Notes on Bregman Iteration

Pascal Getreuer

2009-02-05 (updated 2016-12-07)

Abstract

These notes discuss the Bregman iteration technique for constrained minimization. We review the basic theoretical results (with proofs), in particular, the monotonic convergence of the constraint. Second, we discuss several Bregman-related algorithms.

1 Bregman Iteration

Bregman iteration [1] is a technique for solving constrained minimizations of the form

$$\arg\min_u J(u) \text{ subject to } H(u) = 0$$

or with an inequality constaint

$$\arg\min_u J(u) \text{ subject to } H(u) < \sigma$$

where J and H are convex functions with $\min H(u) = 0$ defined on a Hilbert space.

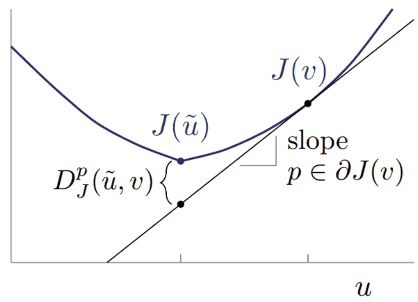
Notations We denote inner product by $\langle u, v\rangle$. The subdifferential of convex function J at v is $\partial J(v) := \big\{p : J(u) \geq J(v) + \langle p, u - v\rangle\,,\ \forall u\big\}$. If J is differentiable at v, then p is the usual gradient, $p = \nabla J(v)$.

Definition 1. The *Bregman distance* of J between u and v is

$$D_J^p(u, v) = J(u) - J(v) - \langle p, u - v\rangle\,. \quad p \in \partial J(v),$$

That is, $D_J^p(u, v)$ is the difference between $J(u)$ and the tangent plane $J(v) + \langle p, u - v\rangle$.

Here is an illustration in one dimension, where the horizontal axis is the u variable:



The Bregman distance $D_J^p(\tilde{u}, v)$.

The geometric meaning is that Bregman distance $D_J^p(u, v)$ compares $J(u)$ with the tangent plane $J(v) + \langle p, u - v\rangle$. Suppose that u is a minimizer of J, then Bregman distance compares the "closeness" of u and v in the sense that $D_J^p(u, v) \geq 0$ and that $D_J^p(u, v) \geq D_J^p(w, v)$ for all points w on the line segment between u and v [3].

Result 1. Bregman distance satisfies

(a) $D_J^p(v, v) = 0$

(b) $D_J^p(u, v) \geq 0$

(c) $D_J^p(u, v) + D_J^{\tilde{p}}(v, \tilde{v}) - D_J^{\tilde{p}}(u, \tilde{v}) = \langle p - \tilde{p}, v - u\rangle$

Proof. Results (a) and (c) follow immediately from Definition 1 and the convexity of J. Result (b) follows by definition of subgradient. □

Definition 2. Given parameter $\lambda > 0$, *Bregman iteration* is

$$u_{k+1} = \arg\min_u D_J^{p_k}(u, u_k) + \lambda H(u), \quad p_k \in \partial J(u_k). \tag{1}$$

or equivalently

$$u_{k+1} = \arg\min_u J(u) - \langle p, u - u_k\rangle + \lambda H(u), \quad p_k \in \partial J(u_k). \tag{2}$$

Each iteration in this process can be seen as minimizing J while approximately imposing the constraint $H(u) \approx 0$, and then adding a linear term to the objective to weaken it so that the $\lambda H(u)$ penalty term is more influential on the next iteration.

Result 2. Suppose that $u^\star$ is a minimizer of H, then Bregman iteration (1) decreases $H(u_k)$ monotonically with

$$H(u_{k+1}) \leq H(u_{k+1}) + \tfrac{1}{\lambda} D_J^{p_k}(u_{k+1}, u_k) \leq H(u_k).$$

Proof. Since u_{k+1} minimizes $D_J^{p_k}(u, u_k) + \lambda H(u)$,

$$\begin{aligned}\lambda H(u_{k+1}) &\leq D_J^{p_k}(u_{k+1}, u_k) + \lambda H(u_{k+1}) \\ &\leq D_J^{p_k}(u_k, u_k) + \lambda H(u_k) \qquad = \lambda H(u_k).\end{aligned}$$

□

Result 3. Suppose H is differentiable. Then $p_k - \lambda \nabla H(u_{k+1}) \in \partial J(u_{k+1})$ and a special case of Bregman iteration is

$$\begin{cases} u_{k+1} = \arg\min_u D_J^{p_k}(u, u_k) + \lambda H(u) \\ p_{k+1} = p_k - \lambda \nabla H(u_{k+1}). \end{cases} \tag{3}$$

Furthermore, let $u^\star$ be a minimizer of H, then iteration (3) satisfies

$$H(u_k) \leq H(u^\star) + \frac{D_J^{p_0}(u^\star, u_0)}{\lambda k}.$$

Proof. Since u_{k+1} minimizes $D_J^{p_k}(u, u_k) + \lambda H(u)$, it satisfies Euler-Lagrange equation

$$\begin{aligned} 0 &\in \partial_u \big[D_J^{p_k}(u, u_k) + \lambda H(u)\big](u_{k+1}) \\ 0 &\in \partial J(u_{k+1}) - p_k + \lambda \nabla H(u_{k+1}) \\ p_k - \lambda \nabla H(u_{k+1}) &\in \partial J(u_{k+1}). \end{aligned}$$

From Result 1(c), iteration (3) satisfies

$$\begin{aligned} D_J^{p_k}(u^\star, u_k) - D_J^{p_{k-1}}(u^\star, u_{k-1}) &\leq D_J^{p_k}(u^\star, u_k) + D_J^{p_{k-1}}(u_k, u_{k-1}) - D_J^{p_{k-1}}(u^\star, u_{k-1}) \\ &= \langle p_k - p_{k-1}, u_k - u^\star \rangle \\ &= \langle \lambda \nabla H(u_k), u^\star - u_k \rangle \\ &\leq \lambda\big(H(u^\star) - H(u_k)\big), \end{aligned}$$

where the last inequality follows by the convexity of H. Summing over $k = 1, \ldots K$ and applying Result 2,

$$\begin{aligned} \sum_{k=1}^{K} D_J^{p_k}(u^\star, u_k) - D_J^{p_{k-1}}(u^\star, u_{k-1}) + \lambda\big(H(u_k) - H(u^\star)\big) &\leq 0 \\ D_J^{p_K}(u^\star, u_K) - D_J^{p_0}(u^\star, u_0) + \lambda K\big(H(u_K) - H(u^\star)\big) &\leq 0 \\ -D_J^{p_0}(u^\star, u_0) + \lambda K\big(H(u_K) - H(u^\star)\big) &\leq 0. \end{aligned}$$

□

See [2] for more details and other results.

2 Bregman Algorithms

Since iteration (3) guarantees convergence (Result 3), Bregman iteration is usually applied in this form. Thus, to find a minimizer of

$$\min_u J(u) \text{ subject to } H(u) = 0,$$

the Bregman algorithm is

$p \leftarrow 0$

while "not converged"

$u \leftarrow \arg\min_u J(u) - \langle p, u\rangle + \lambda H(u)$

$p \leftarrow p - \lambda \nabla H(u)$

where u is initialized for example as a minimizer of H. The computational performance of Bregman iteration depends on how quickly the subproblems $\arg\min_u J(u) - \langle p, u\rangle + \lambda H(u)$ can be solved.

Result 4 ("Adding back the noise")**.** Suppose $H(u, f) = \frac{1}{2}\|Au - f\|_2^2$ where A is a linear operator. Then iteration (3) with $p_0 = 0$ is equivalent to

$$\begin{cases} u_{k+1} = \arg\min_u J(u) + \lambda H(u, f_k) \\ f_{k+1} = f_k + (f - Au_{k+1}), \quad f_0 = f \end{cases} \tag{4}$$

Proof. On iteration K,

$$\begin{aligned} \lambda H(u,f) - \langle p_K, u\rangle &= \lambda H(u,f) - \big\langle p_0 - \lambda \textstyle\sum_{k=1}^K \nabla_u H(u_k, f), u\big\rangle \\ &= \tfrac{\lambda}{2}\|Au\|_2^2 + \tfrac{\lambda}{2}\|f\|_2^2 - \lambda\langle f, Au\rangle + \lambda\big\langle \textstyle\sum_{k=1}^K A^*(Au_k - f), u\big\rangle \\ &= \tfrac{\lambda}{2}\|Au\|_2^2 + \tfrac{\lambda}{2}\|f\|_2^2 - \lambda\big\langle f + \textstyle\sum_{k=1}^K (f - Au_k), Au\big\rangle \\ &= \tfrac{\lambda}{2}\|Au - f_K\| + \mathit{constant}. \end{aligned}$$

□

2.1 Linearized Bregman

Suppose that $J(u)$ is separable (particularly, $J(u) = \|u\|_1$). Linearized Bregman [3] approximates H in the subproblem by

$$H(u) \approx H(u_k) + \nabla H(u_k) \cdot (u - u_k).$$

The approximation is accurate only for u near u_k, so a penalty term $\frac{1}{2\delta}\|u - u_k\|_2^2$ is added.

$$u_{k+1} = \arg\min_u J(u) + \langle \lambda \nabla H(u_k) - p_k, u\rangle + \tfrac{1}{2\delta}\|u - u_k\|_2^2.$$

The linearized subproblem is separable, so it can be solved efficiently. If $J(u) = \|u\|_1$, the solution may be expressed in closed-form as a shrinkage.

2.2 Split Bregman

Suppose E is convex and Φ is convex and differentiable. Split Bregman [4] solves the problem

$$\min_u \|\Phi(u)\|_1 + E(u)$$

by the operator splitting

$$\min_{u,d} \|d\|_1 + E(u) \text{ subject to } \Phi(u) = d.$$

Applying Bregman iteration with $J(u,d) = \|d\|_1 + E(u)$ and $H(u,d) = \frac{1}{2}\|d - \Phi(u)\|_2^2$ yields

$$\begin{cases} (u^{k+1}, d^{k+1}) = \min_{u,d} J(u,d) - \langle p_u^k, u - u^k \rangle - \langle p_d^k, d - d^k \rangle + \lambda H(u,d) \\ \qquad p_u^{k+1} = p_u^k - \lambda \nabla_u H(u^{k+1}, d^{k+1}) \\ \qquad p_d^{k+1} = p_d^k - \lambda \nabla_d H(u^{k+1}, d^{k+1}) \end{cases}$$

where $\nabla_u H(u,d) = \big(\nabla\Phi(u)\big)^* \big(\Phi(u) - d\big)$ and $\nabla_d H(u,d) = d - \Phi(u)$. Analogous to Result 4, Split Bregman is equivalently

$$\begin{cases} (u_{k+1}, d_{k+1}) = \min_{u,d} \|u\|_1 + E(u) + \frac{\lambda}{2}\|d - \Phi(u) - b_k\|_2^2 \\ \qquad b_{k+1} = b_k + \big(\Phi(u^{k+1}) - d^{k+1}\big) \end{cases} \tag{5}$$

The subproblems may then be solved by alternatingly minimizing u and then d,

$$\begin{cases} u_{j+1} = \arg\min_u E(u) + \frac{\lambda}{2}\|d_j - \Phi(u) - b_k\|_2^2 \\ d_{j+1} = \arg\min_d \|d\|_1 + \frac{\lambda}{2}\|d - \Phi(u_{j+1}) - b_k\|_2^2 \end{cases}$$

In the step for d_{j+1}, the minimizer can be expressed in closed-form as a shrinkage.

References

[1] L. Bregman, "The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming," USSR Computational Mathematics and Mathematical Physics, 7, p. 200–217, 1967.

[2] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin, "An Iterative Regularization Method for Total Variation Based Image Restoration," UCLA CAM Report 04–13, 2004.
ftp://ftp.math.ucla.edu/pub/camreport/cam04-13.pdf

[3] Wotao Yin, Stanley Osher, Donald Goldfarb, and Jerome Darbon, "Bregman Iterative Algorithms for ℓ_1-Minimization with Applications to Compressed Sensing," UCLA CAM Report 07–37, 2007.
ftp://ftp.math.ucla.edu/pub/camreport/cam07-37.pdf

[4] Tom Goldstein and Stanley Osher, "The Split Bregman Method for L1 Regularized Problems," UCLA CAM Report 08–29, 2008.
ftp://ftp.math.ucla.edu/pub/camreport/cam08-29.pdf