The Bregman Methods: Review and New Results

Wotao Yin

Department of Computational and Applied Mathematics, Rice University

(Work supported by NSF, ONR, and Sloan Foundation)

December 8, 2009

Acknowledgements: results come in part from discussions with Stan Osher, Christ Brune, and Martin Burger.

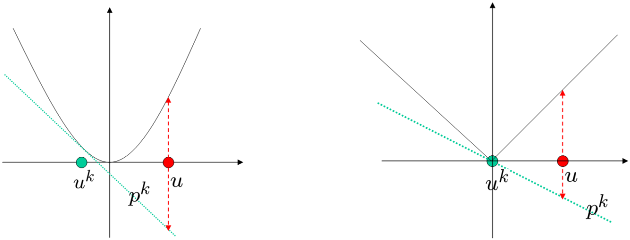
Bregman iteration has been unreasonably successful in

1. Improving the solution quality of regularizers such as ℓ_1, total variation, ...
2. Giving fast, accurate methods for *constrained* ℓ_1-like minimization.

Bregman Distance

- Original model: $\min J(u) + f(u)$. Regularizer $J(\cdot)$
- Given u^k, $p^k \in \partial J(u^k)$
- Bregman distance:

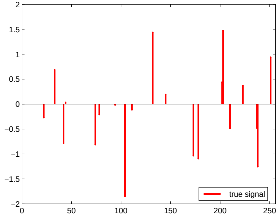
$$D(u, u^k) := J(u) - \left(J(u^k) + \langle p^k, u - u^k \rangle\right)$$



- New model: $u^{k+1} \leftarrow \min \alpha D(u, u^k) + f(u)$. E.g.: $\alpha = 5$. p^k is obtainable from previous iteration.

Example: Compressive Sensing with Noise

- Sparse original signal u

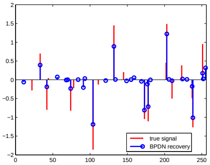


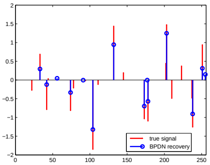
- *Noisy* Gaussian measurements: $b \leftarrow Au + \omega$. A: 100×250.

Models:

- ~~ℓ_0 minimization: $\min \mu \|u\|_0 + \frac{1}{2}\|Au - b\|_2^2$~~. Computationally intractable!
- Basis pursuit: $u \leftarrow \min \mu \|u\|_1 + \frac{1}{2}\|Au - b\|_2^2$
- Bregman: $u^{k+1} \leftarrow \min D(u, u^k) + \frac{1}{2}\|Au - b\|_2^2$

- Basis pursuit:

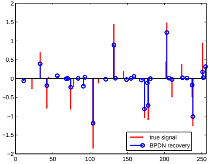




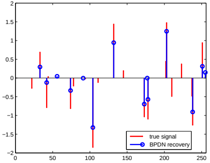
$\mu = 48.5$
Not sparse

$\mu = 49$
Poor fitting

- Basis pursuit:

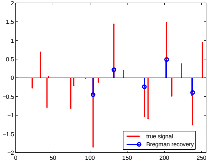


$\mu = 48.5$
Not sparse

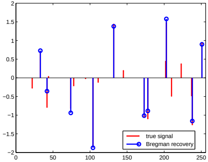


$\mu = 49$
Poor fitting

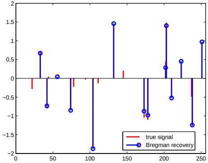
- Bregman: over-regularization $\mu = 150$



Itr 1



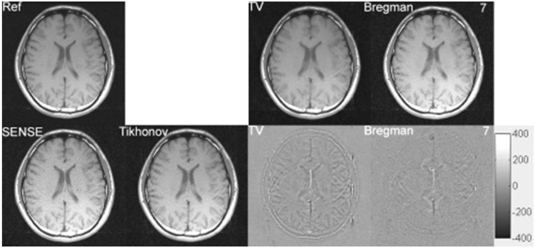
Itr 3



Itr 5

Example: image deblurring and/or denoising

- $J(u) = \mu TV(u)$
- $f(u) = \frac{1}{2}\|Au - b\|_2^2$
- Stop when $\|Au^k - b\|_2^2 \approx$ est. $\|Au^{true} - b\|_2^2$



(UWM-CMRI Lab)

MR Image Reconstruction from Very Few Data

Cut from R. Chartrand's paper. Applied Bregman for ℓ_p minimization.

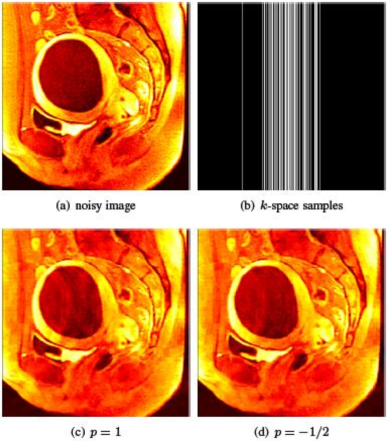


Fig. 2. Synthetic k-space samples are generated for a noisy image, using Gaussian random phase encoding. Reconstruction with $p = 1$ shows greater aliasing than with $p = -1/2$.

- For ℓ_1: Bregman gives sparser, better fitted signals
- For TV: Bregman gives less staircasing, higher contrast

- Reason: *iterative boosting*
  1. Over-regularized u^k: have correct locations for larger nonzeros/edges
  2. $D(u, u^k)$: no regularization for correctly located entries of u

$$D(u, u^k) = J(u) - \left(J(u^k) + \langle p^k, u - u^k \rangle \right)$$

1. Improving the performance of ℓ_1, total variation, ...
   - Work for noisy data
   - Start with over-regularization
   - $f(u^k)\downarrow$, stop $f(u^k) \approx f(\mathit{true}\ u)$ est.

1. Improving the performance of ℓ_1, total variation, ...
   - Work for noisy data
   - Start with over-regularization
   - $f(u^k)\downarrow$, stop $f(u^k) \approx f(\textit{true } u)$ est.

2. Giving fast, accurate methods for constrained ℓ_1 and TV minimization.
   - Work for *noiseless* data
   - $f(u^k)\downarrow$, stop $f(u^k) = 0$.

Applied to Constrained Minimization

Y.-Osher-Goldfarb-Burger 07

- Purpose: $u^{true} \leftarrow \min\{J(u) : Au = b\}$, constrained
- Bregman: $u^{k+1} \leftarrow \min \mu D(u, u^k) + \frac{1}{2}\|Au - b\|_2^2$, unconstrained
- Properties:
  - $u^k \to u^{true}$
  - Fast, finite convergence for ℓ_1–like $J(u)$
  - Accurate, even if subproblems are solved inexactly

However, Bregman iteration has been around since 1967. Moreover, it is equivalent to augmented Lagrangian (when constraints are linear), used in optimization and computation without great success in e.g., Navier-Stokes (NS), because NS involves basically L2 minimization.

Bregman turns out to work very well for ℓ_1, TV, and related minimization; nothing special otherwise.

Reason: Error Cancellation.

- Error cancellation is a happy result due to *adding back*!
- Bregman maintains $p^k \in J(u^k)$:

$$u^{k+1} \leftarrow \min J(u) - J(u^k) - \langle p^k, u - u^k \rangle + \frac{1}{2}\|Au - b\|_2^2$$

$$p^{k+1} \leftarrow p^k - A^\top (Au^k - b).$$

Can rewrite equivalently as

$$u^{k+1} \leftarrow \min J(u) + \frac{1}{2}\|Au - b^{k+1}\|_2^2 \tag{1}$$

$$b^{k+1} \leftarrow b + (b^k - Au^k).$$

- Suppose we make an error w^k and get $u^k_{inexact} = u^k + w^k$. The above update gives:

$$\min J(u) + \frac{1}{2}\|A(u + w^k) - b^{k+1}\|_2^2$$

Subproblem has *model error* compared to (1)!

Theorem

Let w be a model error, and consider solving

$$\min J(u) + f(u + w).$$

Let

- u_{exact}: *exact solution*
- $u_{inexact} = u_{exact} + v$, *where v is solution error*

If u_{exact} and $u_{exact} - w$ are on the same linear piece of J (a face of $\mathrm{graph}(J)$), then

$$u_{inexact} - \underbrace{\arg\min\{J(u) + f(u)\}}_{exact\ sol.\ of\ true\ model} = v - w.$$

Important implication: certain solvers enable <u>v to almost cancel w</u>.

Error Cancellation Example

- u^{true}: 500 entries, 25 nonzero, sparse
- $b = Au^{true}$: 250 linear projections of u^{true}, with a Gaussian random A
- Model: $\min\{\|u\|_1 : Au = b\}$
- Bregman Method: solve subproblems inexactly with tolerance $\equiv$ 1e-6

Error Cancellation Example

- u^{true}: 500 entries, 25 nonzero, sparse
- $b = Au^{true}$: 250 linear projections of u^{true}, with a Gaussian random A
- Model: $\min\{\|u\|_1 : Au = b\}$
- Bregman Method: solve subproblems inexactly with tolerance $\equiv$ 1e-6

Itr k	1	2	3	4	5
$\frac{\|u^{true} - u^k_{inexact}\|}{\|u^{true}\|}$	6.5e-2	2.3e-7	6.2e-14	7.9e-16	5.6e-16.

Error Cancellation Example

- u^{true}: 500 entries, 25 nonzero, sparse
- $b = Au^{true}$: 250 linear projections of u^{true}, with a Gaussian random A
- Model: $\min\{\|u\|_1 : Au = b\}$
- Bregman Method: solve subproblems inexactly with tolerance $\equiv$ 1e-6

Itr k	1	2	3	4	5
$\frac{\|u^{true} - u^k_{inexact}\|}{\|u^{true}\|}$	6.5e-2	2.3e-7	6.2e-14	7.9e-16	5.6e-16.

- Above high accuracy obtainable with subproblem solvers:
  FPC, FPC-BB, GPSR, GPSR–BB, SpaRSA

Generalization to Bregman Iterations

- Inverse scale space (Burger, Gilboa, Osher, Xu, etc.)
- Linearized Bregman (Yin, Osher, Mao, etc.)
- Logistic Regression (Shi, et al.)
- Split Bregman (Goldstein, Osher)
- More ... People use the words “Bregmanize” and “Bregmanized”

Linearized Bregman

Idea: Linearize the fitting term at u^k

Work: Y.-Osher-Goldfarb-Darbon 07, Osher-Mao-Dong-Y. 08, Cai-Osher-Shen 08, Y. 09

Linearized Bregman

Idea: Linearize the fitting term at u^k

Work: Y.-Osher-Goldfarb-Darbon 07, Osher-Mao-Dong-Y. 08, Cai-Osher-Shen 08, Y. 09

- Example: data fitting $= \frac{1}{2}\|Au - b\|_2^2$

$$u^{k+1} \leftarrow \min_u D(u, u^k) + \langle A^\top (Au^k - b), u\rangle + \frac{1}{2\delta}\left\|u - u^k\right\|_2^2$$

Linearized Bregman

Idea: Linearize the fitting term at u^k

Work: Y.-Osher-Goldfarb-Darbon 07, Osher-Mao-Dong-Y. 08, Cai-Osher-Shen 08, Y. 09

- Example: data fitting $= \frac{1}{2}\|Au - b\|_2^2$

$$u^{k+1} \leftarrow \min_u D(u, u^k) + \langle A^\top (Au^k - b), u\rangle + \frac{1}{2\delta}\left\|u - u^k\right\|_2^2$$

- For $D(u, u^k)$ induced by $J(u) = \mu\|u\|_1$, iterations become

$$\begin{aligned} u^{k+1} &\leftarrow \delta\,\mathrm{shrink}(v^k, \mu) \\ v^{k+1} &\leftarrow v^k + A^\top (b - Au^{k+1}). \end{aligned}$$

RICE

Linearized Bregman

Idea: Linearize the fitting term at u^k

Work: Y.-Osher-Goldfarb-Darbon 07, Osher-Mao-Dong-Y. 08, Cai-Osher-Shen 08, Y. 09

- Example: data fitting $= \frac{1}{2}\|Au - b\|_2^2$

$$u^{k+1} \leftarrow \min_u D(u, u^k) + \langle A^\top (Au^k - b), u\rangle + \frac{1}{2\delta}\left\|u - u^k\right\|_2^2$$

- For $D(u, u^k)$ induced by $J(u) = \mu\|u\|_1$, iterations become

$$\begin{aligned} u^{k+1} &\leftarrow \delta\,\mathrm{shrink}(v^k, \mu) \\ v^{k+1} &\leftarrow v^k + A^\top(b - Au^{k+1}). \end{aligned}$$

- Application: non-negative least–squares, matrix completion

Linearized Bregman, Cont'd

Properties:

- gradient-ascend the dual of $\min\{\mu\|u\|_1 + \frac{1}{2\delta}\|u\|^2 : Au = b\}$
- Exact regularization: $\exists\bar{\delta}$: if $\delta > \bar{\delta}$, then solves $\min\{\|u\|_1 : Au = b\}$
- $\#$ nonzeros of u^k often grows monotonically in k

Split Bregman and Alternating Direction Method

Split Bregman (Goldstein–Osher 08): variable splitting + aug. Lagrangian

- Splitting (Wang–Yang–Y.–Zhang 07,08): $\min_u f(Lu) + g(u) \Longrightarrow \min_{u,v}\{f(v) + g(u) : v = Lu\}$ Great payoff for many imaging problems
- Aug. Lag.: λ – multiplier
  1. $\min_{u,v} f(v) + \frac{c}{2}\|v - Lu - \lambda\|_2^2 + g(u)$
  2. update λ
- A special case of the alternating direction method

Split Bregman and Alternating Direction Method

Alternating direction method: (Douglas–Rachford 60s, Glowinski–Marocco, Gabay–Mercier, 70s)

1. fix u, minimize w.r.t. v
2. fix v, minimize w.r.t. u
3. update λ

Example (Wang–Yang–Y.–Zhang 07,08) Compressed MRI, image deblr

$$\min_u \mu TV(u) + \frac{1}{2}\|Au - b\|_2^2 \Leftrightarrow \min_u \{\mu\|w\|_1 + \frac{1}{2}\|Au - b\|_2^2 : w = Du\}$$

where A is partial Fourier or convolution. ADM extends to color images, duals, rank-minimization

Summary

1. Bregman improves ℓ_1-like regularization quality for noisy data
2. Bregman applied to constrained ($Au = b$) minimization is not new but is fast and accurate due *adding back*
3. Various extensions take advantages of model structures

More details and solvers at *Rice L_1-Related Optimization Project*

RICE