Bregman Method from PEP Perspective

Zhenghao Xu

Zhejiang University

September 12, 2021

Table of Contents

1 BPG from PEP

Progress

Finished:

- Derived Bound for Bregman Proximal Gradient under RSC & RS
- Table of (possible) valid improvement

In progress:

- general step size (uneasy to convert dual bilinear to linear SDP)

Table of Contents

1. BPG from 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 is σ-strongly convex relative to h, i.e., $f - \sigma h$ convex.[Lu et al., 2018]
- 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

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

Formulation

(relaxed) PEP:

$$
\max \quad F_N - F_* \tag{PEP-1}
$$

$$
s.t. \begin{cases}
\text{convexity of } Lh - f \text{ between } x_i \text{ and } x_{i-1} \\
\text{convexity of } Lh - f \text{ between } x_i \text{ and } x_{i+1} \\
\text{convexity of } Lh - f \text{ between } x_* \text{ and } x_i \\
\text{convexity of } f - \sigma h \text{ between } x_i \text{ and } x_{i+1} \\
\text{convexity of } f - \sigma h \text{ between } x_* \text{ and } x_i \\
\text{convexity of } \phi \text{ between } x_i \text{ and } x_{i+1} \\
\text{convexity of } \phi \text{ between } x_* \text{ and } x_i \\
s_i - \lambda(g_i + w_{i+1}) - s_{i+1} = 0 \\
g_* + w_* = 0
\end{cases}
$$

where $s_i \in \partial h(x_i)$, $w_i \in \partial \phi(x_i)$, $g_i = \nabla f(x_i)$.

Upper Bound

We get the upper bound

$$F(X_N) - F_* \leq \frac{\sigma}{(\frac{L}{L-\sigma})^N - 1} D_h(x_*, x_0).$$

This result was already obtained in [Lu et al., 2018]. Same rate.

Other Trials

$Lh - f$ conv., with same step size fixed: improved?

h \ f	None	SC	smooth	SC& smooth
None	F	F	F	F
SC	F	F	F	F
smooth	F	T(linear)	F	T(linear)
SC& smooth	T	T(linear)	T	T(linear)

Since "f SC"+"$Lh - f$ conv" $\implies$ "h SC", we obtain that better rate must be taken under "h strongly convex".

"h smooth"($\frac{L_h}{2}\|\cdot\|^2 - h$ conv) + "f SC"($f - \frac{\sigma_f}{2}\|\cdot\|^2$) $\implies$ $f - \frac{\sigma_f}{L_h} h$ conv, thus linear rate.

Other Trials

$Lh - f$ conv., maximal possible step size?

h \ f	None	SC	smooth	SC& smooth
None	1/L	1/L	1/L	>1/L
SC	1/L	1/L	>1/L	>1/L
smooth	1/L	-	1/L	-
SC& smooth	1/L	-	>1/L	-

Above results are not accurate.
Best step size ongoing.

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.

Lu, H., Freund, R., and Nesterov, Y. (2018).
Relatively smooth convex optimization by first-order methods, and applications.
SIAM J. Optim., 28:333–354.