MA3111: Mathematical Image Processing
Split Bregman Iterations

Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 32001, Taiwan

syyang@math.ncu.edu.tw
http://www.math.ncu.edu.tw/∼syyang/

Outline of “Split Bregman iterations”

In this lecture, we will briefly introduce the split Bregman iterations for solving constrained convex minimization problems.

The material of this lecture is mainly based on the following papers:

- J.-F. Cai, S. Osher, and Z. Shen, Split Bregman methods and frame based image restoration, *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal,* 8 (2009), pp. 337-369.
- P. Getreuer, Rudin-Osher-Fatemi total variation denoising using split Bregman, *Image Processing On Line,* 2 (2012), pp. 74-95.
- T. Goldstein and S. Osher, The split Bregman method for $L1$ regularized problems, *SIAM Journal on Imaging Sciences,* 2 (2009), pp. 323-343.

Constrained convex minimization problem

- Bregman iteration is a technique for solving constrained convex minimization problems of the form:

$$\min_{u \in \mathbb{R}^N} J(u) \;\; \text{subject to} \;\; H(u) = 0,$$

where J and H are convex functions defined on the Hilbert space $\mathbb{R}^N$ with the inner product $\langle \cdot, \cdot \rangle$, that is, $J, H : \mathbb{R}^N \to \mathbb{R}$ are convex.

- The key idea is the Bregman distance. *The Bregman distance associated with a convex function J at the point $v \in \mathbb{R}^N$* is defined as follows: for $u \in \mathbb{R}^N$ and $p \in \partial J(v)$, we define

$$D_J^p(u, v) := J(u) - J(v) - \langle p, u - v \rangle, \qquad (\geq 0)$$

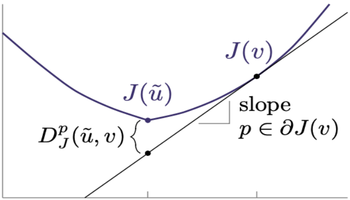
where p is called *a subgradient at v* and $\partial J(v)$ is the set of all subgradients at v, called *the subdifferential of J at v*, defined as

$$\partial J(v) := \{p \in \mathbb{R}^N : \; J(u) \geq J(v) + \langle p, u - v \rangle, \; \forall \, u \in \mathbb{R}^N\}.$$

Bregman distance

Note that the Bregman distance associated with the convex function J at the point v is

$$\begin{aligned} D_J^p(u,v) &:= J(u) - J(v) - \langle p, u - v\rangle \\ &= J(u) + \langle p, v - u\rangle - J(v) \geq 0. \end{aligned}$$

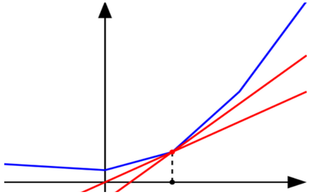


1-D Bregman distance $D_J^p(\tilde{u}, v)$

Subgradients and subdifferential

- Consider the convex function $J : \mathbb{R} \to \mathbb{R}$ defined by $J(u) = |u|$.

$$\begin{array}{lll} (1) & v = 0: & \partial J(v) = \{p \in \mathbb{R} : -1 \leq p \leq 1\}; \\ (2) & v < 0: & \partial J(v) = \{p \in \mathbb{R} : p = -1\}; \\ (3) & v > 0: & \partial J(v) = \{p \in \mathbb{R} : p = 1\}. \end{array}$$



1-D subgradients

- *The subdifferential $\partial J(v)$ is a nonempty, convex, and compact set,*

$$\partial J(v) := \{p \in \mathbb{R}^N : \; J(u) \geq J(v) + \langle p, u - v \rangle, \; \forall\, u \in \mathbb{R}^N\}.$$

Distance-like properties

- Bregman distance is not a distance in the usual sense because it is not in general symmetric,
  (1) $D_J^p(u,v) \neq D_J^q(v,u)$
  (2) If $p \in \partial J(u) \cap \partial J(v)$ and $D_J^p(v,u) = D_J^p(u,v)$, then $D_J^p(v,u) = D_J^p(u,v) = 0$.

  And the triangle inequality is not satisfied.
- From the definition of the distance and the convexity of J, we have the following distance-like properties:
  (1) $D_J^p(v,v) = 0$
  (2) $D_J^p(u,v) \geq 0$
  (3) $D_J^p(u,v) + D_J^q(v,w) - D_J^q(u,w) = \langle p-q, v-u \rangle,\ p,q \in \partial J(v)$.

  Proof of (3): By a direct computation, we have $J(u) - J(v) - \langle p, u-v \rangle + J(v) - J(w) - \langle q, v-w \rangle - J(u) + J(w) + \langle q, u-w \rangle = \langle p, v-u \rangle - \langle q, v-w \rangle + \langle q, u-w \rangle = \langle p-q, v-u \rangle$.

Basic idea of the Bregman iterations

- We consider the constrained convex minimization problems of the form:
$$\min_{u\in\mathbb{R}^N} J(u) \ \text{ subject to } \ H(u) = 0.$$
The associated unconstrained minimization problem is
$$\min_{u\in\mathbb{R}^N} \Big\{ J(u) + \gamma H(u) \Big\},$$
where $\gamma > 0$ is a penalty parameter.
- Given a starting point u^0 and parameter $\gamma > 0$, the Bregman iteration algorithm is formally
$$\begin{aligned} u^{k+1} &= \arg\min_u \Big\{ D_J^{p^k}(u, u^k) + \gamma H(u) \Big\}, \quad p^k \in \partial J(u^k), \\ &= \arg\min_u \Big\{ J(u) - \underbrace{J(u^k)}_{\text{constant}} - \langle p^k, u - u^k\rangle + \gamma H(u) \Big\}, \end{aligned}$$
as was suggested by Bregman in 1967. *(The existence of solutions u^{k+1} is nontrivial if the search space is infinite dimensional)*

$H(u^k)$ is decreasing in k

The iteration has the property that $H(u^k)$ is decreasing in k.

Proof: Since $D_J^{p^k}(u^{k+1}, u^k) \geq 0$ and u^{k+1} is a minimizer of the problem

$$u^{k+1} = \arg\min_u \left\{ D_J^{p^k}(u, u^k) + \gamma H(u) \right\},$$

and $D_J^{p^k}(u^k, u^k) = 0$, we have

$$\begin{aligned} \gamma H(u^{k+1}) &\leq D_J^{p^k}(u^{k+1}, u^k) + \gamma H(u^{k+1}) \\ &\leq D_J^{p^k}(u^k, u^k) + \gamma H(u^k) = \gamma H(u^k). \end{aligned}$$

Therefore, $H(u^k)$ is decreasing in k.

Iteration of the subgradients p at u^{k+1}

For simplicity, we assume that H is differentiable. Then we have the subdifferential at u,

$$\partial\Big(\underbrace{J(u) - J(u^k) - \langle p^k, u - u^k\rangle + \gamma H(u)}_{=D_J^{p^k}(u,u^k)+\gamma H(u)}\Big) = \partial J(u) - p^k + \gamma \nabla H(u).$$

Since u^{k+1} minimizing $D_J^{p^k}(u, u^k) + \gamma H(u)$, the optimality condition gives

$$\begin{aligned} 0 &\in \partial J(u^{k+1}) - p^k + \gamma \nabla H(u^{k+1}) \\ \Leftrightarrow &\quad p^k - \gamma \nabla H(u^{k+1}) \in \partial J(u^{k+1}). \end{aligned}$$

Therefore, we can select p^{k+1} as

$$p^{k+1} := p^k - \gamma \nabla H(u^{k+1}).$$

Bregman iteration algorithm

The Bregman iteration algorithm is defined as follows:

$$p^0 \in \partial J(u^0) \qquad (u^0 = p^0 = 0)$$

for $k = 0, 1, \ldots$ **do**

$$u^{k+1} = \arg\min_u D_J^{p^k}(u, u^k) + \gamma H(u)$$

$$p^{k+1} = p^k - \gamma \nabla H(u^{k+1})$$

Convergence results

- Suppose that H is differentiable and solutions u^{k+1} exist that are obtained by the Bregman iterations, then the convergence results hold: for any $\tilde{u}$ such that $H(\tilde{u}) = 0$ and $J(\tilde{u}) < \infty$,

$$D_J^{p^{k+1}}(\tilde{u}, u^{k+1}) \le D_J^{p^k}(\tilde{u}, u^k) \quad \textit{and} \quad H(u^k) \le \frac{J(\tilde{u})}{\gamma^k}.$$

  Particularly, $\{u^k\}$ is a minimizing sequence of H.

- The limiting solution satisfies the constraint $H(u) = 0$ exactly for any $\gamma > 0$. However, the value of γ does affect the convergence speed and numerical conditioning of the minimization problem, so γ should be selected according to these consideration.

Back to the discretization of ROF model

Applying the operator splitting technique, we obtain the constrained approximate minimization of the ROF model:

$$\min_{d,\,u} \underbrace{\Big\{ \sum_{i,j} |d_{i,j}| + \frac{\lambda}{2} \sum_{i,j} (f_{i,j} - u_{i,j})^2 \Big\}}_{:=J(u)} \text{ subject to } \underbrace{\nabla u_{i,j} - d_{i,j} = 0,\ \forall\, i,j}_{Au=g}.$$

It should be understood that here we use u in $J(u)$ and $H(u)$ to denote (d, u) and $g = 0$. Introducing a penalty parameter $\gamma > 0$, we obtain the unconstrained minimization problem:

$$\min_{u} \Big\{ J(u) + \underbrace{\frac{\gamma}{2} \|Au - g\|_2^2}_{:=\gamma H(u)} \Big\}.$$

Note that here we have $\nabla H(u) = A^\top (Au - g)$.

Supplement: $\nabla H(u) = A^\top (Au - g)$

Assume that

$$A = [a_{ij}]_{N\times N} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}_{N\times N}$$

$u = [u_1, u_2, \cdots, u_N]^\top$, and $g = [g_1, g_2, \cdots, g_N]^\top$. Then we have

$$H(u) = \frac{1}{2}\|Au - g\|_2^2 = \frac{1}{2}\sum_{j=1}^{N}\Big(\big(a_{j1}u_1 + a_{j2}u_2 + \cdots a_{jN}u_N\big) - g_j\Big)^2,$$

which by chain rule implies

$$\frac{\partial}{\partial u_i}H(u) = \sum_{j=1}^{N} a_{ji}\Big(\big(a_{j1}u_1 + a_{j2}u_2 + \cdots a_{jN}u_N\big) - g_j\Big) = \langle (A^\top)_{i\cdot}, (Au - g)\rangle.$$

Therefore, we obtain

$$\nabla H(u) = A^\top (Au - g).$$

Bregman iteration algorithm

Applying the Bregman iteration algorithm to the unconstrained minimization problem, we obtain

$$\begin{aligned} u^{k+1} &= \arg\min_u \Big\{ J(u) - J(u^k) - \langle p^k, u - u^k \rangle + \frac{\gamma}{2} \|Au - g\|_2^2 \Big\}, \quad (\star) \\ p^{k+1} &= p^k - \gamma A^\top (Au^{k+1} - g). \quad (\star\star) \end{aligned}$$

The above Bregman iterations can be reformulated into a compact form. By $p^0 = 0$ and $(\star\star)$, we obtain

$$p^{k+1} = -\gamma A^\top \sum_{i=1}^{k+1} (Au^i - g). \qquad (\star\star\star)$$

Substitute this into $(\star)$, and it yields

$$u^{k+1} = \arg\min_u \Big\{ J(u) + \frac{\gamma}{2} \|Au - g + \sum_{i=1}^{k} (Au^i - g)\|_2^2 \Big\}.$$

Supplement 1

By $p^0 = 0$ and $p^{k+1} = p^k - \gamma A^\top (Au^{k+1} - g)$, we obtain

$$\begin{aligned} p^{k+1} &= \left(p^{k-1} - \gamma A^\top (Au^k - g)\right) - \gamma A^\top (Au^{k+1} - g) \\ &= p^{k-1} - \gamma A^\top \Big((Au^k - g) + (Au^{k+1} - g)\Big) \\ &= \cdots \\ &= p^0 - \gamma A^\top \sum_{i=1}^{k+1} (Au^i - g) \\ &= -\gamma A^\top \sum_{i=1}^{k+1} (Au^i - g). \qquad (\star\star\star) \end{aligned}$$

By $(\star\star\star)$, we have

$$p^k = -\gamma A^\top \sum_{i=1}^{k} (Au^i - g).$$

Supplement 2

By $(\star\star\star)$, we have

$$p^k = -\gamma A^\top \sum_{i=1}^{k} (Au^i - g).$$

Substituting p^k into $(\star)$,

$$u^{k+1} = \mathrm{AM}_u \left\{ J(u) - J(u^k) - \langle p^k, u - u^k \rangle + \frac{\gamma}{2} \|Au - g\|_2^2 \right\}, \quad (\star)$$

we have

$$\begin{aligned} u^{k+1} &= \mathrm{AM}_u \left\{ J(u) - J(u^k) + \gamma \langle A^\top \sum_{i=1}^{k} (Au^i - g), u - u^k \rangle + \frac{\gamma}{2} \|Au - g\|_2^2 \right\} \\ &= \mathrm{AM}_u \left\{ J(u) - J(u^k) + \gamma \langle \sum_{i=1}^{k} (Au^i - g), A(u - u^k) \rangle + \frac{\gamma}{2} \|Au - g\|_2^2 \right\} \\ &= \mathrm{AM}_u \left\{ J(u) + \gamma \langle \sum_{i=1}^{k} (Au^i - g), Au \rangle + \frac{\gamma}{2} \|Au - g\|_2^2 + \text{constant} \right\}. \end{aligned}$$

Supplement 3

On the other hand, from the last equation on page 14,

$$
\begin{aligned}
u^{k+1} &= \mathrm{AM}_u\Big\{J(u) + \frac{\gamma}{2}\|Au - g + \sum_{i=1}^{k}(Au^i - g)\|_2^2\Big\} \\
&= \mathrm{AM}_u\Big\{J(u) + \frac{\gamma}{2}\|Au - g\|_2^2 + \frac{\gamma}{2}\|\sum_{i=1}^{k}(Au^i - g)\|_2^2 \\
&\qquad + \gamma\langle Au - g, \sum_{i=1}^{k}(Au^i - g)\rangle\Big\} \\
&= \mathrm{AM}_u\Big\{J(u) + \frac{\gamma}{2}\|Au - g\|_2^2 + \gamma\langle Au, \sum_{i=1}^{k}(Au^i - g)\rangle + \text{constant}\Big\}.
\end{aligned}
$$

Comparing this with the last equation on page 16, we can conclude that the last equation on page 14 is valid!

Supplement 4

On page 16, we have used the indentity: For any $A \in \mathbb{R}^{N \times N}$ and $u, w \in \mathbb{R}^N$, we have

$$\langle A^\top w, u \rangle = \langle w, Au \rangle.$$

Proof: Let $A = [a_{ij}]_{N\times N}$, $w = [w_1, w_2, \cdots, w_N]^\top$, $u = [u_1, u_2, \cdots, u_N]^\top$. Then by the direct computations we have

$$\begin{aligned}
\langle w, Au \rangle &= w_1(a_{11}u_1 + a_{12}u_2 + \cdots + a_{1N}u_N) \\
&\quad + w_2(a_{21}u_1 + a_{22}u_2 + \cdots + a_{2N}u_N) \\
&\quad + \cdots \\
&\quad + w_N(a_{N1}u_1 + a_{N2}u_2 + \cdots + a_{NN}u_N) \\
&= (a_{11}w_1 + a_{21}w_2 + \cdots + a_{N1}w_N)u_1 \\
&\quad + (a_{12}w_1 + a_{22}w_2 + \cdots + a_{N2}w_N)u_2 \\
&\quad + \cdots \\
&\quad + (a_{1N}w_1 + a_{2N}w_2 + \cdots + a_{NN}w_N)u_N = \langle A^\top w, u \rangle.
\end{aligned}$$

Bregman iteration algorithm (cont'd)

Let $u^0 = 0$ and

$$b^k := \sum_{i=1}^{k} (Au^i - g).$$

Then, we get a compact form of the Bregman iterations ($\star$) and ($\star\star$) as follows:

$$\begin{aligned} u^{k+1} &= \arg\min_u \left\{ J(u) + \frac{\gamma}{2} \|Au - g + b^k\|_2^2 \right\}, \\ b^{k+1} &= b^k + (Au^{k+1} - g). \end{aligned}$$

Note that by $(\star\star\star)$ the relation between p^k and b^k is

$$p^k = -\gamma A^\top b^k.$$

Explicit form of the Bregman iterations for TV denoising

Finally, by replacing u by (d, u), we have the following explicit form of the Bregman iterations for the discretization of ROF model:

$$
\begin{aligned}
(d^{k+1}, u^{k+1}) &= \arg\min_{d,\, u} \Big\{ \sum_{i,j} |d_{i,j}| + \frac{\lambda}{2} \sum_{i,j} (f_{i,j} - u_{i,j})^2 \\
&\qquad + \frac{\gamma}{2} \sum_{i,j} |\nabla u_{i,j} - d_{i,j} + b_{i,j}|^2 \Big\}, \\
b^{k+1} &= b^k + (Au^{k+1} - g) = b^k + \nabla u^{k+1} - d^{k+1}.
\end{aligned}
$$

- Note that we need some numerical methods to solve the (d^{k+1}, u^{k+1})-minimization problem at each iteration.
- *We use an alternating direction approach in the "image denoising lecture."* $\Longrightarrow$ *split Bregman iterations!*