Bregman Method from PEP Perspective

Zhenghao Xu

Zhejiang University

August 18, 2021

Table of Contents

Progress

Finished:

- (last time) BPG's PEP formulation (Gram matrix form)
- (last time) Numerical experiment for BPG (modified from [Dragomir et al., 2019])

Progress

Finished:

- (last time) BPG's PEP formulation (Gram matrix form)
- (last time) Numerical experiment for BPG (modified from [Dragomir et al., 2019])
- *Derived Bound for take I, II & III (nonconvex)*

Progress

Finished:

- (last time) BPG's PEP formulation (Gram matrix form)
- (last time) Numerical experiment for BPG (modified from [Dragomir et al., 2019])
- *Derived Bound for take I, II & III (nonconvex)*

In progress (with difficulties):

- General first-order Bregman (proximal) gradient formulation
- Find characteristic inequalities for pre-defined h
- Accelerated method evaluation

Table of Contents

Formulation (take I & II)

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]
- bounded initial distance and well-posedness of problem and method.

Formulation (take I & II)

Problem:

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

Method: Bregman Proximal Gradient
$\lambda \in (0, 1/L]$

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

Formulation (take I)

(relaxed) PEP (take I):

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

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

Formulation (take I)

(relaxed) PEP (take I):

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

$$s.t. \begin{cases} f_{i+1} - f_i + \langle g_{i+1}, x_i - x_{i+1} \rangle \leq 0 \\ f_i - f_* + \langle g_i, x_* - x_i \rangle \leq 0 \\ \frac{1}{\lambda} h_{i-1} - f_{i-1} - (\frac{1}{\lambda} h_i - f_i) + \left\langle \frac{1}{\lambda} s_{i-1} - g_{i-1}, x_i - x_{i-1} \right\rangle \leq 0 \\ \frac{1}{\lambda} h_{i+1} - f_{i+1} - (\frac{1}{\lambda} h_i - f_i) + \left\langle \frac{1}{\lambda} s_{i+1} - g_{i+1}, x_i - x_{i+1} \right\rangle \leq 0 \\ \frac{1}{\lambda} h_i - f_i - (\frac{1}{\lambda} h_* - f_*) + \left\langle \frac{1}{\lambda} s_i - g_i, x_* - x_i \right\rangle \leq 0 \\ \phi_{i+1} - \phi_i + \langle w_{i+1}, x_i - x_{i+1} \rangle \leq 0 \\ \phi_i - \phi_* + \langle w_i, x_* - x_i \rangle \leq 0 \\ \langle s_i - \lambda(g_i + w_{i+1}) - s_{i+1}, x_j \rangle = 0 \end{cases}$$

Dual Solution (take I)

Dual solution:

- $c^{(1)}_{N,i} = \frac{i}{N}, (i = 0, \dots, N-1)$ for $f_{i+1} - f_i + \langle g_{i+1}, x_i - x_{i+1} \rangle \leq 0$
- $c^{(2)}_{N,i} = \frac{1}{N}, (i = 0, \dots, N)$ for $f_i - f_* + \langle g_i, x_* - x_i \rangle \leq 0$
- $c^{(3)}_{N,i} = \frac{i}{N}, (i = 1, \dots, N)$ for
  $\frac{1}{\lambda} h_{i-1} - f_{i-1} - (\frac{1}{\lambda} h_i - f_i) + \left\langle \frac{1}{\lambda} s_{i-1} - g_{i-1}, x_i - x_{i-1} \right\rangle \leq 0$
- $c^{(4)}_{N,i} = \frac{i}{N}, (i = 1, \dots, N-1)$ for
  $\frac{1}{\lambda} h_{i+1} - f_{i+1} - (\frac{1}{\lambda} h_i - f_i) + \left\langle \frac{1}{\lambda} s_{i+1} - g_{i+1}, x_i - x_{i+1} \right\rangle \leq 0$
- $c^{(5)}_{N,i} = \frac{1}{N}, (i = N)$ for $\frac{1}{\lambda} h_i - f_i - (\frac{1}{\lambda} h_* - f_*) + \left\langle \frac{1}{\lambda} s_i - g_i, x_* - x_i \right\rangle \leq 0$
- $c^{(6)}_{N,i} = \frac{i}{N}, (i = 0, \dots, N-1)$ for $\phi_{i+1} - \phi_i + \langle w_{i+1}, x_i - x_{i+1} \rangle \leq 0$
- $c^{(7)}_{N,i} = \frac{1}{N}, (i = 1, \dots, N)$ for $\phi_i - \phi_* + \langle w_i, x_* - x_i \rangle \leq 0$

Upper Bound (take I)

Summing up inequalities with dual variables as weights.
Apply equality constraints to eliminate terms.
With the aid of computer algebra system (CAS), we get the upper bound

$$F(X_N) - F_* \leq \frac{D_h(x_*, x_0)}{\lambda N}$$

Formulation (take II)

(relaxed) PEP (take II):

$$\max \quad \min_{i=1,\dots,N} D_h(x_{i-1}, x_i) \qquad \text{(PEP-2)}$$

$$s.t. \begin{cases} \textit{convexity of } f \textit{ between } x_* \textit{ and } x_i \\ \textit{convexity of } \frac{1}{\lambda}h - f \textit{ between } x_i \textit{ and } x_{i-1} \\ \textit{convexity of } \frac{1}{\lambda}h - f \textit{ between } x_* \textit{ and } x_i \\ \textit{convexity of } \phi \textit{ between } x_i \textit{ and } x_{i+1} \\ \textit{convexity of } \phi \textit{ between } x_* \textit{ and } x_i \\ F_* - F_i \leq 0 \\ s_i - \lambda(g_i + w_{i+1}) - s_{i+1} = 0 \end{cases}$$

Formulation (take II)

(relaxed) PEP (take II):

$$
\begin{aligned}
\max \quad & d \\
s.t. \quad & \begin{cases}
\textit{convexity of } f \textit{ between } x_* \textit{ and } x_i \\
\textit{convexity of } \frac{1}{\lambda}h - f \textit{ between } x_i \textit{ and } x_{i-1} \\
\textit{convexity of } \frac{1}{\lambda}h - f \textit{ between } x_* \textit{ and } x_i \\
\textit{convexity of } \phi \textit{ between } x_i \textit{ and } x_{i+1} \\
\textit{convexity of } \phi \textit{ between } x_* \textit{ and } x_i \\
F_* - F_i \leq 0 \\
d - D_h(x_{i-1}, x_i) \leq 0 \\
s_i - \lambda(g_i + w_{i+1}) - s_{i+1} = 0
\end{cases}
\end{aligned}
\tag{PEP-2}
$$

Dual Solution (take II)

Dual solution:

- $c^{(1)}_{N,i} = \begin{cases} \frac{2\lambda}{N(N-1)}, & i = 0, \ldots, N-1 \\ K_1, & i = N \end{cases}$ for conv. of f between x_* and x_i
- $c^{(2)}_{N,i} = \begin{cases} \frac{2\lambda i}{N(N-1)}, & i = 1, \ldots, N-1 \\ K_1 + \frac{2\lambda}{N}, & i = N \end{cases}$ for conv. of $\frac{1}{\lambda}h - f$ between x_i and x_{i-1}
- $c^{(3)}_{N,i} = \begin{cases} \frac{2\lambda}{N-1} - K1, & i = N-1 \\ K_1 + \frac{2\lambda}{N}, & i = N \end{cases}$ for conv. of $\frac{1}{\lambda}h - f$ between x_* and x_i
- $c^{(4)}_{N,i} = \frac{2\lambda i}{N(N-1)}, i = 0, \ldots, N-1$ for conv. of ϕ between x_i and x_{i+1}
- $c^{(5)}_{N,i} = \begin{cases} \frac{2\lambda}{N(N-1)}, & i = 1, \ldots, N-1 \\ K_1, & i = N \end{cases}$ for conv. of ϕ between x_* and x_i

Dual Solution (take II)

Dual solution (continue):

- $c^{(6)}_{N,N} = \frac{2\lambda}{N} + K_1$ for $F_* - F_N \leq 0$
- $c^{(7)}_{N,i} = \frac{2i}{N(N-1)}, (i = 0, \ldots, N-1)$ for $d - D_h(x_i, x_{i+1}) \leq 0$

Upper Bound (take II)

Summing up inequalities with dual variables as weights.
Apply equality constraints to eliminate terms.
With the aid of CAS, we get the upper bound

$$\begin{aligned}\min_{i=1,\dots,N} D_h(x_{i-1}, x_i) &\leq \frac{2D_h(x_*, x_0)}{N(N-1)} - \frac{2D_{h-\lambda f}(x_*, x_{N-1})}{N} \\ &\leq \frac{2D_h(x_*, x_0)}{N(N-1)}.\end{aligned}$$

Table of Contents

Formulation (take III)

Problem:

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

Assumptions:

- f is *proper*, *lsc* and continuously differentiable (possibly nonconvex).
- ϕ is *convex*, *proper*, *lsc* (possibly nonsmooth).
- h is a Legendre kernel function.
- f is L-smooth adaptable (L-smad) relative to h [Bolte et al., 2018], i.e., both $Lh - f$ and $Lh + f$ are convex.
- bounded initial distance and well-posedness of problem and method.

Formulation (take III)

(relaxed) PEP (take III):

$$
\begin{aligned}
\max \quad & d \\
s.t. \quad & \begin{cases}
\textit{convexity of } \frac{1}{\lambda}h + f \textit{ between } x_* \textit{ and } x_i \\
\textit{convexity of } \frac{1}{\lambda}h - f \textit{ between } x_{i+1} \textit{ and } x_i \\
\textit{convexity of } \frac{1}{\lambda}h - f \textit{ between } x_* \textit{ and } x_i \\
\textit{convexity of } \phi \textit{ between } x_i \textit{ and } x_{i+1} \\
\textit{convexity of } \phi \textit{ between } x_* \textit{ and } x_i \\
F_* - F_i \leq 0 \\
d - D_h(x_{i-1}, x_i) \leq 0 \\
s_i - \lambda(g_i + w_{i+1}) - s_{i+1} = 0
\end{cases}
\end{aligned}
\tag{PEP-3}
$$

Dual Solution (take III)

Dual solution:

- $c^{(1)}_{N,i} = \begin{cases} \frac{\lambda}{2^i}, & i = 0, \ldots, N-1 \\ \frac{\lambda}{2^{N+1}}, & i = N \end{cases}$ for conv. of $\frac{1}{\lambda}h - f$ between x_* and x_i
- $c^{(2)}_{N,i} = \begin{cases} \lambda(2 - \frac{1}{2^i}), & i = 0, \ldots, N-2 \\ \lambda(2 - \frac{1}{2^N} - \frac{1}{2^{N-1}}), & i = N \end{cases}$ for conv. of $\frac{1}{\lambda}h - f$ between x_{i+1} and x_i
- $c^{(3)}_{N,i} = \frac{\lambda}{2^{i+1}}, (i = N-1, N)$ for conv. of $\frac{1}{\lambda}h - f$ between x_* and x_i
- $c^{(4)}_{N,i} = \lambda(2 - \frac{1}{2^{i-1}}), (i = 1, \ldots, N-1)$ for conv. of ϕ between x_i and x_{i+1}
- $c^{(5)}_{N,i} = \begin{cases} \frac{\lambda}{2^{i-1}}, & i = 1, \ldots, N-1 \\ \frac{\lambda}{2^N}, & i = N \end{cases}$ for conv. of ϕ between x_* and x_i

Dual Solution (take III)

Dual solution (continue):

- $c_{N,i}^{(6)} = \lambda(2 - \frac{1}{2^N} - \frac{1}{2^{N-1}})$ for $F_* - F_N \leq 0$
- $c_{N,i}^{(7)} = 2 - \frac{1}{2^{i-1}}, (i = 1, \ldots, N-1)$ for $d - D_h(x_i, x_{i+1}) \leq 0$

Upper Bound (take III)

Summing up inequalities with dual variables as weights.
Apply equality constraints to eliminate terms.
With the aid of CAS, we get the upper bound

$$\min_{i=1,\dots,N} D_h(x_{i-1}, x_i) \le \frac{D_h(x_*, x_0)}{N - 2 - \frac{1}{2^{N-1}}}.$$

[Teboulle, 2018] has proved for $\min_{i=1,\dots,N} \|x_{i-1}, x_i\|^2 \le O(\frac{1}{\sigma N}\kappa(\lambda, L))$ for σ-strongly convex h.

Table of Contents

Difficulties

- General form might need three sequences.
- Common h, e.g., entropy is not well for acceleration as well. (triangle scaling gain not bound in general) [Hanzely et al., 2021]

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.

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.

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.

Bibliography II

Hanzely, F., Richtárik, P., and Xiao, L. (2021).
Accelerated bregman proximal gradient methods for relatively smooth convex optimization.
Comput. Optim. Appl., 79:405–440.

Teboulle, M. (2018).
A simplified view of first order methods for optimization.
Mathematical Programming, 170:67–96.