Bregman Method from PEP Perspective

Zhenghao Xu

Zhejiang University

August 22, 2021

Table of Contents

1 BPP from PEP

2 In Progress

Progress

Finished:

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

Progress

Finished:

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

In progress:

- Bregman Halpern iteration (numerically unbounded, lack of resources)

Progress

Finished:

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

In progress:

- Bregman Halpern iteration (numerically unbounded, lack of resources)
- general step size λ_i for BPP's PEP, applications[Yang and Toh, 2021]

Progress

Finished:

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

In progress:

- Bregman Halpern iteration (numerically unbounded, lack of resources)
- general step size λ_i for BPP's PEP, applications[Yang and Toh, 2021]
- complete proof (slow with LaTeX)

Table of Contents

1 BPP from PEP

2 In Progress

Formulation

Problem:

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

Assumptions:

- ϕ is *convex*, *proper*, *lsc* (possibly nonsmooth).
- h is a Legendre kernel function.
- bounded initial distance and well-posedness of problem and method.

Formulation

Problem:

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

Method: Bregman Proximal Point (fixed step size) $\lambda \in (0, \infty)$

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

Formulation

(relaxed) PEP:

$$\max \quad \phi_N - \phi_* \tag{PEP-1}$$
$$s.t. \begin{cases} \text{convexity of } h \text{ between } x_i \text{ and } x_{i-1} \\ \text{convexity of } h \text{ between } x_i \text{ and } x_{i+1} \\ \text{convexity of } h \text{ between } x_* \text{ and } x_i \\ \text{convexity of } \phi \text{ between } x_i \text{ and } x_{i+1} \\ \text{convexity of } \phi \text{ between } x_* \text{ and } x_i \\ s_i - \lambda w_{i+1} - s_{i+1} = 0 \end{cases}$$

where $s_i \in \partial h(x_i)$, $w_i \in \partial \phi(x_i)$.

Formulation

(relaxed) PEP:

$$\max \quad \phi_N - \phi_* \tag{PEP-1}$$

$$s.t. \begin{cases} h_{i-1} - h_i + \langle s_{i-1}, x_i - x_{i-1} \rangle \leq 0 \\ h_{i+1} - h_i + \langle s_{i+1}, x_i - x_{i+1} \rangle \leq 0 \\ h_i - h_* + \langle s_i, x_* - x_i \rangle \leq 0 \\ \phi_{i+1} - \phi_i + \langle w_{i+1}, x_i - x_{i+1} \rangle \leq 0 \\ \phi_i - \phi_* + \langle w_i, x_* - x_i \rangle \leq 0 \\ \langle s_i - \lambda w_{i+1} - s_{i+1}, x_j \rangle = 0 \end{cases}$$

where $s_i \in \partial h(x_i)$, $w_i \in \partial \phi(x_i)$.

Formulation

(relaxed) PEP:

$$\max \quad \phi_N - \phi_* \qquad \text{(sdp-PEP-1)}$$

$$s.t. \begin{cases} h_{i-1} - h_i + G^{s,x}_{i-1,i} - G^{s,x}_{i-1,i-1} \le 0 \\ h_{i+1} - h_i + G^{s,x}_{i+1,i} - G^{s,x}_{i+1,i+1} \le 0 \\ h_i - h_* + G^{s,x}_{i,*} - G^{s,x}_{i-1,i-1} \le 0 \\ \phi_{i+1} - \phi_i + G^{w,x}_{i+1,i} - G^{w,x}_{i+1,i+1} \le 0 \\ \phi_i - \phi_* + G^{w,x}_{i,*} - G^{w,x}_{i,i} \le 0 \\ \langle s_i - \lambda w_{i+1} - s_{i+1}, x_j \rangle = 0 \end{cases}$$

(assume $x_* = \phi_* = h_* = 0$)
$P = \big(P_x, P_s, P_w\big) = \big(x_0, \ldots, x_N, s_*, s_0, \ldots, s_N, w_*, w_0, \ldots, w_N\big)$,
$G = P^T P$, $G^{s,x} = P_s^T P_x$, $G^{w,x} = P_w^T P_x$.

Dual Solution

Dual solution:

- $c_{N,i}^{(1)} = \frac{i}{\lambda N}, (i = 1, \ldots, N)$ for $h_{i-1} - h_i + \langle s_{i-1}, x_i - x_{i-1} \rangle \leq 0$
- $c_{N,i}^{(2)} = \frac{i}{\lambda N}, (i = 1, \ldots, N-1)$ for $h_{i+1} - h_i + \langle s_{i+1}, x_i - x_{i+1} \rangle \leq 0$
- $c_{N,N}^{(3)} = \frac{1}{\lambda N}$, for $h_N - h_* + \langle s_N, x_* - x_N \rangle \leq 0$
- $c_{N,i}^{(4)} = \frac{i}{N}, (i = 0, \ldots, N-1)$ for $\phi_{i+1} - \phi_i + \langle w_{i+1}, x_i - x_{i+1} \rangle \leq 0$
- $c_{N,i}^{(5)} = \frac{1}{N}, (i = 1, \ldots, N)$ for $\phi_i - \phi_* + \langle w_i, x_* - x_i \rangle \leq 0$

Upper Bound

Summing up inequalities with dual variables as weights.

$$
\begin{aligned}
&\sum_{i=1}^{N} \frac{i}{\lambda N} * \left(h_{i-1} - h_i + \langle s_{i-1}, x_i - x_{i-1} \rangle\right) \\
&+ \sum_{i=1}^{N-1} \frac{i}{\lambda N} * \left(h_{i+1} - h_i + \langle s_{i+1}, x_i - x_{i+1} \rangle\right) \\
&+ \frac{1}{\lambda N} * \left(h_N - h_* + \langle s_N, x_* - x_N \rangle\right) \\
&+ \sum_{i=0}^{N-1} \frac{i}{N} * \left(\phi_{i+1} - \phi_i + \langle w_{i+1}, x_i - x_{i+1} \rangle\right) \\
&+ \sum_{i=1}^{N} \frac{1}{N} * \left(\phi_i - \phi_* + \langle w_i, x_* - x_i \rangle\right) \leq 0.
\end{aligned}
$$

Upper Bound

We get the upper bound

$$\phi(X_N) - \phi_* \leq \frac{D_h(x_*, x_0)}{\lambda N}.$$

For flexible step size λ_i at each step, under some assumptions we have $O\left(\frac{D_h(x_*, x_0)}{\sum_{i=0}^{N-1} \lambda_i}\right)$ for inexact BPP algorithm (iBPPA)[Yang and Toh, 2021]

Table of Contents

1 BPP from PEP

2 In Progress

Bregman Halpern

Bregman Halpern iteration:

$$x_{k+1} = \nabla h^*[\lambda_k \nabla h(x_0) + (1-\lambda_k)\nabla h(A(x_k))]$$

Bregman resolvent $R_T^h = (\nabla h + \alpha T)^{-1} \circ \nabla h$
$y = R_{\partial\phi}^h(x)$ iff $y \in \arg\min\{\phi(z) + \frac{1}{\alpha} D_h(z,x)\}$

Let $A = R_{\partial\phi}^h$, let $y_k = A(x_k)$, we have

$$\begin{cases} y_k = A(x_k) = (\nabla h + \alpha\partial\phi)^{-1} \circ \nabla h \\ x_{k+1} = \nabla h^*[\lambda_k \nabla h(x_0) + (1-\lambda_k)\nabla h(y_k)] \end{cases}$$

Bregman Halpern

$$\begin{cases} y_k = A(x_k) = (\nabla h + \alpha\partial\phi)^{-1} \circ \nabla h \\ x_{k+1} = \nabla h^*[\lambda_k \nabla h(x_0) + (1-\lambda_k)\nabla h(y_k)] \end{cases}$$

$$\Longrightarrow \begin{cases} \nabla h(y_k) + \alpha\partial\phi(y_k) = \nabla h(x_k) \\ \nabla h(x_{k+1}) = \lambda_k \nabla h(x_0) + (1-\lambda_k)\nabla h(y_k) \end{cases}$$

$$\Longrightarrow \begin{cases} \nabla h(y_0) + \alpha\partial\phi(y_0) = \nabla h(x_0) \\ \nabla h(y_k) + \alpha\partial\phi(y_k) = \lambda_{k-1} \nabla h(x_0) + (1-\lambda_{k-1})\nabla h(y_{k-1}) \end{cases}$$

$$\Longrightarrow \begin{cases} s_0 + \alpha w_0 = \nabla h(x_0) \\ s_k + \alpha w_k = \lambda_{k-1} \nabla h(x_0) + (1-\lambda_{k-1}) s_{k-1} \end{cases}$$

$$\Longrightarrow s_k + \alpha w_k = \lambda_{k-1}(s_0 + \alpha w_0) + (1-\lambda_{k-1}) s_{k-1} \quad (k \geq 1)$$

Bregman Halpern

$$s_k + \alpha w_k = \lambda_{k-1}(s_0 + \alpha w_0) + (1 - \lambda_{k-1})s_{k-1} \quad (k \geq 1)$$

Try to plug this equality constraint into the PEP.

But numerically unbounded.
Cannot decide if it is because of the lack of constraints in the program, or this method has no bound for fixed step N.

Others

Try to formulate general step size λ_i PEP for Bregman Proximal Point.

Complete the full proof for BPG convergence bound.

Bibliography I

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.

Yang, L. and Toh, K. (2021).
Bregman proximal point algorithm revisited: A new inexact version and its variant.