Bregman Method from PEP Perspective

Zhenghao Xu

Zhejiang University

September 26, 2021

Table of Contents

1. Performance
2. Step Size

Progress

Finished:

- evidence of improvement (numerical)

In progress:

- optimal fixed step size (dual matrix)
- valid one-step analysis
- toolbox for explicit dual matrix (in Python)

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

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

Table of Contents

1 Performance

2 Step Size

Performance

$Lh - f$ conv., with same step size fixed: $\lambda = 1/L$.

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)

case 1: h: L_h-smooth and σ_h-strongly convex.

Performance

$Lh - f$ conv., with same step size fixed: $\lambda = 1/L$.

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)*

case 1: h: L_h-smooth and σ_h-strongly convex.
case 2: h: L_h-smooth; f: σ_f-strongly convex -> relatively SC.

Performance

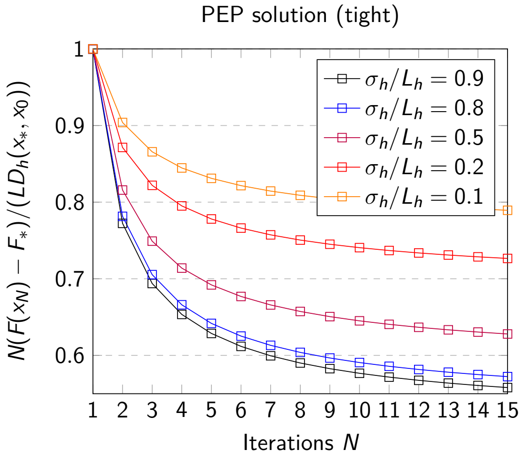
$Lh - f$ conv., with same step size fixed: $\lambda = 1/L$.

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)

case 1: h: L_h-smooth and σ_h-strongly convex.
case 2: h: L_h-smooth; f: σ_f-strongly convex -> relatively SC.
case 3: h: L_h-smooth and σ_h-strongly convex; f: L_f-smooth -> replace $Lh - f$ conv. when $L > \frac{L_f}{\sigma_h}$.

Case 1: h - SC & Smooth

Assume $\lambda = 1/L$. (L arbitrary)



Numerical result:

$$F(x_N) - f(x_*) \leq \frac{LD_N(x_*, x_0)}{N}\left(\frac{c}{N} + 1 - c\right)$$

No simple multipliers.
No obvious potential function.

Case 1: h - SC & Smooth

Try to solve PEP for $N = 1$. with some constraints on dual variables we get ($0 \leq b \leq 1$, $0 \leq c \leq L$):

$$
\begin{aligned}
& F(x_1) - F_* + \frac{(L-c)\left((s_1 - s_0)^2 + L_h \sigma_h (x_1 - x_0)^2\right)}{2(L_h - \sigma_h)} \\
& + \frac{c\left((s_1 - s_*)^2 + (x_1 - x_*)^2 L_h \sigma_h\right)}{2(L_h - \sigma_h)} - \frac{c\sigma_h}{L_h - \sigma_h} D_h(x_1, x_*) \\
& - \frac{(L-c)\sigma_h}{L_h - \sigma_h} D_h(x_1, x_0) - \left(\frac{(L-c)\sigma_h}{L_h - \sigma_h} - c + bL\right) D_h(x_0, x_1) \\
& + \left(\frac{-cL_h}{L_h - \sigma_h} + bL\right) D_h(x_*, x_1) + (c - bL) D_h(x_*, x_0) \leq 0.
\end{aligned}
$$

Case 1: h - SC & Smooth

Try to solve PEP for $N = 1$. with some constraints on dual variables we get ($b = c/L$, $0 \leq c \leq L$):

$$
\begin{aligned}
&F(x_1) - F_* + \frac{(L-c)\left((s_1 - s_0)^2 + L_h\sigma_h(x_1 - x_0)^2\right)}{2(L_h - \sigma_h)} \\
&+ \frac{c\left((s_1 - s_*)^2 + (x_1 - x_*)^2 L_h\sigma_h\right)}{2(L_h - \sigma_h)} - \frac{c\sigma_h}{L_h - \sigma_h} D_h(x_1, x_*) \\
&- \frac{(L-c)\sigma_h}{L_h - \sigma_h} D_h(x_1, x_0) - \frac{(L-c)\sigma_h}{L_h - \sigma_h} D_h(x_0, x_1) \\
&- \frac{c\sigma_h}{L_h - \sigma_h} D_h(x_*, x_1) \leq 0.
\end{aligned}
$$

Case 1: h - SC & Smooth

Try to solve PEP for $N = 1$. with some constraints on dual variables we get ($b = c/L$, $0 \leq c \leq L$):

$$
\begin{aligned}
&F(x_1) - F_* + \frac{(L-c)\left((s_1 - s_0)^2 - L_h\sigma_h(x_1 - x_0)^2\right)}{2(L_h - \sigma_h)} \\
&+ \frac{c\left((s_1 - s_*)^2 - (x_1 - x_*)^2 L_h\sigma_h\right)}{2(L_h - \sigma_h)} \leq 0.
\end{aligned} \tag{form-1}
$$

(Apply $\langle s_i - s_j, x_i - x_j \rangle \leq L_h \|x_i - x_j\|^2$.)

Cannot telescope?

Case 1: h - SC & Smooth

Try to solve PEP for $N = 1$. with some constraints on dual variables we get ($b = c/L$, $0 \leq c \leq L$):

$$
\begin{aligned}
&F(x_1) - F_* + \frac{(L-c)(s_1 - s_0)^2}{2(L_h - \sigma_h)} + \frac{c(s_1 - s_*)^2}{2(L_h - \sigma_h)} \\
\leq &\frac{c\sigma_h}{2(L_h - \sigma_h)} \left(D_h(x_1, x_*) + D_h(x_*, x_1) \right) \qquad \text{(form-2)} \\
&+ \frac{(L-c)\sigma_h}{2(L_h - \sigma_h)} \left(D_h(x_1, x_0) + D_h(x_0, x_1) \right).
\end{aligned}
$$

(Apply $\langle s_i - s_j, x_i - x_j \rangle \leq L_h \|x_i - x_j\|^2$.)

Only use half.

Case 1: h - SC & Smooth

Let $b = c/L$, $c = L/2$:

$$F(x_1) - F_* + \frac{L\left((s_1 - s_0)^2 + (s_1 - s_*)^2\right)}{4(L_h - \sigma_h)}$$
$$\leq \frac{L\sigma_h}{4(L_h - \sigma_h)}(D_h(x_0, x_*) + D_h(x_*, x_0)$$
$$+ \langle s_1 - s_0, x_1 - x_* \rangle + \langle s_1 - s_*, x_1 - x_0 \rangle)$$
$$\implies$$
$$F(x_1) - F_* + \frac{L\left(\|s_1 - s_0 - \frac{\sigma_h}{2}(x_1 - x_*)\|^2 + \|s_1 - s_* - \frac{\sigma_h}{2}(x_* - x_0)\|^2\right)}{4(L_h - \sigma_h)}$$
$$\leq \frac{L\sigma_h}{4(L_h - \sigma_h)}(D_h(x_0, x_*) + D_h(x_*, x_0) + \langle s_1 - s_*, x_1 - x_* \rangle +$$
$$\frac{\sigma_h}{4}(\|x_1 - x_*\|^2 + \|x_* - x_0\|^2)) \tag{form-2}$$

Case 1: h - SC & Smooth

Let $b = c/L$, $c = L/2$:

$$\begin{aligned} F(x_1) - F_* \leq & \frac{L\sigma_h}{4(L_h - \sigma_h)}(D_h(x_0, x_*) + D_h(x_*, x_0) \\ & + L_h\|x_1 - x_*\|^2 + \frac{\sigma_h}{4}(\|x_1 - x_*\|^2 + \|x_0 - x_*\|^2)) \end{aligned} \tag{form-2}$$

Still not fully clear.

Other criteria does not give clear potential function.

Table of Contents

1. Performance

2. Step Size

Step Size

$Lh - f$ conv.

f \ h	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	-

case 1: h: σ_h-strongly convex; f: L_f-smooth -> replace $Lh - f$ conv. when $L > \frac{L_f}{\sigma_h}$.

Step Size

$Lh - f$ conv.

f \ h	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	-

case 1: h: σ_h-strongly convex; f: L_f-smooth -> replace $Lh - f$ conv. when $L > \frac{L_f}{\sigma_h}$.
case 2: f: L_f-smooth and σ_f-strongly convex. -> replace $Lh - f$ conv. when $L > \frac{LL_f}{\sigma_f}$

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.