Primal-Dual Hybrid Gradient algorithm Preconditioning

Evangelos Papoutsellis

University of Manchester

evangelos.papoutsellis@manchester.ac.uk

CCPi/SyneRBI vhackathon
June 2020

Primal-Dual Hybrid Gradient (PDHG) algorithm

$$x^* \in \underset{x \in X}{\arg\min}\, \mathcal{F}(Kx) + \mathcal{G}(x) \tag{1}$$

1. X, Y finite dimensional vector spaces, $K : X \to Y$ is a linear operator.
2. $\mathcal{F}, \mathcal{G}$ proper, convex & lower-semicontinuous.
3. $\mathcal{G} : X \to [0, \infty)$, $\mathcal{F} : Y \to [0, \infty)$

Examples

$$\underset{u \in X}{\arg\min} \left\{ \alpha \mathrm{TV}(u) + \frac{1}{2} \|Au - g\|_2^2 \right\} \tag{2}$$

$$\underset{u \in X}{\arg\min} \left\{ \alpha \mathrm{TV}(u) + \sum Au - g \log(Au + \eta) + \mathbb{I}_{\{u>0\}}(u) \right\} \tag{3}$$

1. $\mathrm{TV}(u) := \|\nabla u\|_{2,1} = \sum \sqrt{(\partial_x u)^2 + (\partial_y u)^2}$
2. g: Acquisition Data
3. α: Regularising parameter
4. A: Projection operator

Primal-Dual Hybrid Gradient (PDHG) algorithm

PDHG algorithm solves non-smooth convex optimization problems with *saddle-point* structure:

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

$$\min_{x\in X}\Big\{\mathcal{F}(Kx) + \mathcal{G}(x)\Big\} \qquad \text{(Primal problem)}$$

$$\max_{y\in Y}\Big\{-\mathcal{G}^*(-K^*y) - \mathcal{F}^*(y)\Big\} \qquad \text{(Dual problem)}$$

$\mathcal{F}^*$, $\mathcal{G}^*$ convex conjugates of functionals $\mathcal{F}$, $\mathcal{G}$.

- (x^*, y^*) is a saddle point, i.e.,

  $$\mathcal{L}(x^*, y) \leq \mathcal{L}(x^*, y^*) \leq \mathcal{L}(x, y^*), \forall (x, y) \in X \times Y$$

  for $\mathcal{L}(x, y) := \langle Kx, y\rangle - \mathcal{F}^*(y) + \mathcal{G}(x)$.
- (x^*, y^*) is a saddle point $\Rightarrow$ x^* is a solution of the Primal Problem, y^* is a solution of the Dual problem.
- *Primal-Dual Gap* $:= \mathcal{F}(Kx) + \mathcal{G}(x) + \mathcal{G}^*(-K^*y) + \mathcal{F}^*(y) \geq 0$
- If (x^*, y^*) is a saddle point then *PDgap* vanishes.

Primal-Dual Hybrid Gradient (PDHG) algorithm

PDHG (or Chambolle-Pock) algorithm

- $(x^0, y^0) \in (X \times Y), \bar{x}^0 = x^0$.
- $\tau, \sigma > 0$ (stepsize parameters) such that $\tau\sigma\|K\|^2 < 1$, $\theta \in [0, 1]$.
- $\|K\| = \sup \Big\{ \|Kx\| : \|x\| \leq 1 \Big\}$.

Update:

- $y^{n+1} = (I + \sigma\partial F^*)^{-1}(y^n + \sigma K\bar{x}^n)$ (Gradient ascend dual variable)
- $x^{n+1} = (I + \tau\partial G)^{-1}(x^n - \tau K^*\bar{y}^{n+1})$ (Gradient descent primal variable)
- $\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$ (Over-relaxation step)

$$
\begin{aligned}
(I + \sigma\partial F^*)^{-1}(y) &:= \underset{z\in Y}{\arg\min} \left\{ \frac{1}{2}\|y - z\|^2 + \sigma\mathcal{F}^*(z) \right\} =: \mathrm{prox}_{\sigma\mathcal{F}^*}(y) \\
(I + \tau\partial G)^{-1}(x) &:= \underset{z\in X}{\arg\min} \left\{ \frac{1}{2}\|x - z\|^2 + \tau\mathcal{G}(z) \right\} =: \mathrm{prox}_{\tau\mathcal{G}}(x)
\end{aligned} \tag{4}
$$

Assumption: Proximal operators are "easy to solve", i.e. have closed form solution or can be efficiently computed with a high precision.

Primal-Dual Hybrid Gradient (PDHG) algorithm

- PDHG is a powerful and flexible algorithm with simple subproblems to solve.
- Requires careful choice of stepsize parameters (σ, τ) to accelerate convergence e.g., $(\sigma, \tau) = (\frac{1}{\|K\|}, \frac{1}{\|K\|})$, $(\sigma, \tau) = (10, \frac{1}{\sigma\|K\|^2})$
  - Tradeoff between primal & dual residuals. Large τ value $\Rightarrow$ Strong minimization step in the primal variable $\Rightarrow$ Slow maximization in the dual variable.
- σ, τ depends on $\|K\|$ which in practice cannot be computed exactly $\Rightarrow$ **Power iteration**. Large operator norm $\Rightarrow$ Slow convergence.

Alternatives

- Adaptive update (σ, τ) based on (convex) properties of $\mathcal{F}^*, \mathcal{G}$. [1]
- Adaptive update (σ, τ) based on primal & dual residuals. [2].
- Preconditioning PDHG $\Rightarrow$ Replace scalars τ, σ by T, Σ symmetric and positive definite matrices. [3]

---

[1] Chambolle and Pock 2010, [2] Goldstein, Li, Yuan, Esser, Baraniuk, 2015, [3] Chambolle and Pock 2011, Pock, Cremers, Bischof, Chambolle 2009, Esser, Zhang, Chan, 2010

Preconditioning PDHG

PDHG

- $(x^0, y^0) \in (X \times Y), \bar{x}^0 = x^0$.
- $\tau, \sigma > 0$ (stepsize parameters) such that $\tau\sigma\|K\|^2 < 1$, $\theta \in [0, 1]$.

Update:

- $y^{n+1} = (I + \sigma\partial F^*)^{-1}(y^n + \sigma K\bar{x}^n)$
- $x^{n+1} = (I + \tau\partial G)^{-1}(x^n - \tau K^*\bar{y}^{n+1})$
- $\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$

Preconditioning PDHG

Preconditioning - PDHG

- $(x^0, y^0) \in (X \times Y), \bar{x}^0 = x^0$.
- Σ, T (symmetric, positive definite matrices) such that $\|\Sigma^{1/2} K \mathrm{T}^{1/2}\|^2 < 1$, $\theta = 1$.

Update:

- $y^{n+1} = (I + \Sigma \partial F^*)^{-1}(y^n + \Sigma K \bar{x}^n)$
- $x^{n+1} = (I + \mathrm{T} \partial G)^{-1}(x^n - \mathrm{T} K^* \bar{y}^{n+1})$
- $\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$

$$(I + \Sigma \partial F^*)^{-1}(y) := \operatorname*{arg\,min}_{z \in Y} \left\{ \frac{1}{2} \|y - z\|_{\Sigma}^2 + \mathcal{F}^*(z) \right\}, \quad (5)$$

where,

$$\|x\|_{\Sigma}^2 := \|\Sigma^{-1/2} x\|^2 = \left\langle \Sigma^{-1} x, x \right\rangle.$$

Preconditioning PDHG

Preconditioning - PDHG

- $(x^0, y^0) \in (X \times Y), \bar{x}^0 = x^0$.
- Σ, T (symmetric, positive definite matrices) such that $\|\Sigma^{1/2} K \mathrm{T}^{1/2}\|^2 < 1$, $\theta = 1$.

Update:

- $y^{n+1} = (I + \Sigma \partial F^*)^{-1}(y^n + \Sigma K \bar{x}^n)$
- $x^{n+1} = (I + \mathrm{T} \partial G)^{-1}(x^n - \mathrm{T} K^* y^{n+1})$
- $\bar{x}^{n+1} = x^{n+1} + \theta(x^{n+1} - x^n)$

Choice of preconditioners: Let $a \in [0, 2]$, $\mathrm{T} = \mathrm{diag}(\boldsymbol{\tau}),\ \Sigma = \mathrm{diag}(\boldsymbol{\sigma})$

$$\begin{aligned} \boldsymbol{\tau} &= (\tau_j)_{1 \le j \le n} = \frac{1}{\sum_{i=1}^m |K_{i,j}|^{2-a}}, \quad j = 1, \ldots, n \quad \text{(Sum over columns)} \\ \boldsymbol{\sigma} &= (\sigma_i)_{1 \le i \le m} = \frac{1}{\sum_{j=1}^n |K_{i,j}|^{a}}, \quad i = 1, \ldots, m \quad \text{(Sum over rows)} \end{aligned} \tag{5}$$

Example: (Precond) PDHG

$$\underset{u\in X}{\arg\min}\left\{\alpha\|\nabla u\|_{2,1}+\sum Au-g\log(Au+\eta)+\mathbb{I}_{\{u>0\}}(u)\right\}$$
$$\Updownarrow \tag{6}$$
$$\underset{x\in X}{\arg\min}\,\mathcal{F}(Kx)+\mathcal{G}(x)$$

(Explicit) PDHG

$$K=\begin{bmatrix}\nabla\\ A\end{bmatrix},\ \mathcal{G}(x):=\mathcal{G}(u)=\mathbb{I}_{\{u>0\}}(u),\ \sigma=\tau=\frac{1}{\|K\|}$$

$$\mathcal{F}(z)=\sum_{i=1}^{2}F_i(z_i):=\alpha\|z_1\|_{2,1}+\sum z_2-g\log(z_2+\eta)\quad\text{(Separable Sum)}$$

$$\mathrm{prox}_{\tau\mathcal{G}}(\hat{u})=\max\{\hat{u},0\}$$

$$\mathrm{prox}_{\sigma\mathcal{F}^*}(\hat{z})=\begin{bmatrix}\mathrm{prox}_{\sigma F_1^*}(\hat{z}_1)=\frac{\hat{z}_1}{\max\{1,\frac{|\hat{z}_1|_2}{\alpha}\}}\\ \mathrm{prox}_{\sigma F_2^*}(\hat{z}_2)=0.5\left(\hat{z}_2+\sigma\eta+1-\sqrt{(\hat{z}_2+\sigma\eta-1)^2+4\sigma g}\right)\end{bmatrix}$$

Example: (Precond) PDHG

$$\begin{aligned} &\underset{u \in X}{\arg\min} \left\{ \alpha \|\nabla u\|_{2,1} + \sum Au - g \log(Au + \eta) + \mathbb{I}_{\{u>0\}}(u) \right\} \\ &\Updownarrow \\ &\underset{x \in X}{\arg\min}\, \mathcal{F}(Kx) + \mathcal{G}(x) \end{aligned} \tag{6}$$

(Explicit) Precond-PDHG

$$K = \begin{bmatrix} \nabla \\ A \end{bmatrix}, \ \mathcal{G}(x) := \mathcal{G}(u) = \mathbb{I}_{\{u>0\}}(u), \ \Sigma, \mathrm{T} : \ \text{(Diagonal Matrices)}$$

$$\mathcal{F}(z) = \sum_{i=1}^{2} F_i(z_i) := \alpha \|z_1\|_{2,1} + \sum z_2 - g \log(z_2 + \eta) \quad \text{(Separable Sum)}$$

$$\mathrm{prox}_{\mathrm{T}\mathcal{G}}(\hat{u}) = \max\{\hat{u}, 0\}$$

$$\mathrm{prox}_{\Sigma \mathcal{F}^*}(\hat{z}) = \begin{bmatrix} \mathrm{prox}_{\Sigma_1 F_1^*}(\hat{z}_1) = \frac{\hat{z}_1}{\max\{1, \frac{|\hat{z}_1|_2}{\alpha}\}} \\ \mathrm{prox}_{\Sigma_2 F_2^*}(\hat{z}_2) = 0.5\left(\hat{z}_2 + \Sigma_2 \eta + 1 - \sqrt{(\hat{z}_2 + \Sigma_2 \eta - 1)^2 + 4\Sigma_2 g}\right) \end{bmatrix}$$

$$\arg\min_{u\in X}\left\{\alpha\|\nabla u\|_{2,1}+\sum Au-g\log(Au+\eta)+\mathbb{I}_{\{u>0\}}(u)\right\} \Updownarrow \arg\min_{x\in X}\mathcal{F}(Kx)+\mathcal{G}(x) \tag{6}$$

(Explicit) PDHG $\Rightarrow$ in CIL

```
op1 = Gradient(ig)
op2 = AstraProjectorSimple(ig, ag)
K = BlockOperator(op1, op2)

f = BlockFunction(alpha * MixedL21Norm(),
                  KullbackLeibler(b=noisy_data, eta = eta))

normK = K.norm()
sigma = 1/normK
tau = 1/normK

pdhg = PDHG(f = f, g = g, operator = K, tau = tau, sigma = sigma,
            max_iteration = 1000, update_objective_interval = 200)
pdhg.run(very_verbose = True)
```

Example: (Precond) PDHG

$$\underset{u\in X}{\arg\min}\left\{\alpha\|\nabla u\|_{2,1}+\sum Au-g\log(Au+\eta)+\mathbb{I}_{\{u>0\}}(u)\right\}$$
$$\Updownarrow \tag{6}$$
$$\underset{x\in X}{\arg\min}\,\mathcal{F}(Kx)+\mathcal{G}(x)$$

(Implicit) PDHG

$K=A,\ \ \mathcal{G}(x):=\mathcal{G}(u)=\alpha\|\nabla u\|_{2,1}+\mathbb{I}_{\{u>0\}}(u),\quad \mathcal{F}(z)=\sum z-g\log(z+\eta)$

$$\operatorname{prox}_{\sigma F_2^*}(\hat{z})=0.5\left(\hat{z}+\sigma\eta+1-\sqrt{(\hat{z}+\sigma\eta-1)^2+4\sigma g}\right) \tag{7}$$

$$\operatorname{prox}_{\tau(\alpha\|\nabla(\cdot)\|_{2,1})}(\hat{u})=\underset{u>0}{\arg\min}\left\{\frac{1}{2}\|u-\hat{u}\|^2+\alpha\|\nabla u\|_{2,1}\right\} \tag{8}$$

$\Rightarrow$ Solve using **Fast Gradient Projection** (FGP) algorithm.*

Example: (Precond) PDHG

$$\underset{u\in X}{\arg\min}\left\{\alpha\|\nabla u\|_{2,1}+\sum Au-g\log(Au+\eta)+\mathbb{I}_{\{u>0\}}(u)\right\}$$
$$\Updownarrow \tag{6}$$
$$\underset{x\in X}{\arg\min}\,\mathcal{F}(Kx)+\mathcal{G}(x)$$

(Implicit) Precond - PDHG

$K=A,\ \ \mathcal{G}(x):=\mathcal{G}(u)=\alpha\|\nabla u\|_{2,1}+\mathbb{I}_{\{u>0\}}(u),\quad \mathcal{F}(z)=\sum z-g\log(z+\eta)$
Since, K is a positive operator, i.e., $Kx\geq 0,\ \forall x\geq 0$, for $a=1$, we have:

$$\Sigma=\frac{1}{K\mathbf{1}_X},\qquad \mathrm{T}=\frac{1}{K^T\mathbf{1}_Y}$$

$$\mathrm{prox}_{\Sigma F_2^*}(\hat{z})=0.5\left(\hat{z}+\Sigma\eta+1-\sqrt{(\hat{z}+\Sigma\eta-1)^2+4\Sigma g}\right) \tag{7}$$

$$\mathrm{prox}_{\mathrm{T}(\alpha\|\nabla(\cdot)\|_{2,1})}(\hat{u})=\underset{u>0}{\arg\min}\left\{\frac{1}{2}\|u-\hat{u}\|^2_{\mathrm{T}}+\alpha\|\nabla u\|_{2,1}\right\} \tag{8}$$

$\Rightarrow$ Solve using **Fast Gradient Projection** (FGP) algorithm.*

---

* Beck, Teboulle 2009

Example: (Precond) PDHG

$$\arg\min_{u\in X}\left\{\alpha\|\nabla u\|_{2,1}+\sum Au-g\log(Au+\eta)+\mathbb{I}_{\{u>0\}}(u)\right\}$$
$$\Updownarrow \tag{6}$$
$$\arg\min_{x\in X}\mathcal{F}(Kx)+\mathcal{G}(x)$$

(Implicit) PDHG

```
K =  AstraProjectorSimple(ig, ag)

f = KullbackLeibler(b = noisy_data, eta = eta)
g = TotalVariation(alpha, num_iters, lower = 0)

if precond:
    tau = 1/K.adjoint(K.range_geometry().allocate(1))
    sigma = 1/K.direct(K.domain_geometry().allocate(1))
else:
    normK = K.norm()
    sigma = 1/normK
    tau = 1/normK

pdhg = PDHG(f = f, g = g, operator = K, tau = tau, sigma = sigma,
            max_iteration = 1000, update_objective_interval = 200)
pdhg.run(very_verbose = True)
```

Thank you for your attention