Stochastic primal-dual hybrid gradient algorithm

Claire Delplancke

University of Bath
PET++ project

VHackaton 29 June 2020

Motivation: PDHG vs OSEM

PDHG

- $x^{k+1} = \mathrm{prox}_{\tau G}(x^k - \tau K^* \bar{y}^k)$
- $y^{k+1} = \mathrm{prox}_{\sigma F^*}(y^k + \sigma K x^k)$
- $\bar{y}^{k+1} = y^{k+1} + \theta(y^{k+1} - y^k)$

$K = (\nabla, A)$: at each iteration, full forward and back projections A and A^*.

OSEM $A = \begin{bmatrix} A_1 \\ \vdots \\ A_n \end{bmatrix}$, no prior. Update rule: for $i = \mathrm{mod}(k, n)$,

$$x^{k+1} = \frac{x^k}{A_i^* \mathbf{1}} A_i^* \left(\frac{g_i}{A_i x_k + \eta_i} \right)$$

At each iteration: only partial forward and back projections.

Saddle point and separable structure

$$\min_{x\in X} F(Kx) + G(x) = \min_{x\in X} \sup_{y\in Y} \langle Kx, y\rangle - F^*(y) + G(x).$$

Let's assume that

$$K = \begin{bmatrix} K_1 \\ \vdots \\ K_m \end{bmatrix} : X \to Y = Y_1 \times \cdots \times Y_m, \quad F(y) = \sum_{i=1}^m F_i(y_i).$$

Then,

$$\begin{aligned} \min_{x\in X} F(Kx) + G(x) &= \min_{x\in X} \sum_{i=1}^m F_i(K_i x) + G(x) \\ &= \min_{x\in X} \sup_{y\in Y} \sum_{i=1}^m \langle K_i x, y_i\rangle - F_i^*(y_i) + G(x). \end{aligned}$$

Stochastic Primal-Dual Hybrid Gradient[1]

PDHG (with extrapolation on dual variable y)

- $x^{k+1} = \text{prox}_{\tau G}(x^k - \tau K^* \bar{y}^k)$ (Gradient descent primal variable)
- $y^{k+1} = \text{prox}_{\sigma F^*}(y^k + \sigma K x^k)$ (Gradient ascent dual variable)
- $\bar{y}^{k+1} = y^{k+1} + \theta(y^{k+1} - y^k)$ (Extrapolation step dual variable).

SPDHG

- $x^{k+1} = \text{prox}_{\tau G}(x^k - \tau K^* \bar{y}^k)$ (Gradient descent primal variable)
- Choose index i with probability p_i
- $y_i^{k+1} = \text{prox}_{\sigma_i F_i^*}(y_i^k + \sigma_i K_i x^k)$ (Gradient ascent on i-th dual variable)
- $\bar{y}_i^{k+1} = y_i^{k+1} + \frac{\theta}{p_i}(y_i^{k+1} - y_i^k)$ (Extrapolation step dual variable).

---

[1] *Stochastic primal-dual hybrid gradient algorithm with arbitrary sampling and imaging applications*, A. Chambolle, M. J. Ehrhardt, P. Richtárik and C.-B. Schönlieb, SIAM J. Optim, 2018
Faster PET Reconstruction with Non-Smooth Priors by Randomization and Preconditioning, M. J. Ehrhardt, P. Markiewicz and C.-B. Schönlieb, PMB, 2019

Stochastic Primal-Dual Hybrid Algorithm

Notice that

$$K^*\bar{y}^{k+1} = K^*\bar{y}^k + \left(1 + \frac{\theta}{p_i}\right) K_i^*(y_i^{k+1} - y_i^k),$$

hence one needs to compute only a partial backprojection K_i^*, and to store the variable $A^*\bar{y}^k$ (at same memory requirement than x).

Theorem

For $\theta = 1$, and for each i, $p_i > 0$ and $\sigma_i \tau \|A_i\|^2 < p_i$, SPDHG converges to a saddle-point of the primal-dual problem almost surely in Bregman distance.

SPDHG with preconditioning

For a symmetric, positive-definite T, define

$$\mathrm{prox}_G^T(x) = (I + TG)^{-1}(x) = \arg\min_{u \in X} \frac{1}{2}\langle (u-x), T^{-1}(u-x)\rangle + G(u).$$

- $x^{k+1} = \mathrm{prox}_G^T(x^k - TK^*\bar{y}^k)$ (Gradient descent primal variable)
- Choose index i with probability p_i
- $y_i^{k+1} = \mathrm{prox}_{F_i^*}^{\Sigma_i}(y_i^k + \Sigma_i K_i x^k)$ (Gradient ascent on i-th dual variable)
- $\bar{y}_i^{k+1} = y_i^{k+1} + \frac{\theta}{p_i}(y_i^{k+1} - y_i^k)$ (Extrapolation step dual variable).

Convergence ensured if

- for all i, $\|\Sigma_i^{1/2} A_i T^{1/2}\|^2 < p_i$
- or[2] $\|\Sigma^{1/2} A T^{1/2}\|^2 < 1$ where $\Sigma = \begin{bmatrix}\Sigma_1 & \cdots & \Sigma_n\end{bmatrix}$.

[2] C. Delplancke, J. Latz, P. J. Markiewicz, C.-B. Schönlieb, M. J. Ehrhardt *Improving a stochastic algorithm for regularized PET image reconstruction*, submitted

Choice of step-sizes

Take $\gamma > 0$ and $\rho < 1$.

- Scalar step-sizes:

$$\sigma_i = \frac{\gamma\rho}{\|A_i\|}, \quad \tau = \frac{\rho}{\gamma \sum_{i=1}^m \|A_i\|};$$

- Diagonal step-sizes:

$$\Sigma_i = \mathrm{Diag}\left(\frac{\gamma\rho}{\sum_l |A_i(k,l)|}\right), \quad \tau = \mathrm{Diag}\left(\frac{\rho}{\gamma \sum_i \sum_k |A_i(k,l|}\right).$$

Remark: if A_i has non-negative elements, then $(\sum_l |A_i(k,l)|)_k = A_i\mathbf{1}$ and $(\sum_k |A_i(k,l)|)_l = A_i^*\mathbf{1}$.

SPDHG for PET reconstruction

Let's consider a partition of all angles in n subsets. Call A_i the forward projector corresponding to the i-th subset, and call

$$F_i(y_i) = \sum_{j \in i\text{-th subset}} y_j - g_j \log(y_j + \eta_j).$$

Then:

$$\begin{aligned} &\arg\min \sum Ax - g \log(Ax + \eta) + \alpha \|\nabla x\|_{1,2} + i^+(x) \\ = &\arg\min \sum_{i=1}^{m} F_i(K_i x) + G(x), \end{aligned}$$

with $G = i^+$, $m = n + 1$,

- for $i = 1 \dots n$, $K_i = A_i$ and F_i defined above
- $K_{n+1} = \nabla$ and $F_{n+1} = \|\|_{1,2}$.

A good choice of probabilities is $p_i = 1/(2n)$ for $i = 1, \dots, n$ and $p_{n+1} = 1/2$.