Generalized Arnoldi-Tikhonov methods for Regularization and Sparse Reconstruction

Silvia Gazzola
Joint work with: **James Nagy, Paolo Novati**

Department of Mathematics
University of Padova - Italy

25th Biennial Numerical Analysis Conference

Outline

1 The Arnoldi-Tikhonov method
- The problem
- The standard AT and the generalized AT method
- The parameter choice strategy

Outline

1 The Arnoldi-Tikhonov method
- The problem
- The standard AT and the generalized AT method
- The parameter choice strategy

2 Sparse Reconstruction
- The problem
- The $\| \cdot \|_1$ case
- The TV case
- Examples and Comparisons

Outline

The problem

Consider

$$Ax = b, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^{N},$$

The problem

Consider

$$Ax = b, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^{N},$$

where

- A is assumed to have singular values that decay and cluster at zero;
- $b = b^{ex} + e = Ax^{ex} + e$;
- $\varepsilon \simeq \|e\|$ known.

The problem

Consider

$$Ax = b, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

where

- A is assumed to have singular values that decay and cluster at zero;
- $b = b^{ex} + e = Ax^{ex} + e$;
- $\varepsilon \simeq \|e\|$ known.

Our examples: Deblurring of Images

The problem

Consider

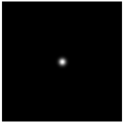
$$Ax = b, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

where

- A is assumed to have singular values that decay and cluster at zero;
- $b = b^{ex} + e = Ax^{ex} + e$;
- $\varepsilon \simeq \|e\|$ known.

Our examples: Deblurring of Images

$$\int K(s-t)\, f(t)dt$$



PSF

$*$

exact

The problem

Consider

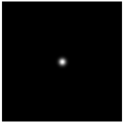
$$Ax = b, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^{N},$$

where

- A is assumed to have singular values that decay and cluster at zero;
- $b = b^{ex} + e = Ax^{ex} + e$;
- $\varepsilon \simeq \|e\|$ known.

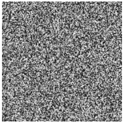
Our examples: Deblurring of Images

$$\int K(s-t)\,f(t)dt \qquad + \qquad n$$



$*$

$+$



PSF exact noise

The problem

Consider

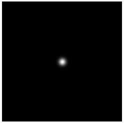
$$Ax = b, \quad A \in \mathbb{R}^{N \times N}, \quad b \in \mathbb{R}^{N},$$

where

- A is assumed to have singular values that decay and cluster at zero;
- $b = b^{ex} + e = Ax^{ex} + e$;
- $\varepsilon \simeq \|e\|$ known.

Our examples: Deblurring of Images

$$\int K(s-t)\,f(t)dt \qquad + \qquad n \qquad = \qquad g(s)$$

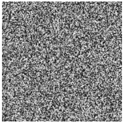


PSF

$*$

exact

$+$



noise

$=$

available

The Arnoldi-Tikhonov (AT) method

Given λ, L, x_0 consider the Tikhonov regularization

$$\min_{x \in \mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}.$$

[1] Calvetti-Morigi-Reichel-Sgallari, *Tikh. reg. and the L-curve for large discr. ill-posed pbs.*, J. Comp. Appl. Math., 2000.

The Arnoldi-Tikhonov (AT) method

Given λ, L, x_0 consider the Tikhonov regularization

$$\min_{x\in\mathbb{R}^N} \left\{\|b - Ax\|^2 + \lambda\|L(x - x_0)\|^2\right\}.$$

Consider the special case: $L = I_N$ and $x_0 = 0$.

Solution by the Arnoldi-Tikhonov method[1]: projection of A onto

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\}, \quad m \ll N.$$

[1] Calvetti-Morigi-Reichel-Sgallari, *Tikh. reg. and the L-curve for large discr. ill-posed pbs.*, J. Comp. Appl. Math., 2000.

The Arnoldi-Tikhonov (AT) method

Given λ, L, x_0 consider the Tikhonov regularization

$$\min_{x\in\mathbb{R}^N}\left\{\|b-Ax\|^2+\lambda\|L(x-x_0)\|^2\right\}.$$

Consider the special case: $L=I_N$ and $x_0=0$.

Solution by the Arnoldi-Tikhonov method[1]: projection of A onto

$$\mathcal{K}_m(A,b)=\text{span}\{b,Ab,\dots,A^{m-1}b\},\quad m\ll N.$$

The Krylov subspaces are built by the Arnoldi algorithm:

$$AV_m=V_{m+1}\bar{H}_m.$$

[1] Calvetti-Morigi-Reichel-Sgallari, *Tikh. reg. and the L-curve for large discr. ill-posed pbs.*, J. Comp. Appl. Math., 2000.

The Arnoldi-Tikhonov (AT) method

Given λ, L, x_0 consider the Tikhonov regularization

$$\min_{x\in\mathbb{R}^N} \left\{\|b - Ax\|^2 + \lambda\|L(x - x_0)\|^2\right\}.$$

Consider the special case: $L = I_N$ and $x_0 = 0$.

Solution by the Arnoldi-Tikhonov method[1]: projection of A onto

$$\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\}, \quad m \ll N.$$

The Krylov subspaces are built by the Arnoldi algorithm:

$$AV_m = V_{m+1}\bar{H}_m.$$

The AT method searches for approximations $x_{m,\lambda} \in \mathcal{K}_m(A, b)$, i.e. $x_{m,\lambda} = V_m y_{m,\lambda}$ where

$$y_{m,\lambda} = \text{argmin}_{y\in\mathbb{R}^m} \left\{\left\|\|b\|e_1 - \bar{H}_m y\right\|^2 + \lambda\left\|y\right\|^2\right\}.$$

[1] Calvetti-Morigi-Reichel-Sgallari, *Tikh. reg. and the L-curve for large discr. ill-posed pbs.*, J. Comp. Appl. Math., 2000.

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P \times N}$ and $x_0 \neq 0$.

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P \times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$
\begin{aligned}
y_{m,\lambda} &= \operatorname{argmin}_{y \in \mathbb{R}^m} \left\{ \left\| r_0 - A V_m y \right\|^2 + \lambda \left\| L V_m y \right\|^2 \right\} \\
&= \operatorname{argmin}_{y \in \mathbb{R}^m} \left\{ \left\| \left\| r_0 \right\| e_1 - \bar{H}_m y \right\|^2 + \lambda \left\| L V_m y \right\|^2 \right\} .
\end{aligned}
$$

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P \times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$
\begin{aligned}
y_{m,\lambda} &= \text{argmin}_{y \in \mathbb{R}^m} \left\{ \left\| r_0 - AV_m y \right\|^2 + \lambda \left\| LV_m y \right\|^2 \right\} \\
&= \text{argmin}_{y \in \mathbb{R}^m} \left\{ \left\| \left\| r_0 \right\| e_1 - \bar{H}_m y \right\|^2 + \lambda \left\| LV_m y \right\|^2 \right\}.
\end{aligned}
$$

Equivalent formulations:

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P\times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$\begin{aligned} y_{m,\lambda} &= \text{argmin}_{y\in\mathbb{R}^m} \left\{ \left\| r_0 - AV_m y \right\|^2 + \lambda \left\| LV_m y \right\|^2 \right\} \\ &= \text{argmin}_{y\in\mathbb{R}^m} \left\{ \left\| \left\| r_0 \right\| e_1 - \bar{H}_m y \right\|^2 + \lambda \left\| LV_m y \right\|^2 \right\}. \end{aligned}$$

Equivalent formulations:

- $$y_{m,\lambda} = \text{argmin}_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} L V_m \end{pmatrix} - \begin{pmatrix} \left\| r_0 \right\| e_1 \\ 0 \end{pmatrix} \right\|^2 .$$

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P\times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$\begin{aligned} y_{m,\lambda} &= \text{argmin}_{y\in\mathbb{R}^m} \left\{ \|r_0 - AV_m y\|^2 + \lambda \|LV_m y\|^2 \right\} \\ &= \text{argmin}_{y\in\mathbb{R}^m} \left\{ \left\| \|r_0\| e_1 - \bar{H}_m y \right\|^2 + \lambda \|LV_m y\|^2 \right\}. \end{aligned}$$

Equivalent formulations:

- $y_{m,\lambda} = \text{argmin}_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} LV_m \end{pmatrix} - \begin{pmatrix} \|r_0\| e_1 \\ 0 \end{pmatrix} \right\|^2.$
- $(\bar{H}_m^T \bar{H}_m + \hat{\lambda} V_m^T L^T LV_m) y_{m,\lambda} = \bar{H}_m^T \|r_0\| e_1$

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P \times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$\begin{aligned} y_{m,\lambda} &= \text{argmin}_{y \in \mathbb{R}^m} \left\{ \|r_0 - AV_m y\|^2 + \lambda \|LV_m y\|^2 \right\} \\ &= \text{argmin}_{y \in \mathbb{R}^m} \left\{ \left\| \|r_0\| e_1 - \bar{H}_m y \right\|^2 + \lambda \|LV_m y\|^2 \right\}. \end{aligned}$$

Equivalent formulations:

- $y_{m,\lambda} = \text{argmin}_{y \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} LV_m \end{pmatrix} - \begin{pmatrix} \|r_0\| e_1 \\ 0 \end{pmatrix} \right\|^2.$
- $(\bar{H}_m^T \bar{H}_m + \hat{\lambda} V_m^T L^T LV_m) y_{m,\lambda} = \bar{H}_m^T \|r_0\| e_1$

Remarks:

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P\times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$\begin{aligned} y_{m,\lambda} &= \text{argmin}_{y\in\mathbb{R}^m} \left\{ \left\| r_0 - AV_m y \right\|^2 + \lambda \left\| LV_m y \right\|^2 \right\} \\ &= \text{argmin}_{y\in\mathbb{R}^m} \left\{ \left\| \left\| r_0 \right\| e_1 - \bar{H}_m y \right\|^2 + \lambda \left\| LV_m y \right\|^2 \right\}. \end{aligned}$$

Equivalent formulations:

- $y_{m,\lambda} = \text{argmin}_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} LV_m \end{pmatrix} - \begin{pmatrix} \|r_0\| e_1 \\ 0 \end{pmatrix} \right\|^2.$
- $(\bar{H}_m^T \bar{H}_m + \hat{\lambda} V_m^T L^T L V_m) y_{m,\lambda} = \bar{H}_m^T \|r_0\| e_1$

Remarks:

- $\lambda = \lambda_m$;

The Generalized Arnoldi-Tikhonov (GAT) method

Incorporate $L \in \mathbb{R}^{P \times N}$ and $x_0 \neq 0$.
Looking for $x_{m,\lambda} \in x_0 + \mathcal{K}_m(A, r_0)$, i.e. $x_{m,\lambda} = x_0 + V_m y_{m,\lambda}$:

$$\begin{aligned} y_{m,\lambda} &= \text{argmin}_{y \in \mathbb{R}^m} \left\{ \|r_0 - AV_m y\|^2 + \lambda \|LV_m y\|^2 \right\} \\ &= \text{argmin}_{y \in \mathbb{R}^m} \left\{ \left\| \|r_0\| e_1 - \bar{H}_m y \right\|^2 + \lambda \|LV_m y\|^2 \right\}. \end{aligned}$$

Equivalent formulations:

- $y_{m,\lambda} = \text{argmin}_{y \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} LV_m \end{pmatrix} - \begin{pmatrix} \|r_0\| e_1 \\ 0 \end{pmatrix} \right\|^2.$
- $(\bar{H}_m^T \bar{H}_m + \hat{\lambda} V_m^T L^T LV_m) y_{m,\lambda} = \bar{H}_m^T \|r_0\| e_1$

Remarks:

- $\lambda = \lambda_m$;
- if $L \in \mathbb{R}^{N \times N}$ we could employ $L_m = V_m^T L V_m$

The parameter choice strategy

The parameter choice strategy

The discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrapprox 1.$$

The parameter choice strategy

The discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrapprox 1.$$

In the AT setting:

$$\begin{aligned}\phi_m(\lambda) &= \left\| \|r_0\|e_1 - \bar{H}_m y_{m,\lambda} \right\| \\ &= \left\| \|r_0\|e_1 - \bar{H}_m(\bar{H}_m^T \bar{H}_m + \lambda V_m^T L^T L V_m)^{-1} \bar{H}_m^T \|r_0\| e_1 \right\|.\end{aligned}$$

The parameter choice strategy

The discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrapprox 1.$$

In the AT setting:

$$\begin{aligned}\phi_m(\lambda) &= \left\| \|r_0\|e_1 - \bar{H}_m y_{m,\lambda} \right\| \\ &= \left\| \|r_0\|e_1 - \bar{H}_m(\bar{H}_m^T\bar{H}_m + \lambda V_m^T L^T L V_m)^{-1}\bar{H}_m^T\|r_0\|e_1 \right\|.\end{aligned}$$

Our choice, at each iteration m

$\phi_m(\lambda) \approx \phi_m(0) + \lambda\beta_m,$

The parameter choice strategy

The discrepancy principle is satisfied as soon as

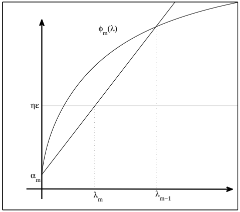
$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrsim 1.$$

In the AT setting:

$$\begin{aligned} \phi_m(\lambda) &= \left\| \|r_0\|e_1 - \bar{H}_m y_{m,\lambda} \right\| \\ &= \left\| \|r_0\|e_1 - \bar{H}_m(\bar{H}_m^T \bar{H}_m + \lambda V_m^T L^T L V_m)^{-1} \bar{H}_m^T \|r_0\| e_1 \right\|. \end{aligned}$$

Our choice, at each iteration m

$$\phi_m(\lambda) \approx \phi_m(0) + \lambda\beta_m,$$



The parameter choice strategy

The discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrapprox 1.$$

In the AT setting:

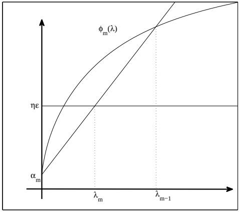
$$\begin{aligned}\phi_m(\lambda) &= \left\| \|r_0\| e_1 - \bar{H}_m y_{m,\lambda} \right\| \\ &= \left\| \|r_0\| e_1 - \bar{H}_m (\bar{H}_m^T \bar{H}_m + \lambda V_m^T L^T L V_m)^{-1} \bar{H}_m^T \|r_0\| e_1 \right\|.\end{aligned}$$

Our choice, at each iteration m

$$\phi_m(\lambda) \approx \phi_m(0) + \lambda\beta_m,$$

where

$$\beta_m \approx \frac{\phi_m(\lambda_{m-1}) - \alpha_m}{\lambda_{m-1}}.$$



The parameter choice strategy

The discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrsim 1.$$

In the AT setting:

$$\begin{aligned} \phi_m(\lambda) &= \left\| \|r_0\| e_1 - \bar{H}_m y_{m,\lambda} \right\| \\ &= \left\| \|r_0\| e_1 - \bar{H}_m (\bar{H}_m^T \bar{H}_m + \lambda V_m^T L^T L V_m)^{-1} \bar{H}_m^T \|r_0\| e_1 \right\|. \end{aligned}$$

Our choice, at each iteration m

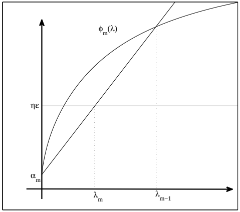
$$\phi_m(\lambda) \approx \phi_m(0) + \lambda\beta_m,$$

where

$$\beta_m \approx \frac{\phi_m(\lambda_{m-1}) - \alpha_m}{\lambda_{m-1}}.$$

Therefore

$$\lambda_m = \left| \frac{\eta\varepsilon - \alpha_m}{\phi_m(\lambda_{m-1}) - \alpha_m} \right| \lambda_{m-1}.$$



The parameter choice strategy

The discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) := \|b - Ax_{m,\lambda}\| \leq \eta\varepsilon, \quad \eta \gtrsim 1.$$

In the AT setting:

$$\begin{aligned}\phi_m(\lambda) &= \left\| \|r_0\| e_1 - \bar{H}_m y_{m,\lambda} \right\| \\ &= \left\| \|r_0\| e_1 - \bar{H}_m (\bar{H}_m^T \bar{H}_m + \lambda V_m^T L^T L V_m)^{-1} \bar{H}_m^T \|r_0\| e_1 \right\|.\end{aligned}$$

Our choice, at each iteration m

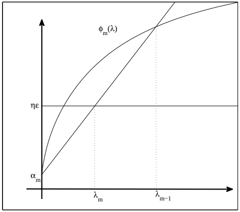
$$\phi_m(\lambda) \approx \phi_m(0) + \lambda\beta_m,$$

where

$$\beta_m \approx \frac{\phi_m(\lambda_{m-1}) - \alpha_m}{\lambda_{m-1}}.$$

Therefore

$$\lambda_m = \left| \frac{\eta\varepsilon - \alpha_m}{\phi_m(\lambda_{m-1}) - \alpha_m} \right| \lambda_{m-1}.$$



Secant update method

The problem

Better result obtained solving

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|_p^p + \lambda \ell(x) \right\}$$

The problem

Better result obtained solving

$$\min_{x\in\mathbb{R}^N}\left\{\|b-Ax\|_p^p+\lambda\ell(x)\right\}$$

- $p=2$;

The problem

Better result obtained solving

$$\min_{x\in\mathbb{R}^N}\left\{\|b - Ax\|_p^p + \lambda\ell(x)\right\}$$

- $p = 2$;
- $\ell(x) = \|x\|_q^q$

The problem

Better result obtained solving

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|_p^p + \lambda \ell(x) \right\}$$

- $p = 2$;
- $\ell(x) = \|x\|_q^q$ ($q = 1$ to force sparsity in the solution);

The problem

Better result obtained solving

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|_p^p + \lambda\ell(x) \right\}$$

- $p = 2$;
- $\ell(x) = \|x\|_q^q$ ($q = 1$ to force sparsity in the solution);
- $\ell(x) = TV(x) = \|\sqrt{(D_h x)^2 + (D_v x)^2}\|_1$.

The problem

Better result obtained solving

$$\min_{x\in\mathbb{R}^N} \left\{\|b - Ax\|_p^p + \lambda\ell(x)\right\}$$

- $p = 2$;
- $\ell(x) = \|x\|_q^q$ ($q = 1$ to force sparsity in the solution);
- $\ell(x) = TV(x) = \|\sqrt{(D_h x)^2 + (D_v x)^2}\|_1$.

ISSUE: nonlinearity!

The problem

Better result obtained solving

$$\min_{x \in \mathbb{R}^N} \left\{ \|b - Ax\|_p^p + \lambda \ell(x) \right\}$$

- $p = 2$;
- $\ell(x) = \|x\|_q^q$ ($q = 1$ to force sparsity in the solution);
- $\ell(x) = TV(x) = \| \sqrt{(D_h x)^2 + (D_v x)^2} \|_1$.

ISSUE: nonlinearity!

- P. Rodríguez, B. Wohlberg.
  An Efficient Algorithm for Sparse Representations with ℓ^p Data Fidelity Term.
  Proceedings of the 4th IEEE Andean Technical Conference (ANDESCON), 2008.
- B. Wohlberg, P. Rodriguez.
  An Iteratively Reweighted Norm Algorithm for Minimization of TV Functionals.
  IEEE Signal Processing Letters, 14 (2007), pp. 948–951.
- 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.

An *iteratively reweighted norm* approach

Exploiting the iterative setting of the AT method:

$$\ell(x) \approx \|\mathcal{W}_m x\|_2^2,$$

An *iteratively reweighted norm* approach

Exploiting the iterative setting of the AT method:

$$\ell(x) \approx \|\mathcal{W}_m x\|_2^2,$$

where

- $\ell(x) = \|x\|_1$ $$\mathcal{W}_m = L_m = \text{diag}\left(\frac{1}{\sqrt{|x_{m-1}|}}\right)$$

An *iterative reweighted norm* approach

Exploiting the iterative setting of the AT method:

$$\ell(x) \approx \|\mathcal{W}_m x\|_2^2,$$

where

- $\ell(x) = \|x\|_1$ $$\mathcal{W}_m = L_m = \mathrm{diag}\left(\frac{1}{\sqrt{|x_{m-1}|}}\right)$$
- $\ell(x) = TV(x) = \|\sqrt{(D_h x)^2 + (D_v x)^2}\|_1,\ x \in \mathbb{R}^N = \mathbb{R}^{n^2}$ $$\mathcal{W}_m = L_m = S_m D_{hv}$$

An *iteratively reweighted norm* approach

Exploiting the iterative setting of the AT method:

$$\ell(x) \approx \|\mathcal{W}_m x\|_2^2,$$

where

- $\ell(x) = \|x\|_1$
$$\mathcal{W}_m = L_m = \text{diag}\left(\frac{1}{\sqrt{|x_{m-1}|}}\right)$$
- $\ell(x) = TV(x) = \|\sqrt{(D_h x)^2 + (D_v x)^2}\|_1,\ x \in \mathbb{R}^N = \mathbb{R}^{n^2}$

$$\mathcal{W}_m = L_m = S_m D_{hv}$$

$$D = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}, \quad D_{hv} = \begin{pmatrix} D_h \\ D_v \end{pmatrix} = \begin{pmatrix} D \otimes I_n \\ I_n \otimes D \end{pmatrix}$$

An *iteratively reweighted norm* approach

Exploiting the iterative setting of the AT method:

$$\ell(x) \approx \|\mathcal{W}_m x\|_2^2,$$

where

- $\ell(x) = \|x\|_1$
$$\mathcal{W}_m = L_m = \mathrm{diag}\left(\frac{1}{\sqrt{|x_{m-1}|}}\right)$$
- $\ell(x) = TV(x) = \|\sqrt{(D_h x)^2 + (D_v x)^2}\|_1,\ x \in \mathbb{R}^N = \mathbb{R}^{n^2}$

$$\mathcal{W}_m = L_m = S_m D_{hv}$$

$$D = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n}, \quad D_{hv} = \begin{pmatrix} D_h \\ D_v \end{pmatrix} = \begin{pmatrix} D \otimes I_n \\ I_n \otimes D \end{pmatrix}$$

$$\widetilde{x}_{m-1} = D_{hv} x_{m-1}, \quad \widetilde{S}_m = \mathrm{diag}\left(\frac{1}{\sqrt[4]{\sum_{i=1}^{2(N-n)} (\widetilde{x}_{m-1})_i}}\right), \quad S_m = \begin{pmatrix} \widetilde{S}_m & 0 \\ 0 & \widetilde{S}_m \end{pmatrix}$$

The $\|\cdot\|_1$ case

Standard form transformation

$$\min_x \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}$$

[2] SAAD, *A flexible inner-outer prec. GMRES alg.*, SIAM J. Sci. Comp., 1993.

The $\|\cdot\|_1$ case

Standard form transformation

$$\min_x \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}$$

$$\min_{\widetilde{x}} \{ \|b - \widetilde{A}\widetilde{x}\|^2 + \widetilde{\lambda} \|\widetilde{x} - \widetilde{x}_0\|^2 \}$$

- $\widetilde{A} = AL^{-1}$
- $\widetilde{x}_0 = Lx_0$
- $x = L^{-1}\widetilde{x}$

[2] Saad, *A flexible inner-outer prec. GMRES alg.*, SIAM J. Sci. Comp., 1993.

The $\|\cdot\|_1$ case

Standard form transformation

$$\min_x \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}$$

$$\min_{\widetilde{x}} \{ \|b - \widetilde{A}\widetilde{x}\|^2 + \widetilde{\lambda} \|\widetilde{x} - \widetilde{x}_0\|^2 \}$$

- $\widetilde{A} = AL^{-1}$
- $\widetilde{x}_0 = Lx_0$
- $x = L^{-1}\widetilde{x}$

Features:

[2] Saad, *A flexible inner-outer prec. GMRES alg.*, SIAM J. Sci. Comp., 1993.

The $\|\cdot\|_1$ case

Standard form transformation

$$\min_x \left\{\|b - Ax\|^2 + \lambda\|L(x - x_0)\|^2\right\}$$

$$\min_{\widetilde{x}}\{\|b - \widetilde{A}\widetilde{x}\|^2 + \widetilde{\lambda}\|\widetilde{x} - \widetilde{x}_0\|^2\}$$

- $\widetilde{A} = AL^{-1}$
- $\widetilde{x}_0 = Lx_0$
- $x = L^{-1}\widetilde{x}$

Features:

- Matrix associated to the projected LS problem completely reduced

[2] SAAD, *A flexible inner-outer prec. GMRES alg.*, SIAM J. Sci. Comp., 1993.

The $\|\cdot\|_1$ case

Standard form transformation

$$\min_{x}\left\{\|b - Ax\|^2 + \lambda\|L(x - x_0)\|^2\right\}$$

$$\min_{\widetilde{x}}\{\|b - \widetilde{A}\widetilde{x}\|^2 + \widetilde{\lambda}\|\widetilde{x} - \widetilde{x}_0\|^2\}$$

- $\widetilde{A} = AL^{-1}$
- $\widetilde{x}_0 = Lx_0$
- $x = L^{-1}\widetilde{x}$

Features:

- Matrix associated to the projected LS problem completely reduced
- Preconditioned Krylov subspaces, with variable preconditioning

$$x_m = x_0 + Z_m y, \quad Z_m = span\{L_1^{-1}\check{v}_1, L_2^{-1}\check{v}_2, \dots, L_m^{-1}\check{v}_m\}$$

[2] SAAD, *A flexible inner-outer prec. GMRES alg.*, SIAM J. Sci. Comp., 1993.

The $\|\cdot\|_1$ case

Standard form transformation

$$\min_x \left\{\|b - Ax\|^2 + \lambda\|L(x - x_0)\|^2\right\}$$

$$\min_{\widetilde{x}}\{\|b - \widetilde{A}\widetilde{x}\|^2 + \widetilde{\lambda}\|\widetilde{x} - \widetilde{x}_0\|^2\}$$

- $\widetilde{A} = AL^{-1}$
- $\widetilde{x}_0 = Lx_0$
- $x = L^{-1}\widetilde{x}$

Features:

- Matrix associated to the projected LS problem completely reduced
- Preconditioned Krylov subspaces, with variable preconditioning

$$x_m = x_0 + Z_m y, \quad Z_m = span\{L_1^{-1}\check{v}_1, L_2^{-1}\check{v}_2, \dots, L_m^{-1}\check{v}_m\}$$

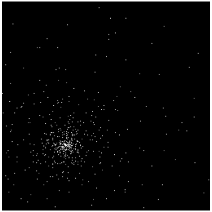
Flexible Arnoldi Algorithm[2]:

$$AZ_m = \check{V}_{m+1}\check{H}_m$$

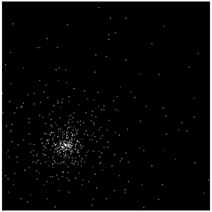
[2] Saad, *A flexible inner-outer prec. GMRES alg.*, SIAM J. Sci. Comp., 1993.

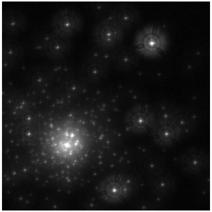
First Example - Star Cluster

First Example - Star Cluster

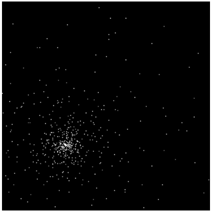


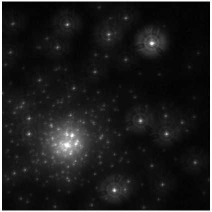
First Example - Star Cluster

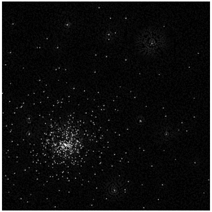




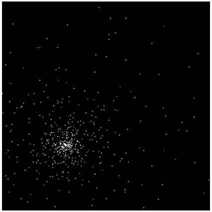
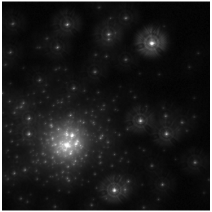
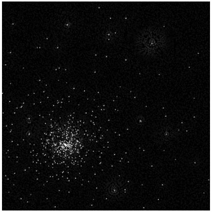
First Example - Star Cluster





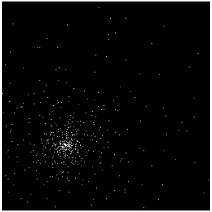
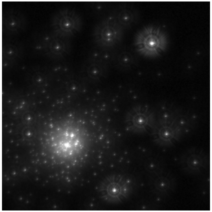
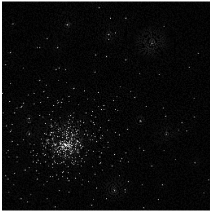


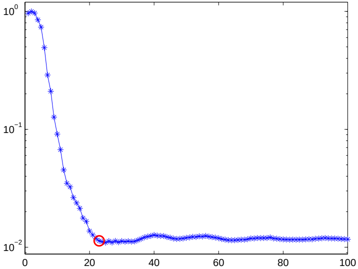
First Example - Star Cluster

Iteration #23; Relative Error $1.1349 \cdot 10^{-2}$; $\widetilde{\lambda} = 1.1976 \cdot 10^{-4}$.

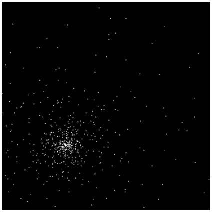
First Example - Star Cluster

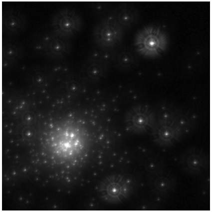




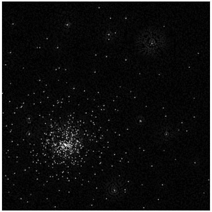


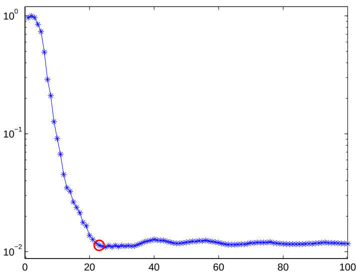
Iteration #23; Relative Error $1.1349 \cdot 10^{-2}$; $\widetilde{\lambda} = 1.1976 \cdot 10^{-4}$.

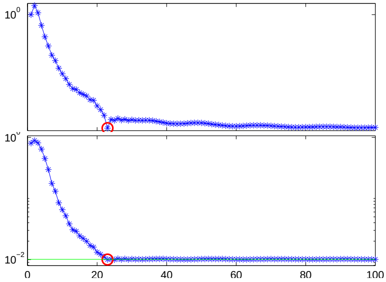
First Example - Star Cluster











Iteration #23; Relative Error $1.1349 \cdot 10^{-2}$; $\widetilde{\lambda} = 1.1976 \cdot 10^{-4}$.

The TV case

Problem: $L = S_m D_{hv}$ not easily "invertible".

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle
- "Wait" for the discrepancy principle to be satisfied

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle
- "Wait" for the discrepancy principle to be satisfied
- Initial guess, at the kth outer cycle

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\|r_0^{(k)}\right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle
- "Wait" for the discrepancy principle to be satisfied
- Initial guess, at the kth outer cycle
  - $x_0 = 0$;

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle
- "Wait" for the discrepancy principle to be satisfied
- Initial guess, at the kth outer cycle
  - $x_0 = 0$;
  - $x_0 = x_m^{(k)}$;

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \; x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle
- "Wait" for the discrepancy principle to be satisfied
- Initial guess, at the kth outer cycle
  - $x_0 = 0$;
  - $x_0 = x_m^{(k)}$;
  - $x_0 = P(x_m^{(k)})$, projection onto $\{x \in \mathbb{R}^N : (x)_i \geq 0, i = 1, \ldots, N\}$;

The *TV* case

Problem: $L = S_m D_{hv}$ not easily "invertible".

$$\min_{x\in\mathbb{R}^N} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}, \quad x_m \in x_0 + \mathcal{K}_m(A, r_0)$$

Given a meaningful initial guess $x_0^{(k)}$, restarting (inner-outer iterations)

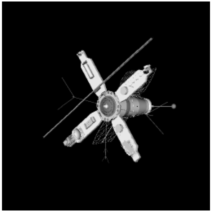
$$\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} S^{(k)} D_{hv} V_m \end{pmatrix} y - \begin{pmatrix} \left\| r_0^{(k)} \right\|_2 e_1 \\ 0 \end{pmatrix} \right\|^2, \quad x_m = x_0^{(k)} + V_m y$$

- Regularization Matrix updated at each outer cycle
- "Wait" for the discrepancy principle to be satisfied
- Initial guess, at the kth outer cycle
  - $x_0 = 0$;
  - $x_0 = x_m^{(k)}$;
  - $x_0 = P(x_m^{(k)})$, projection onto $\{x \in \mathbb{R}^N : (x)_i \geq 0, i = 1, \dots, N\}$;

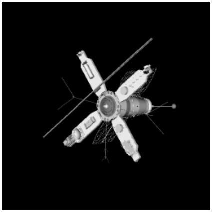
Main difference: unpreconditioned Krylov subspaces

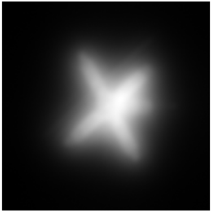
Second Example - Satellite

Second Example - Satellite

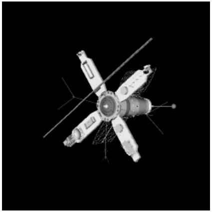
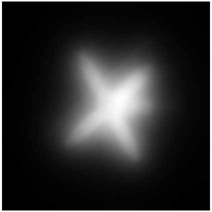
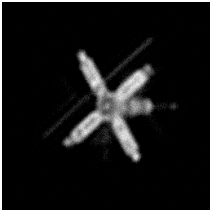


Second Example - Satellite

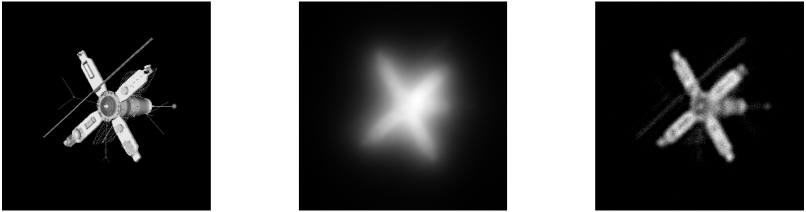




Second Example - Satellite

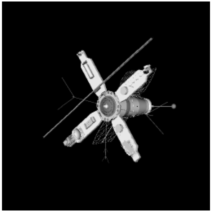




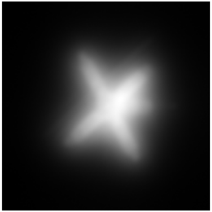
Second Example - Satellite

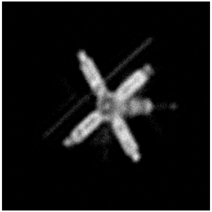


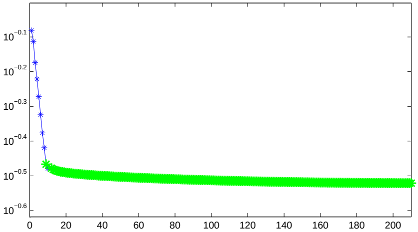
Outer Iterations: 200; Total Iterations: 210; Relative Error: $3.0115 \cdot 10^{-1}$.

Second Example - Satellite





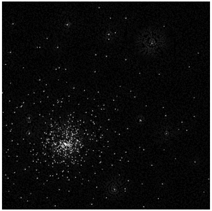




Outer Iterations: 200; Total Iterations: 210; Relative Error: $3.0115 \cdot 10^{-1}$.

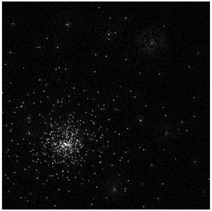
The $\|\cdot\|_1$ case revisited

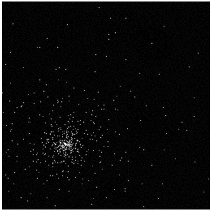
Std form transformation; update and nonnegativity at each restart.



The $\|\cdot\|_1$ case revisited

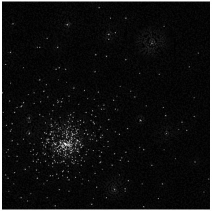
Std form transformation; update and nonnegativity at each restart.

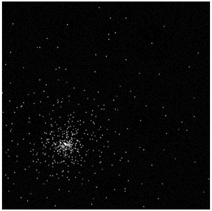


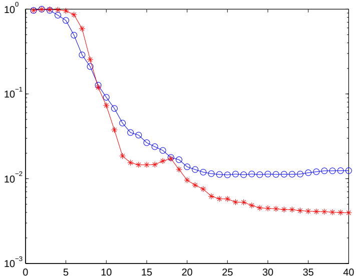


The $\|\cdot\|_1$ case revisited

Std form transformation; update and nonnegativity at each restart.

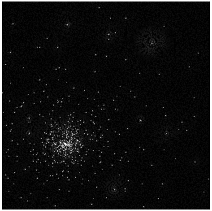


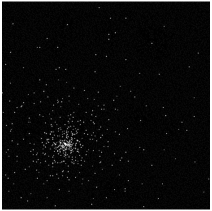


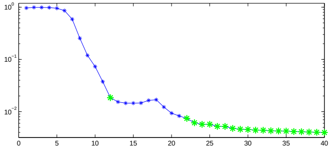


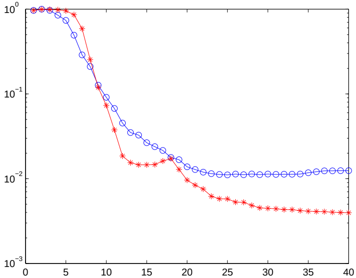
The $\|\cdot\|_1$ case revisited

Std form transformation; update and nonnegativity at each restart.



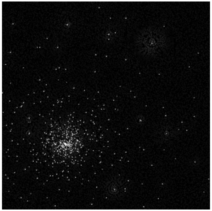


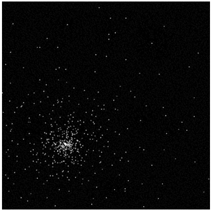


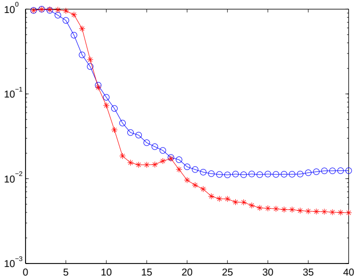


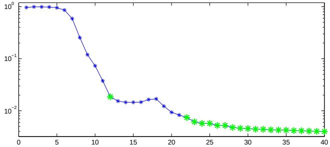
The $\|\cdot\|_1$ case revisited

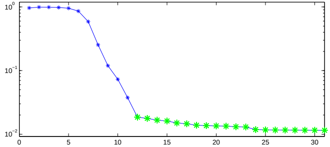
Std form transformation; update and nonnegativity at each restart.





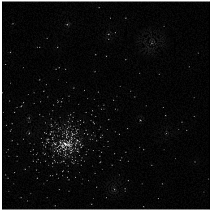


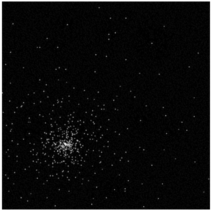


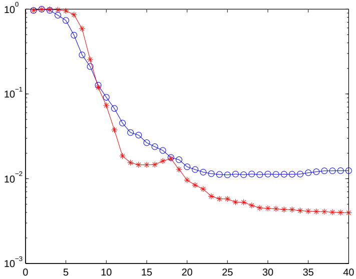


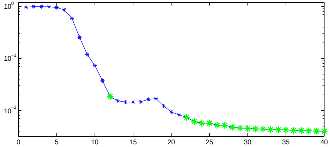
The $\|\cdot\|_1$ case revisited

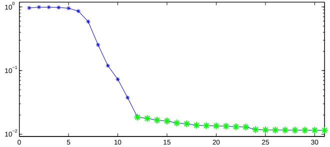
Std form transformation; update and nonnegativity at each restart.

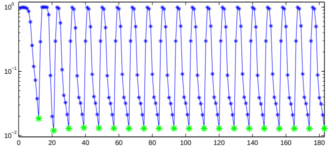












Comparisons

$$\ell(x) = \|x\|_1$$

Method	Relative Error	Iterations	Total Time	Average Time
SpaRSA	$1.1081 \cdot 10^{-2}$	343	45.16	0.13
TwIST	$1.1105 \cdot 10^{-2}$	102	16.39	0.16
l1_l2	$1.1146 \cdot 10^{-2}$	307	378.61	1.23
IRN-BPDN	$1.1146 \cdot 10^{-2}$	791	112.33	0.14
AT	$1.8609 \cdot 10^{-2}$	12	0.65	0.05
Flexi-AT	$1.1610 \cdot 10^{-2}$	100	5.03	0.05
NN-ReSt-GAT	$4.0606 \cdot 10^{-3}$	40	2.56	0.06

Comparisons

$$\ell(x) = \|x\|_1$$

Method	Relative Error	Iterations	Total Time	Average Time
SpaRSA	$1.1081 \cdot 10^{-2}$	343	45.16	0.13
TwIST	$1.1105 \cdot 10^{-2}$	102	16.39	0.16
l1_l2	$1.1146 \cdot 10^{-2}$	307	378.61	1.23
IRN-BPDN	$1.1146 \cdot 10^{-2}$	791	112.33	0.14
AT	$1.8609 \cdot 10^{-2}$	12	0.65	0.05
Flexi-AT	$1.1610 \cdot 10^{-2}$	100	5.03	0.05
NN-ReSt-GAT	$4.0606 \cdot 10^{-3}$	40	2.56	0.06

$$\ell(x) = TV(x)$$

Method	Relative Error	Iterations	Total Time	Average Time
aMM-TV	$2.8834 \cdot 10^{-1}$	631	4754.54	7.53
IRN-TV(0)	$3.2136 \cdot 10^{-1}$	418	15.65	0.04
IRN-TV(wr)	$3.2141 \cdot 10^{-1}$	190	7.66	0.04
NN-ReSt-TV	$3.0110 \cdot 10^{-1}$	210	11.36	0.05
AT	$3.4176 \cdot 10^{-1}$	9	0.17	0.02
GAT	$3.4809 \cdot 10^{-1}$	9	0.36	0.04

Final remarks

Final remarks

- simple and efficient (all the extra computations are performed in reduced dimension);

Final remarks

- simple and efficient (all the extra computations are performed in reduced dimension);
- simultaneously determine the regularization parameter and the number of iterations;

Final remarks

- simple and efficient (all the extra computations are performed in reduced dimension);
- simultaneously determine the regularization parameter and the number of iterations;
- Future Works: generalized to the multi-parameter case; dropping the hypothesis on $\|e\|$.

Final remarks

- simple and efficient (all the extra computations are performed in reduced dimension);
- simultaneously determine the regularization parameter and the number of iterations;
- Future Works: generalized to the multi-parameter case; dropping the hypothesis on $\|e\|$.

Thanks for your attention!