Stability Analysis for Estimating Parameters in the Split Bregman Algorithm for Signal Restoration

Rosemary Renaut
http://math.asu.edu/~rosie

GAMM Meeting
September 2012
Acknowledgement:
Charles University Prague
Illustration: Blurred Signal Restoration

\[ b(t) = \int_{-\pi}^{\pi} h(t, s)x(s)\,ds \quad \text{and} \quad h(s) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-s^2}{2\sigma^2}\right), \quad \sigma > 0. \]

Solve \( Ax \approx b \) where \( A \) describes the blur \( h(s) \).
Sensitivity to Noise in the data: here noise 0.0001

Solution depends on the conditioning of $A$, here $1.8679e + 05$

inverse crime: $x = A^{-1}(b + n)$
Ill-conditioned Least Squares: Tikhonov Regularization

Solve ill-conditioned

\[ Ax \approx b \]

Standard Tikhonov, \( L \) approximates a derivative operator

\[
x(\lambda) = \arg\min_x \left\{ \frac{1}{2} \| Ax - b \|_2^2 + \frac{\lambda^2}{2} \| Lx \|_2^2 \right\}
\]

\( x(\lambda) \) solves normal equations provided \( \text{null}(L) \cap \text{null}(A) = \{0\} \)

\[
(A^T A + \lambda^2 L^T L)x(\lambda) = A^T b
\]

This is not good for preserving edges in solutions.
Not good for extracting features in images.
Consider general regularization \( R(x) \) suited to properties of \( x \):

\[
x(\lambda) = \arg\min_{\{x\}} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} R(x) \right\}
\]

Suppose \( R \) is total variation of \( x \) (general options are possible).

The total variation for a function \( x \) defined on a discrete grid is

\[
TV(x(s)) = \sum_i |x(s_i) - x(s_{i-1})| \approx \Delta \sum_i |dx(s_i)/ds| \approx \Delta \|Lx\|_1
\]

TV approximates a scaled sum of the magnitude of jumps in \( x \).
\( \Delta \) is a scale factor dependent on the grid size.

\[
x(\lambda) = \arg\min_{\{x\}} \left\{ \|Ax - b\|_2^2 + \lambda^2 \|Lx\|_1 \right\}
\]

Problematic for large scale to solve the optimization problem.
Introduce \( d \approx Lx \) and let 
\[
R(x) = \frac{\lambda^2}{2} \| d - Lx \|^2_2 + \mu \| d \|_1
\]

\[
(x, d)(\lambda, \mu) = \text{argmin}_{\{x, d\}} \left\{ \frac{1}{2} \| A x - b \|^2_2 + \frac{\lambda^2}{2} \| d - Lx \|^2_2 + \mu \| d \|_1 \right\}
\]

Alternating minimization separates steps for \( d \) from \( x \)
Various versions of the iteration can be defined. Fundamentally:

\[
S1 : x^{(k+1)} = \text{argmin}_{\{x\}} \left\{ \frac{1}{2} \| A x - b \|^2_2 + \frac{\lambda^2}{2} \| L x - (d^{(k+1)} - g^{(k)}) \|^2_2 \right\}
\]

\[
S2 : d^{(k+1)} = \text{argmin}_{\{d\}} \left\{ \frac{\lambda^2}{2} \| d - (L x^{(k+1)} + g^{(k)}) \|^2_2 + \mu \| d \|_1 \right\}
\]

\[
S3 : g^{(k+1)} = g^{(k)} + L x^{(k+1)} - d^{(k+1)}.
\]
Advantages of the formulation

Update for $g$: updates the Lagrange multiplier $g$

\[ S3: \quad g^{(k+1)} = g^{(k)} + Lx^{(k+1)} - d^{(k+1)}. \]

This is just a vector update

Update for $d$:

\[ S2: \quad d = \arg\min_{\{d\}} \left\{ \mu \|d\|_1 + \frac{\lambda^2}{2} \|d - c\|_2^2 \right\}, \quad c = Lx + g \]

\[ = \arg\min_{\{d\}} \left\{ \|d\|_1 + \frac{\gamma}{2} \|d - c\|_2^2 \right\}, \quad \gamma = \frac{\lambda^2}{\mu}. \]

This is achieved using soft thresholding.
Thresholding for \( \mathbf{d} \)

If \( \mathbf{d} \) has \( q \) components \( (\mathbf{d})_i \) componentwise solution:

\[
(\mathbf{d})_i = \frac{c_i}{|c_i|} \max(|c_i| - \frac{1}{\gamma}, 0) \quad i = 1 : q
\]

If \( \mathbf{d} \) is two dimensional it contains components \( \mathbf{d}_x \) and \( \mathbf{d}_y \). Threshold is applied for each component of \( (\mathbf{d}_x^T, \mathbf{d}_y^T)^T \): we use

\[
\| \mathbf{d} \|_1 = \| \mathbf{d}_x \|_1 + \| \mathbf{d}_y \|_1
\]

The TV norm for the two dimensional case can be written

\[
\| \mathbf{d} \|_{\text{TV}} = \left( \sum_{l=1}^{n} \| \mathbf{d}_l \|_2 \right) \quad q = 2(n - 1)
\]

Intrinsically TV is still local

\[
(\mathbf{d}_{\text{TV}})_i = \frac{c_i}{\| c_i \|_2} \max(\| c_i \|_2 - \frac{1}{\gamma}, 0) \quad i = 1 : n
\]
The Tikhonov Step of the Algorithm

\[ S1 : x^{(k+1)} = \arg\min_{\{x\}} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|Lx - (d^{(k+1)} - g^{(k)})\|_2^2 \right\} \]

Update for \( x \):

\[ x = \arg\min_{\{x\}} \left\{ \frac{1}{2} \|Ax - b\|_2^2 + \frac{\lambda^2}{2} \|Lx - h\|_2^2 \right\}, \quad h = d - g \]

Standard least squares update using a Tikhonov regularizer.

**Disadvantages of the formulation**

- **update for \( x \):** A Tikhonov LS update each step
- Changing right hand side.
- Regularization parameter \( \lambda \) - dependent on \( k \)?
- Threshold parameter \( \mu \) - dependent on \( k \)?
Prior work

*ignores* parameter determination
*ignores* relationship with standard Tikhonov regularization.
**assumes** $\lambda, \mu$ constant over $k$.

An Opportunity

To determine the stability we analyze LS step of the algorithm
Take advantage of parameter estimation techniques to find $\lambda$
Use Generalized Singular Value Decomposition for analysis
The Generalized Singular Value Decomposition

Lemma (GSVD)

Assume invertibility and $m \geq n \geq p$. There exist unitary matrices $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{p \times p}$, and a nonsingular matrix $Z \in \mathbb{R}^{n \times n}$ such that

$$A = U \begin{bmatrix} \Upsilon & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} Z^T, \quad L = V \begin{bmatrix} M, \mathbf{0} \end{bmatrix} Z^T,$$

$$\Upsilon = \text{diag}(\nu_1, \ldots, \nu_p, 1, \ldots, 1) \in \mathbb{R}^{n \times n}, \quad M = \text{diag}(\mu_1, \ldots, \mu_p) \in \mathbb{R}^{p \times p},$$

with

$$0 \leq \nu_1 \leq \cdots \leq \nu_p \leq 1, \quad 1 \geq \mu_1 \geq \cdots \geq \mu_p > 0, \quad \nu_i^2 + \mu_i^2 = 1, \quad i = 1 : p.$$

Generalized singular values: $\gamma_i = \frac{\nu_i}{\mu_i}$
Solution in terms of the GSVD Basis

Update for $x$: using the GSVD : basis $Z$

$$
x^{(k+1)} = \sum_{i=1}^{p} \left( \frac{\nu_i u_i^T b}{\nu_i^2 + \lambda^2 \mu_i^2} + \frac{\lambda^2 \mu_i v_i^T h^{(k)}}{\nu_i^2 + \lambda^2 \mu_i^2} \right) z_i + \sum_{i=p+1}^{n} (u_i^T b) z_i
$$

A weighted combination of basis vectors $z_i$: weights

$$
(i) \frac{\nu_i^2}{\nu_i^2 + \lambda^2 \mu_i^2} u_i^T b \quad (ii) \frac{\lambda^2 \mu_i^2}{\nu_i^2 + \lambda^2 \mu_i^2} v_i^T h^{(k)}
$$

Notice (i) is fixed by $b$, but (ii) depends on the updates $h^{(k)}$

If (i) dominates (ii) solution will converge slowly or not at all

$\lambda$ impacts the solution and must not over damp (ii)
Solution $Lx$

Update for $Lx$: using the GSVD: basis $V$

$$Lx(\lambda) = \sum_{i=1}^{p} \left( \frac{\nu_i \mu_i u_i^T b}{\nu_i^2 + \lambda^2 \mu_i^2} + \frac{\lambda^2 \nu_i^2 v_i^T h^{(k)}}{\nu_i^2 + \lambda^2 \mu_i^2} \right) v_i$$

Clearly important that the terms also in $Lx(\lambda)$ are not over damped - this determines the shrinkage step

Let $\phi_i = \frac{\nu_i^2}{\nu_i^2 + \lambda^2 \mu_i^2} = \frac{1}{\gamma_i^2 + \lambda^2}$

$$x^{(k+1)} = \sum_{i=1}^{p} \left( \phi_i \frac{u_i^T b}{\nu_i} + (1 - \phi_i) \frac{v_i^T h^{(k)}}{\mu_i} \right) z_i + \sum_{i=p+1}^{n} (u_i^T b) z_i$$

$$Lx^{(k+1)} = \sum_{i=1}^{p} \left( \phi_i \frac{u_i^T b}{\gamma_i} + (1 - \phi_i) v_i^T h^{(k)} \right) v_i.$$

If $\phi_i$ constant then these are update equations for $x$ and $Lx$
Returning to the SVD for Solution $Ax \approx b$

Singular value decomposition (SVD) of $A$ (full column rank)

$$A = U \Sigma V^T = \sum_{i=1}^{n} u_i \sigma_i v_i^T, \quad \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_n).$$

gives expansion for the solution

$$x = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i$$

Solution is a weighted linear combination of the basis vectors $v_i$. 
The Discrete Picard Condition to examine the SVD

Problem size 84. condition, $9.3873e + 20$

Figure: Picard condition for a problem size 84. Ratios $|u_i^T b| / \sigma_i$ order 1 with no noise
The same problem with noise added in $b$ - low noise variance $0.0001$

Figure: With small noise the weights **blow up**
Filtered SVD - used to handle the exponential decaying spectrum

Recall the filtered solution which corresponds to the Tikhonov regularization

\[ x = \sum_{i=1}^{n} \phi_i \frac{u_i^T b}{\sigma_i} v_i \]

Notice relation with the GSVD solutions for the SB algorithm:

\[ x^{(k+1)} = \sum_{i=1}^{p} \left( \phi_i \frac{u_i^T b}{\nu_i} + (1 - \phi_i) \frac{v_i^T h^{(k)}}{\mu_i} \right) z_i + \sum_{i=p+1}^{n} (u_i^T b) z_i \]

\[ Lx^{(k+1)} = \sum_{i=1}^{p} \left( \phi_i \frac{u_i^T b}{\gamma_i} + (1 - \phi_i) v_i^T h^{(k)} \right) v_i. \]

Can expect that \( \lambda \) is \( k \)-dependent.
Examining the solution with $\lambda$ increasing

Coefficients $|u_i^Tb|$ and $|v_i^T h|$, $\lambda = 0.001$, 0.1 and 10
Examining the solution with $\lambda$ increasing

Filter Coefficients for $x^{(k)}$ and $Lx^{(k)}$ for $\lambda = .001, .1$ and $10$
Examining the solution with $\lambda$ increasing

Solutions: $\lambda = 0.001, 0.1$ and $10$
The weights on both $x$ and $Lx$ are crucial to avoid stagnation.

$\lambda$ needs to be chosen appropriately.

Wealth of possibilities for choosing $\lambda$ from standard Tikhonov regularization.

Examples here will use Unbiased Predictive Risk Estimation (UPRE) - assume variance of noise in $b$ is known.

For regularization in general the choice of $\lambda$ depends on the right hand side vector.

In this case $h$ changes each step.

Thus we might update $\lambda$ each step. (Adaptive choice for $\lambda$)
Numerical Experiments

Some Results

- SB UPRE uses the estimated $\lambda$ from UPRE for all SB steps.
- Update SB, updates $\lambda$ each step.
- We also use iteratively reweighted norm approach for the $d$ update (IRN)
- SB IRN updates and iteratively reweights $\|d\|_1$. 
Example Solution: 1D: No Updates for the parameters

Noise level 7.2769e-05 Comparing fixed $\lambda$

- Exact
- Data
- LS
- SB UPRE
- UPRE IRN
Example Solution: 1D Updates for the parameters

Figure: \( \gamma = 200 \) Low noise. Without updating \( \lambda \) left and updated right.

SB UPRE uses the estimated \( \lambda \) from UPRE for all SB steps. Update SB, updates \( \lambda \) each step. SB IRN updates and iteratively reweights \( \|d\|_1 \).
Example Solution: 2D - similar blurring operator (SNR)

Noise level 0.024857 Updating λ
LS 9.2355
SB Fixed λ 8.9701

SB Update 10.645
SB IRN Update 10.124
Example Solution: 2D - a projection case

Noise level 0.05 Updating
LS 8.905
SB Fixed 12.5759

SB Update 9.697
SB IRN Update 11.1868
Some theoretical results: using the Unbiased Predictive Risk

Lemma
Suppose that noise in $h^{(k)}$ is stochastic and both data fit $Ax \approx b$ and derivative data fit $\ell x \approx h$ are weighted by their inverse covariance matrices for normally distributed noise in $b$ and $h$; then the optimal choice for $\lambda$ at all steps is $\lambda = 1$.

Remark
Experiments violate the above assumption.

Lemma
Suppose that the vector $h^{(k+1)}$ is regarded as deterministic, then UPRE applied to find the the optimal choice for $\lambda$ at each step leads to a different optimum at each step, namely it depends on $h$.

Remark
Because $h$ changes each step the optimal choice for $\lambda$ using UPRE will thus change with each iteration, confirming that one should not fix $\lambda$ over all steps.
Further Observations and Future Work

Results: demonstrate basic analysis of problem is worthwhile.

Parameter: estimation from basic LS should be used to find appropriate parameter at least at step 1.

Overhead of optimal $\lambda$ for the first step - reasonable.

Overhead of subsequent steps - UPRE requires matrix trace - but can be optimized, or use $\chi^2$ approach.

Practically: use standard LSQR algorithms rather than GSVD.

Convergence: testing is based on $h$.

Future: significant scope to include more NLA in terms of subspace recycling or solutions for shifted systems.