Bregman Method from PEP Perspective

Zhenghao Xu

Zhejiang University

October 24, 2021

Table of Contents

1. BPG's PEP

Progress

Finished:

-

In progress:

-

Table of Contents

1. BPG's PEP

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.
- f is L-smooth relative to h, i.e., $Lh - f$ convex.[Bauschke et al., 2017]
- f, h more restricted.
- 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 Gradient

$$x_{k+1} = \arg\min_{x \in \mathbb{R}^d} \{ f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \phi(x) + L D_h(x, x_k) \} \tag{BPG}$$

PEP Formulation Flaw

Previous PEP formulation:

$$\max \quad f_N - f_* \tag{PEP-1}$$

$$s.t. \begin{cases} \text{convexity of } f \\ \text{convexity of } Lh - f \\ L_h\text{-smoothness } \sigma_h\text{-strongly convexity of } h \\ D_h(x_*, x_0) \le R \\ s_i - \frac{1}{L} g_i = s_{i+1} \end{cases}$$

Problem: *three* sets of interpolation conditions here.
Consistency not warranted: interpolated f, $Lh - f$ and h might not satisfy

$$f + (Lh - f) = Lh.$$

PEP Formulation Flaw

This causes a strict relaxation:
the optimal solution (one step) to (PEP-1) takes equality in expression like

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle \,.$$

but with relatively smoothness condition, f is LL_h-smooth, it should have

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2LL_h} \|\nabla f(x) - \nabla f(y)\|^2.$$

This does not hold, thus $\mathrm{val}(\text{PEP-1}) > \mathrm{val}(PEP)$.

PEP Formulation Flaw

$$
\begin{aligned}
f(x) &\geq f(y) + \langle \nabla f(y), x - y \rangle \\
f(x) &\geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2LL_h} \|\nabla f(x) - \nabla f(y)\|^2.
\end{aligned}
$$

This also gives an explanation to abnormal curves of improvement we got last time: some of the multipliers are zero in one step PEP, so corresponding *not tight* (like above) inequalities do not show up more steps are given; with steps going this non-tightness of interpolation affects less to the result.

PEP Formulation Flaw

In [Dragomir et al., 2019], they only consider the existence of convex h, so only f and $Lh - f$ are interpolated. Existence of convex h is warranted by summing up f and $Lh - f$.

This leads to two remedies:

- Derive a "relatively smooth strongly convex interpolation condition" for given h.
- Put the smooth strongly convex restrictions onto $d := Lh - f$.

Attempt 1: RSSC Interpolation

Problem of "relatively smooth strongly convex interpolation" states like this:

relatively smooth strongly convex

For $0 \leq \mu < L \leq +\infty$, f is L-smooth μ-strongly convex relative to h, denoted as $f \in \mathcal{B}_{\mu,L}(\mathbb{R}^d)$, when both $Lh - f$ and $f - \mu h$ are convex.

relatively smooth strongly convex interpolation

Given Legendre function h. Given $S = \{x_i, g_i, f_i\}_{i \in I} \subseteq \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}$. S is called L-smooth μstrongly convex interpolable relative to h if

$$\exists f \in \mathcal{B}_{\mu,L}(\mathbb{R}^d), \forall i \in I, \quad g_i \in \partial f(x_i).$$

Attempt 1: RSSC Interpolation

How is SSC ($h = \frac{1}{2}\|\cdot\|^2$) case possible?
According to [Taylor et al., 2017], a key step is to perform conjugation:
$f \in \mathcal{F}_{0,L}(\mathbb{R}^d) \iff f^* \in \mathcal{F}_{\frac{1}{L},+\infty}(\mathbb{R}^d)$.
and this leads to the condition mentioned

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{1}{2L}\|\nabla f(x) - \nabla f(y)\|_*^2.$$

This can be done because we can perform Fenchel-Young's and exchange x and $\nabla f(x)$ through ∇f and $\nabla f^* = (\nabla f)^{-1}$.

Attempt 1: RSSC Interpolation

But in Bregman case (SCC relative to h), should we use ∇h or ∇f? - this asymmetric makes the conjugate argument failed.

$$
\begin{aligned}
&\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L \langle \nabla h(x) - \nabla h(y), x - y \rangle \\
&\langle \nabla f(\nabla h^*(x)) - \nabla f(\nabla h^*(y)), \nabla h^*(u) - \nabla h^*(v) \rangle \\
&\quad \leq L \langle u - v, \nabla h^*(u) - \nabla h^*(v) \rangle .
\end{aligned}
$$

Cannot find a way to convert $Lh - f_1$ convex to $f_2 - \mu h$ convex, unless $(\nabla f \circ \nabla h^*)$ has something to do with ∇h in general.

$D_h(x, y)$ lacks important properties:

- spherical symmetric $(D_h(x, y) \not\equiv \psi(|x - y|))$
- shift invariant $(D_h(x, y) \not\equiv D_h(x - c, y - c))$

Attempt 2: Shifting SCC Condition

Shift the $h \in \mathcal{F}_{\mu,L}(\mathbb{R}^d)$ to $Lh - f \in \mathcal{F}_{\mu',L'}(\mathbb{R}^d)$.
This makes a restriction on PEP.

$$\left\{ \begin{array}{l} \exists f \in \mathcal{F}_{0,\infty} \\ \exists h \in \mathcal{F}_{\sigma_h, L_h} \\ Lh - f \in \mathcal{F}_{0,\infty} \end{array} \right. \iff \left\{ \begin{array}{l} \exists f \in \mathcal{F}_{0,LL_h} \\ \exists h \in \mathcal{F}_{\sigma_h, L_h} \\ Lh - f \in \mathcal{F}_{0,LL_h} \end{array} \right.$$

$$\overset{\text{strengthen}}{\Longleftarrow} \left\{ \begin{array}{l} \exists f \in \mathcal{F}_{0,LL_h - L\xi} \\ \exists d \in \mathcal{F}_{L\sigma_h, L\xi} \\ h := \frac{f+d}{L} \in \mathcal{F}_{\sigma_h, L_h}. \end{array} \right.$$

$0 \le \sigma_h \le \xi \le L_h \le +\infty$.

This only gives a lower bound on PEP when $\xi < \infty$.
When $\xi = \infty$, no improvement, even this lower bound.

Bibliography I

Bauschke, H. H., Bolte, J., and Teboulle, M. (2017).
A descent lemma beyond lipschitz gradient continuity: First-order methods revisited and applications.
Math. Oper. Res., 42:330–348.

Dragomir, R., Taylor, A. B., d'Aspremont, A., and Bolte, J. (2019).
Optimal complexity and certification of bregman first-order methods.
ArXiv, abs/1911.08510.

Taylor, A. B., Hendrickx, J., and Glineur, F. (2017).
Smooth strongly convex interpolation and exact worst-case performance of first-order methods.
Mathematical Programming, 161:307–345.