Bregman Method from PEP Perspective

Zhenghao Xu

Zhejiang University

October 31, 2021

Table of Contents

Table of Contents

Formulation

Problem:

$$\min_{x\in\mathbb{R}^d} \quad F(x) \triangleq f(x) + \phi(x) \tag{P}$$

Assumptions:

- f is *convex*, *proper*, *lsc* and continuously differentiable.
- ϕ is *convex*, *proper*, *lsc* (possibly nonsmooth).
- h is a Legendre kernel function and is σ-strongly convex.
- f is L-Lipschitz continuous.
- bounded initial distance and well-posedness of problem and method.

Formulation

Problem:

$$\min_{x \in \mathbb{R}^d} \quad F(x) \triangleq f(x) + \phi(x) \tag{P}$$

Method: Bregman Proximal Point Subgradient Method.

$$y_k = \arg\min_{u \in \mathbb{R}^d} \phi(u) + \frac{1}{\lambda_k} D_h(u, x_k) \tag{1}$$

$$x_{k+1} = \nabla h^* \left(\nabla h(y_k) - \lambda_k f'(y_k) \right), \quad f'(y_k) \in \partial(y_k) \tag{2}$$

$$y_k = \underset{u\in\mathbb{R}^d}{\arg\min}\, \phi(u) + \frac{1}{\lambda_k} D_h(u, x_k)$$
$$x_{k+1} = \nabla h^* \left(\nabla h(y_k) - \lambda_k f'(y_k)\right), \quad f'(y_k) \in \partial(y_k)$$

$\Longrightarrow$

$$\begin{cases} \nabla h(y_0) = \nabla h(x_0) - \lambda_k \phi'(y_0), \\ \nabla h(y_{k+1}) = \nabla h(y_k) - \lambda_k f'(y_k) - \lambda_{k+1}\phi'(y_{k+1}) \quad k = 0, 1, 2, \ldots. \end{cases} \tag{3}$$

When $\lambda_k \equiv \lambda$, it is equivalent to BPG except the first step and replacement from gradient to subgradient of f.

Convergence Result

Convergence rate derived through PEP:

$$\min_{i=0,\dots,N} F(y_i) - F(x_*) \le \frac{D_h(x_*, x_0) + \frac{L^2}{2\sigma}\sum_{i=0}^N \lambda_i^2}{\sum_{i=0}^N \lambda_i}$$

this matches the usual subgradient method rate with
$D_h(x_*, x_0) = \frac{1}{2}\|x_* - x_0\|_2^2$.
Better constant than degenerated case in [Boț and Böhm, 2019]:

$$\mathbb{E}\left(\min_{0\le k\le N-1}\left(\sum_{i=1}^m f_i + g\right)(x_{k+1}) - \left(\sum_{i=1}^m f_i + g\right)(y)\right) \le \frac{2\sigma D_H(y, x_0) + \left(2\left(\sum_{i=1}^m \frac{1}{p_i^2}\right)^{\frac{1}{2}} + 3 + 2m\right)\left(\sum_{i=1}^m L_{f_i}\right)^2 \sum_{k=0}^{N-1} t_k^2}{2\sigma \sum_{k=0}^{N-1} t_k}.$$

Table of Contents

L-smad BPG

L-smad [Bolte et al., 2018]: $Lh - f$ and $Lh + f$ convex.
Result (checked, suitable for any $N \geq 1$, $\lambda \in (0, 1/L]$):

$$\min_{i=0,\ldots,N-1} D_h(x_i, x_{i+1}) \leq \frac{D_h(x_*, x_0)}{N - 2 + 2^{1-N}}$$

When $\lambda = 1/L$, exchanging the criteria to min $D_h(x_{i+1}, x_i)$ will make PEP unbounded. If $\lambda < 1/L$ then finite solution is given.
Result in [Bolte et al., 2018] is recovered:

$$\min_{i=0,\ldots,N-1} D_h(x_{i+1}, x_i) \leq \frac{\lambda(F(x_*) - F(x_0))}{N(1 - \lambda L)}$$

Another criteria min $F(x_i) - F_*$ is also bounded, but may not converge.

L-smad BPG

With symmetric Bregman distance as criteria, PEP gives the same upper bound.

$$\min_{i=0,\ldots,N-1} D_h(x_{i+1}, x_i) + D_h(x_i, x_{i+1}) \leq \frac{\lambda(F(x_*) - F(x_0))}{N(1-\lambda L)}$$

The other side the bound has slightly different bound, and $\lambda = 1/L$ is attainable.

$$\min_{i=0,\ldots,N-1} D_h(x_i, x_{i+1}) \leq \frac{\lambda(F(x_*) - F(x_0))}{N}$$

Bibliography I

Bolte, J., Sabach, S., Teboulle, M., and Vaisbourd, Y. (2018).
First order methods beyond convexity and lipschitz gradient continuity with applications to quadratic inverse problems.
ArXiv, abs/1706.06461.

Boț, R. I. and Böhm, A. (2019).
An incremental mirror descent subgradient algorithm with random sweeping and proximal step.
Optimization, 68:33 – 50.