{\|x^k - \hat{x}\|}{\|\hat{x}\|}$$



The converge rate ϑ depends on the batch size

Parallel MRI Reconstruction

Given a fixed b, how many partitions of $\{1, ..., n\}$ exist of subsets of size b?

Given a fixed b, how many partitions of $\{1, ..., n\}$ exist of subsets of size b?

A lot.

Given a fixed b, how many partitions of $\{1, ..., n\}$ exist of subsets of size b?

A lot.

Example: $n = 12$, $b = 6$, there are 462 different partitions of $\{1, ..., n\}$ into two subsets of size 6.

$$\Big\{ (1, 2, 3, 4, 5, 6),\ (7, 8, 9, 10, 11, 12) \Big\}$$

$$\Big\{ (1, 2, 3, 4, 5, 7),\ (6, 8, 9, 10, 11, 12) \Big\}$$

$$\vdots$$

$$\Big\{ (1, 8, 9, 10, 11, 12),\ (2, 3, 4, 5, 6, 7) \Big\}$$

Parallel MRI Reconstruction

Given a fixed b, how many partitions of $\{1, ..., n\}$ exist of subsets of size b?

A lot.

Example: $n = 12$, $b = 6$, there are 462 different partitions of $\{1, ..., n\}$ into two subsets of size 6.

$$\Big\{ (1, 2, 3, 4, 5, 6),\ (7, 8, 9, 10, 11, 12) \Big\}$$

$$\Big\{ (1, 2, 3, 4, 5, 7),\ (6, 8, 9, 10, 11, 12) \Big\}$$

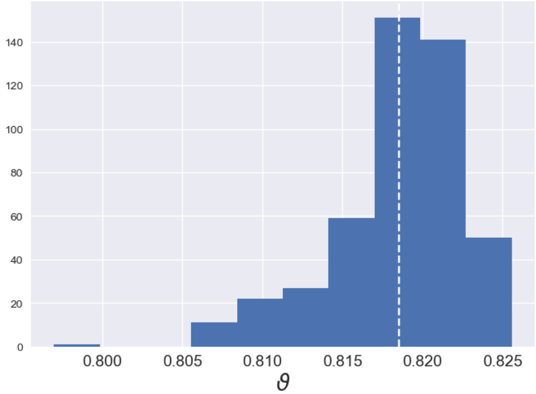
$$\vdots$$

$$\Big\{ (1, 8, 9, 10, 11, 12),\ (2, 3, 4, 5, 6, 7) \Big\}$$

Example: For $n = 12$, $b = 2$, the number of partitions is 10,395.

Example: $n = 12$, $b = 6$

For each possible sampling of size $b = 6$, we compute ϑ:

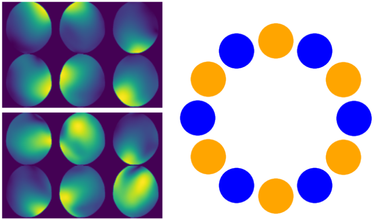


The convergence rate ϑ depends on which partition we sample from

Parallel MRI Reconstruction

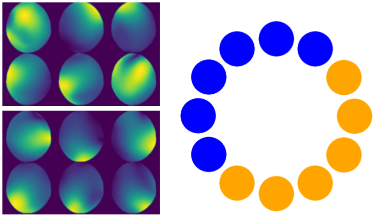
$$n = 12,\ b = 6$$

Best Partition:



$\vartheta = 0.7969$

Worst Partition:



$\vartheta = 0.8255$

The location of the coils gives us a clue on how to find the best partition

SPDHG

- Wide variety of applications
- Non-smooth functionals, large number of data
- Faster than PDHG

Our Contributions

- We prove convergence for any sampling
- We propose optimal step size parameters for different samplings
- For **parallel MRI** we show how convergence speed is affected by
    - Batch size b
    - The partition of $\{1, ..., n\}$
    - The physical location of the receiver coils