AFOSR Review

Hybrid Constrained Iterative Methods for Deconvolution

James Nagy

Mathematics and Computer Science
Emory University
Atlanta, GA USA

Outline

1. Introduction
   - Define Notation
   - Regularization Basics
2. Iterative Methods
3. Conjugate Gradient and Hybrid Methods
   - Golub-Kahan Based Hybrid Method
   - Arnoldi Based Hybrid Method and Sparsity Constraints
4. Concluding Remarks

Deconvolution

Consider general problem

$$\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$$

where

- $\mathbf{b}$ is known vector (measured data)
- $\mathbf{x}$ is unknown vector (want to find this)
- $\mathbf{A}$ is large, usually ill-conditioned matrix (may be known or unknown)
- $\boldsymbol{\eta}$ is unknown vector (noise)

Examples with this general model: $\mathbf{b} = \mathbf{A}\mathbf{x} + \eta$

- Spatially invariant deconvolution
  - $\mathbf{x}$ is unknown, true image
  - $\mathbf{A}\mathbf{x}$ is convolution, where $\mathbf{A}$ is known (defined by PSF)
  - $\mathbf{b}$ is observed, blurred, noisy image
- Spatially variant blurs
  - same as above, except
  - need mathematical model of $\mathbf{A}\mathbf{x}$ (blurring operation)
- Multi-Frame problems

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \vdots \\ \mathbf{b}_m \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_m \end{bmatrix}$$

Basics of Regularization: SVD Analysis

Suppose $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, and $\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}} + \boldsymbol{\eta} = \mathbf{b}_{\text{true}} + \boldsymbol{\eta}$.

Naive inverse solution, $\mathbf{x}_{\text{inv}} = \mathbf{A}^{-1}\mathbf{b} = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^T\mathbf{b}$

$$\mathbf{x}_{\text{inv}} = \sum_{i=1}^{n} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i = \underbrace{\sum_{i=1}^{n} \frac{\mathbf{u}_i^T\mathbf{b}_{\text{true}}}{\sigma_i}\mathbf{v}_i}_{\mathbf{x}_{\text{true}}} + \underbrace{\sum_{i=1}^{n} \frac{\mathbf{u}_i^T\boldsymbol{\eta}}{\sigma_i}\mathbf{v}_i}_{\text{error}}$$

The goal is to balance:

- reconstructing "good" SVD components: $\dfrac{\mathbf{u}_i^T\mathbf{b}_{\text{true}}}{\sigma_i}$ (large σ_i)
- avoid reconstructing "bad" SVD components $\dfrac{\mathbf{u}_i^T\boldsymbol{\eta}}{\sigma_i}$ (small σ_i)

One approach: TSVD:

$$\mathbf{x}_{\text{tsvd}} = \sum_{i=1}^{k} \frac{\mathbf{u}_i^T\mathbf{b}}{\sigma_i}\mathbf{v}_i$$

SVD Filter-based Regularization

The TSVD idea can be generalized:

$$\mathbf{x}_{\text{filt}} = \sum_{i=1}^{n} \phi_i \frac{\mathbf{u}_i \mathbf{b}}{\sigma_i} \mathbf{v}_i, \quad \text{where} \quad \phi_i \approx \begin{cases} 1 & \text{for "large" } \sigma_i \\ 0 & \text{for "small" } \sigma_i \end{cases}$$

Examples:

- TSVD: $\phi_i = \begin{cases} 1 & i = 1, 2, \ldots, k \\ 0 & i = k+1, \ldots, n \end{cases}$
  We must choose regularization parameter k.
- Tikhonov/Wiener: $\phi_i = \dfrac{\sigma_i^2}{\sigma_i^2 + \alpha^2}$
  We must choose regularization parameter α
- Exponential: $\phi_i = 1 - e^{-\sigma_i^2/\alpha^2}$
  We must choose regularization parameter α

Filtering and Variational Regularization

- Tikhonov filtering is often written in variational form:

$$\min_{\mathbf{x}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \alpha^2 \|\mathbf{x}\|_2^2 \right\} \quad \text{or} \quad \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A} \\ \alpha \mathbf{I} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

- Tikhonov filtering can be generalized to:

$$\min_{\mathbf{x}} \left\{ \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \alpha^2 \|\mathbf{L}\mathbf{x}\|_2^2 \right\} \quad \text{or} \quad \min_{\mathbf{x}} \left\| \begin{bmatrix} \mathbf{A} \\ \alpha \mathbf{L} \end{bmatrix} \mathbf{x} - \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \right\|_2^2$$

  where $\mathbf{L}$ can be, e.g., a differentiation operator.

Remarks

We can use SVD filtering if we assume:

- Spatially invariant PSFs and periodic boundary conditions.
- The blur is separable.

We may need iterative methods when:

- Blur is spatially variant.
- Important information is near boundary of field of view.
- Additional constraints need to be imposed on solution.

Iterative Methods

Some approaches to using iterative methods:

1. Apply iterative method to variational form of regularization:

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2 + \alpha^2 \|\mathbf{L}\mathbf{x}\|_2^2$$

2. Apply iterative method directly to

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$

   enforce regularization by stopping iteration early.

3. Combine the two approaches $\Rightarrow$ Hybrid Method

Iterative Regularization

- Consider the inverse problem: $\mathbf{b} = \mathbf{A}\mathbf{x} + \boldsymbol{\eta}$

- Apply an iterative method to:

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2$$

  enforce regularization by stopping iteration early.

- Why should this work? Does it always work?

Iterative Regularization

Need to show solution at each iteration satisfies:

$$\mathbf{x}^{(k)} = \sum_{i=1}^{n} \phi_i^{(k)} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i \,,$$

where the filter factors satisfy:

$$\phi_i^{(k)} \approx \begin{cases} 1 & \text{for "large" } \sigma_i \\ 0 & \text{for "small" } \sigma_i \end{cases}$$

- More filtering is done when k is small.
- Less filtering when k is large.

Simple Example: Landweber (gradient descent)

- Iteration: $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \tau \mathbf{A}^T \left(\mathbf{b} - \mathbf{A}\mathbf{x}^{(k)} \right)$
- SVD filter: $\phi_i^{(k+1)} = 1 - (1 - \tau \sigma_i^2)^{k+1}$

Iterative Regularization

Remarks:

- Landweber is easy to analyze, but is very slow.
- Conjugate gradient methods:
  - Much faster to converge
  - But much harder to analyze
  - Example: LSQR

Golub-Kahan Bidiagonalization

LSQR is based on the Golub-Kahan bidiagonalization (GKB)

Given $m \times n$ $\mathbf{A}$, vector $\mathbf{b}$, k-th GKB iteration computes

$$\begin{aligned} \mathbf{A}^T\mathbf{W}_k &= \mathbf{Y}_k\mathbf{B}_k^T + \gamma_{k+1}\mathbf{y}_{k+1}\mathbf{e}_{k+1}^T \\ \mathbf{A}\mathbf{Y}_k &= \mathbf{W}_k\mathbf{B}_k, \end{aligned}$$

where

- $\mathbf{W}_k$ and $\mathbf{Y}_k$ have orthonormal columns
- $\mathbf{B}_k$ is bidiagonal:

$$\mathbf{B}_k = \begin{bmatrix} \gamma_1 & & & \\ \beta_2 & \gamma_2 & & \\ & \ddots & \ddots & \\ & & \beta_k & \gamma_k \\ & & & \beta_{k+1} \end{bmatrix}.$$

GKB Properties

Using GKB iterates, can compute approximate solution $\mathbf{x}^{(k)}$

$$\min_{\mathbf{x} \in R(\mathbf{Y}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \min_{\hat{\mathbf{x}}} \|\mathbf{B}_k \hat{\mathbf{x}} - \beta \mathbf{e}_1\|_2^2$$

where $\beta = \|\mathbf{b}\|_2$, and $\mathbf{x}^{(k)} = \mathbf{Y}_k \hat{\mathbf{x}}$.

Important property:

Singular values of $\mathbf{B}_k$ approximate large singular values of $\mathbf{A}$

GKB Properties

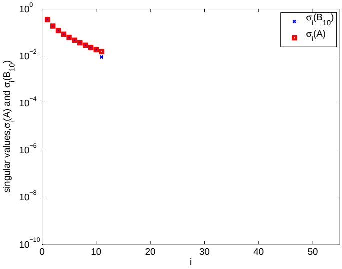
Using GKB iterates, can compute approximate solution $\mathbf{x}^{(k)}$

$$\min_{\mathbf{x} \in R(\mathbf{Y}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \min_{\hat{\mathbf{x}}} \|\mathbf{B}_k \hat{\mathbf{x}} - \beta \mathbf{e}_1\|_2^2$$

where $\beta = \|\mathbf{b}\|_2$, and $\mathbf{x}^{(k)} = \mathbf{Y}_k \hat{\mathbf{x}}$.

Important property:

Singular values of $\mathbf{B}_k$ approximate large singular values of $\mathbf{A}$



GKB Properties

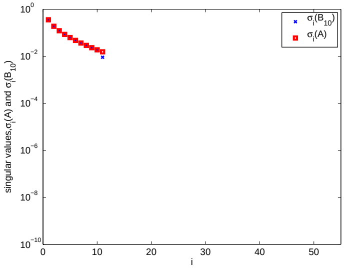
Using GKB iterates, can compute approximate solution $\mathbf{x}^{(k)}$

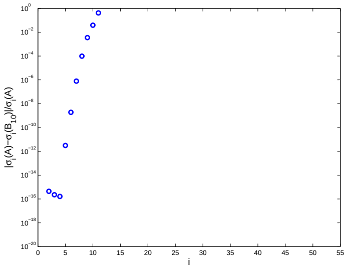
$$\min_{\mathbf{x}\in R(\mathbf{Y}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \min_{\hat{\mathbf{x}}} \|\mathbf{B}_k\hat{\mathbf{x}} - \beta\mathbf{e}_1\|_2^2$$

where $\beta = \|\mathbf{b}\|_2$, and $\mathbf{x}^{(k)} = \mathbf{Y}_k\hat{\mathbf{x}}$.

Important property:

Singular values of $\mathbf{B}_k$ approximate large singular values of $\mathbf{A}$





GKB Properties

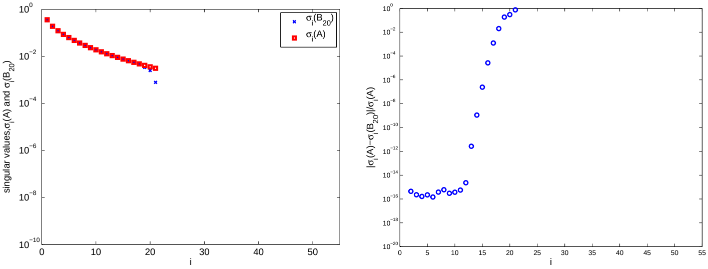
Using GKB iterates, can compute approximate solution $\mathbf{x}^{(k)}$

$$\min_{\mathbf{x}\in R(\mathbf{Y}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \min_{\hat{\mathbf{x}}} \|\mathbf{B}_k\hat{\mathbf{x}} - \beta\mathbf{e}_1\|_2^2$$

where $\beta = \|\mathbf{b}\|_2$, and $\mathbf{x}^{(k)} = \mathbf{Y}_k\hat{\mathbf{x}}$.

Important property:

Singular values of $\mathbf{B}_k$ approximate large singular values of $\mathbf{A}$



GKB Properties

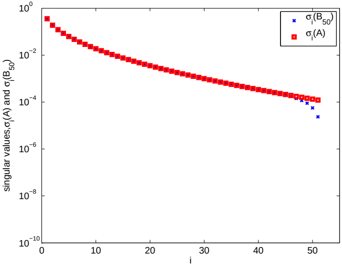
Using GKB iterates, can compute approximate solution $\mathbf{x}^{(k)}$

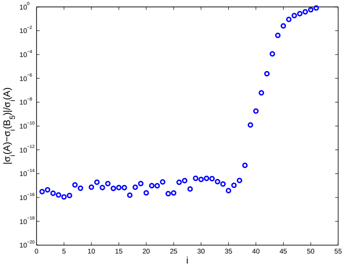
$$\min_{\mathbf{x} \in R(\mathbf{Y}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 = \min_{\hat{\mathbf{x}}} \|\mathbf{B}_k \hat{\mathbf{x}} - \beta \mathbf{e}_1\|_2^2$$

where $\beta = \|\mathbf{b}\|_2$, and $\mathbf{x}^{(k)} = \mathbf{Y}_k \hat{\mathbf{x}}$.

Important property:

Singular values of $\mathbf{B}_k$ approximate large singular values of $\mathbf{A}$





LSQR and Iterative Regularization

This property implies:

- Early iterations $\mathbf{x}^{(k)}$ in a subspace that approximates large singular components of $\mathbf{A}$.
- Thus for $k \ll n$, $\mathbf{x}^{(k)}$ is a regularized solution.
- Eventually $\mathbf{x}^{(k)} \to \mathbf{x}_{\text{inv}} = \mathbf{x}_{\text{true}} + \text{error}$ (bad)
- Iterative regularization $\Rightarrow$ determine good stopping iteration.

GKB-based Hybrid Methods

Hybrid approach:

- Enforce regularization at each GKB iteration
- Regularize projected least squares problem involving $\mathbf{B}_k$

$$\min_{\hat{\mathbf{x}}} \left\{ \|\mathbf{B}_k \hat{\mathbf{x}} - \beta \mathbf{e}_1\|_2^2 + \alpha^2 \|\hat{\mathbf{x}}\|_2^2 \right\}$$

- Can be done very cheaply

Some references:

O'Leary and Simmons, SISSC, 1981.
Björck, BIT 1988.
Björck, Grimme, and Van Dooren, BIT, 1994.
Larsen, PhD Thesis, 1998.
Hanke, BIT 2001.
Kilmer and O'Leary, SIMAX, 2001.
Kilmer, Hansen, Español, SISC 2007.
Hnětynkovà, Plešinger and Strakoš, BIT 2009
Chung, N, O'Leary, ETNA 2007

Hybrid Methods

Advantages of hybrid approach:

- Powerful regularization parameter choice methods can be implemented efficiently on the projected problem.
- Semi-convergence avoided, less sensitive to stopping iteration.
- Our implementation: HyBR
  - Can automatically choose regularization parameters (GCV)
  - Can automatically suggest stopping iteration

Arnoldi-based Hybrid Methods

- Instead of Golub-Kahan bidiagonalization (LSQR),
- Use Arnoldi Hessenberg reduction (GMRES)

  Calvetti, Morigi, Riechel, Sgallari, JCAM, 2000.
  Hochstenback, Reichel, J. Comput. Appl. Math., 2010.
  Reichel, Sgallari, Ye, Appl. Numer. Math, 2012.

- Arnoldi advantages:
  - Easier to implement for general Tikhonov regularization:

$$\mathbf{x}_m = \arg\min_{\mathbf{x}\in\mathcal{K}_m} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{L}\mathbf{x}\|_2^2$$

    Gazzola, Novati, 2013.
  - Flexible Krylov subspaces can be used to solve

$$\mathbf{x}_m = \arg\min_{\mathbf{x}\in\mathcal{K}_m} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{L}\mathbf{x}\|_p$$

    More on this later.

More General Regularization Methods

Consider the general problem:

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \mathcal{R}(\mathbf{x})$$

where

- $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_p^p = \sum |x_i|^p \,, \quad p \geq 1$
- $\mathcal{R}(x) = \left\| \sqrt{(D_h\mathbf{x})^2 + (D_v\mathbf{x})^2} \right\|_1 \quad$ (Total Variation)

Many Previous Works ...

S. Becker, J. Bobin, and E. Candès.
NESTA: A Fast and Accurate First-Order Method for Sparse Recovery.
SIAM J. Imaging Sciences, 4(1):1–39, 2011.

J.M. Bioucas-Dias and M.A.T. Figueiredo.
A new TwIST: two step iterative shrinkage/thresholding algorithms for image restoration.
IEEE Trans. Image Proc.,, 16 (2007), pp. 2992–3004.

S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinvesky.
An interior-point method for large-scale ℓ_1-regularized lest squares.
IEEE J. Selected Topics in Image Processing, 1 (2007), pp. 606–617.

J.P. Oliveira, J.M. Bioucas-Dias, M.A.T. Figueiredo.
Adaptive total vatiation image deblurring: A majorization-minimization approach.
Signal Processing, 89 (2009), pp. 1683–1693.

P. Rodríguez and B. Wohlberg.
An iteratively reweighted norm algorithm for total variation regularization.
In *Proceedings of the 40th Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2006.

S.J. Wright, R.D. Nowak, M.A.T. Figueiredo.
Sparse Reconstruction by Separable Approximation.
IEEE Transactions on Signal Processing, Vol. 57 No. 7 (2009), pp. 2479–2493.

Iteratively Reweighted Norm Approach (Wohlberg, Rodríguez)

- Iteratively construct $\mathbf{L}_m$ so that $\|\mathbf{L}_m\mathbf{x}\|_2^2 \approx \mathcal{R}(\mathbf{x})$, and compute

$$\mathbf{x}_m = \arg\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_m\|\mathbf{L}_m\mathbf{x}\|_2^2$$

- For example, $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_1$

$$\mathbf{L}_m = \mathrm{diag}\left(\frac{1}{\sqrt{|\mathbf{x}_{m-1}|}}\right) = \texttt{diag(1 ./ sqrt(abs(}\mathbf{x}_{m-1}\texttt{)))}$$

Arnoldi-based Hybrid Methods

Our approach: Similar to Wholberg and Rodriguez, combined with Arnoldi:

- Iteratively construct $\mathbf{L}_m$ so that $\|\mathbf{L}_m\mathbf{x}\|_2^2 \approx \mathcal{R}(\mathbf{x})$, and compute

$$\mathbf{x}_m = \arg\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_m \|\mathbf{L}_m\mathbf{x}\|_2^2$$

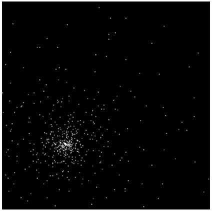
  using Arnoldi-based hybrid method.
- This is equivalent to flexible (variable) preconditioning:

$$\min_{\mathbf{x}} \|\mathbf{b} - \mathbf{A}\mathbf{L}_m^{-1}\mathbf{L}_m\mathbf{x}\|_2^2 + \lambda_m \|\mathbf{L}_m\mathbf{x}\|_2^2 \quad \Leftrightarrow \quad \left\{ \begin{array}{l} \min_{\tilde{\mathbf{x}}} \|\mathbf{b} - \tilde{\mathbf{A}}\,\tilde{\mathbf{x}}\|_2^2 + \lambda \|\tilde{\mathbf{x}}\|_2^2 \\ \tilde{\mathbf{A}} = \mathbf{A}\mathbf{L}_m^{-1}, \quad \tilde{\mathbf{x}} = \mathbf{L}_m\,\mathbf{x} \end{array} \right.$$

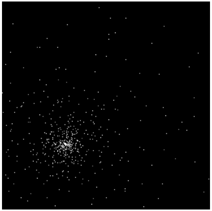
- Here, purpose of preconditioning:
  - not to improve condition number of the iteration matrix
  - instead, ensure iteration vector lies in "correct" subspace
- Use "flexible" Arnoldi-based hybrid method (similar to flexible GMRES).

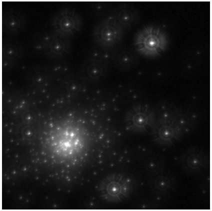
Sparsity Example

Sparsity Example

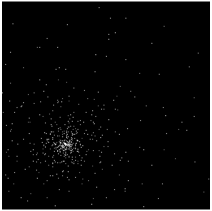
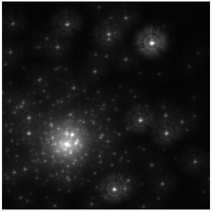
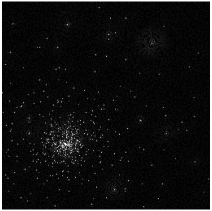


Sparsity Example

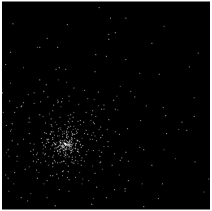
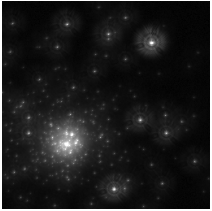
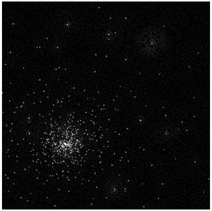




Sparsity Example

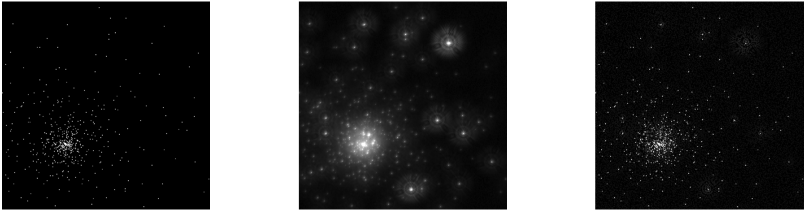




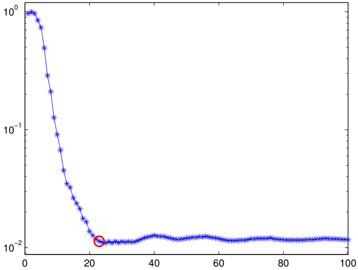
Sparsity Example

Stopping Iteration: 23 $\quad \widetilde{\lambda} = 1.1976 \cdot 10^{-4}$.

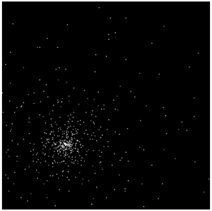
Sparsity Example

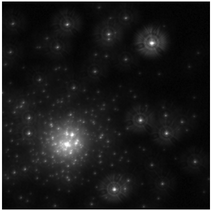


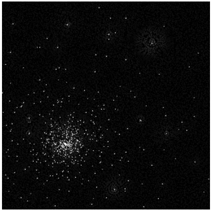


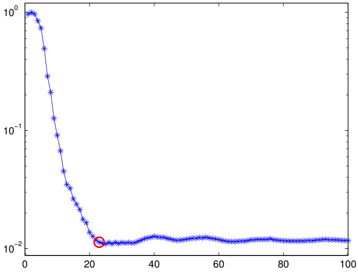
Stopping Iteration: 23 $\quad \widetilde{\lambda} = 1.1976 \cdot 10^{-4}$.

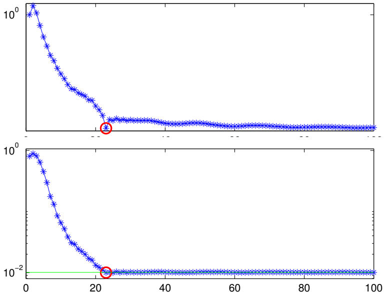
Sparsity Example











Stopping Iteration: 23 $\widetilde{\lambda} = 1.1976 \cdot 10^{-4}$.

Comparison with other method: Sparse Reconstructions

Method	Relative Error	Iterations	Total Time
SpaRSA	$2.2365 \cdot 10^{-2}$	94	24.76
NESTA	$1.7800 \cdot 10^{-2}$	248	306.17
TwIST	$1.1089 \cdot 10^{-2}$	104	28.02
l1_ls	$2.2257 \cdot 10^{-2}$	298	683.55
IRN-BPDN	$2.2294 \cdot 10^{-2}$	103	35.72
Flexi-AT	$1.1345 \cdot 10^{-2}$	23	2.44

SpaRSA: Wright, Nowak, Figueiredo, 2007
NESTA: Becker, Bobin, Candès, 2011
TwIST: Bioucas-Dias, Figueiredo, 2009
l1_ls: Kim, Koh, Lustig, Boyd, Gorinvesky, 2007
IRN-BPDN: Rodríguez, Wohlberg, 2009
Flexi-AT: Our method

Restarting Strategy

- For sparse reconstruction, $\mathbf{L}_m$ is diagonal $\Rightarrow$ it is easy to invert.
- In the Total Variation case,

$$\mathbf{L}_m = \mathbf{S}_m \mathbf{D}_{hv}$$

  is complicated, and not easy to invert.
- If $\mathbf{L}_m$ is not easy to invert, cost per iteration increases dramatically.
- So, we incorporate a restart strategy:
  - Restart when discrepancy principle is satisfied (residual reaches noise level).
  - Apply $\mathbf{L}_m$ at each restart.
  - Can also enforce projection constraints with each restart:

$$\mathbf{x}_0^{\text{new}} = \mathcal{P}(\mathbf{x}_k)$$

  where, e.g., $\mathcal{P}(\mathbf{x}_k) = \arg\min_{\mathbf{y}} \frac{1}{2}\|\mathbf{x}_k - \mathbf{y}\|_2^2$, s.t. $\mathbf{y} \geq \mathbf{0}$ and $\|\mathbf{y}\|_1 = v$

Comparison with other method: Sparse Reconstructions

Method	Relative Error	Iterations	Total Time
SpaRSA	$2.2365 \cdot 10^{-2}$	94	24.76
NESTA	$1.7800 \cdot 10^{-2}$	248	306.17
TwIST	$1.1089 \cdot 10^{-2}$	104	28.02
`l1_ls`	$2.2257 \cdot 10^{-2}$	298	683.55
IRN-BPDN	$2.2294 \cdot 10^{-2}$	103	35.72
Flexi-AT	$1.1345 \cdot 10^{-2}$	23	2.44
Flexi-AT (NN)	$3.7530 \cdot 10^{-3}$	60	6.25

AT: Standard Tikhonov regularization
SpaRSA: Wright, Nowak, Figueiredo, 2007
NESTA: Becker, Bobin, Candès, 2011
TwIST: Bioucas-Dias, Figueiredo, 2009
l1_ls: Kim, Koh, Lustig, Boyd, Gorinvesky, 2007
IRN-BPDN: Rodríguez, Wohlberg, 2009
Flexi-AT: Our method
Flexi-AT (NN): Our method with nonnegative projection restart

Concluding Remarks

- Hybrid iterative solvers can be effective.
  - Can use sparse constraint ($\| \cdot \|_1$) or TV regularization.
  - Requires flexible Krylov subspace framework.
  - Can incorporate regularization parameter choice methods and stopping criteria.
  - Restarting may be needed, but can be useful when enforcing projection constraints (e.g., nonnegativity).
- Current work:
  - Adapt approach to Calef's iteratively reweighted blind deconvolution.
  - Handles outliers (e.g., glints) in measure data.