Fast Nonnegative Least Squares through flexible Krylov subspaces

Silvia Gazzola

Department of Mathematical Sciences

UNIVERSITY OF BATH

NL2A, CIRM (France)
October 28, 2016

What is this talk about?

What is this talk about?

Solution of

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

coming from suitable discretization of

$$\int_\Omega k(s,t)f(t)dt = g(s).$$

What is this talk about?

Solution of

$$\min_{x \in \mathbb{R}^N} \|b - Ax\|_2, \quad A \in \mathbb{R}^{N \times N}, \quad b \in \mathbb{R}^N,$$

coming from suitable discretization of

$$\int_\Omega k(s,t) f(t) dt = g(s).$$

Modeling inverse problems:

- the process k, the output g ($g = g^{ex} + \varepsilon$) are known;
- the input f is unknown.

What is this talk about?

Solution of

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

coming from suitable discretization of

$$\int_\Omega k(s,t)f(t)dt = g(s).$$

Modeling inverse problems:

- the process k, the output g ($g = g^{ex} + \varepsilon$) are known;
- the input f is unknown.

Ill-posed problems; appear in a variety of applications.

What is this talk about?

Solution of

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

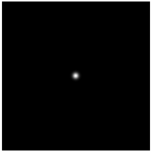
coming from suitable discretization of

$$\int_\Omega k(s,t)f(t)dt = g(s).$$

Modeling inverse problems:

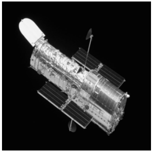
- the process k, the output g ($g = g^{ex} + \varepsilon$) are known;
- the input f is unknown.

Ill-posed problems; appear in a variety of applications.
An example: image deblurring and denoising.



PSF

$*$



exact

What is this talk about?

Solution of

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

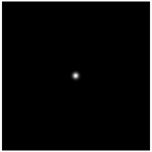
coming from suitable discretization of

$$\int_\Omega k(s,t)f(t)dt = g(s).$$

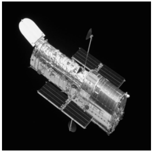
Modeling inverse problems:

- the process k, the output g ($g = g^{ex} + \varepsilon$) are known;
- the input f is unknown.

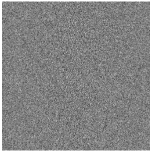
Ill-posed problems; appear in a variety of applications.
An example: image deblurring and denoising.


PSF

$*$


exact

$+$


noise

What is this talk about?

Solution of

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2, \quad A \in \mathbb{R}^{N\times N}, \quad b \in \mathbb{R}^N,$$

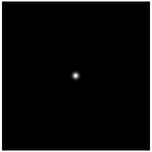
coming from suitable discretization of

$$\int_\Omega k(s,t)f(t)dt = g(s).$$

Modeling inverse problems:

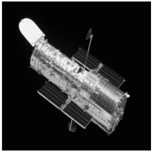
- the process k, the output g ($g = g^{ex} + \varepsilon$) are known;
- the input f is unknown.

Ill-posed problems; appear in a variety of applications.
An example: image deblurring and denoising.



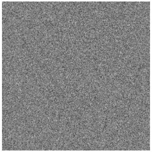
PSF

$*$



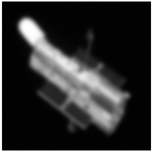
exact

$+$



noise

$=$



available

$$x = A^{\dagger} b$$

Regularization...

$$x = A^{\dagger} b$$

Regularization...

$x = A^{\dagger} b$

$x = (A^T A + \lambda I)^{-1} A^T b$

$x = A_k^{\dagger} b$

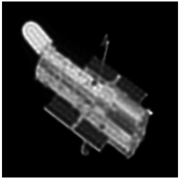
Regularization...

$x = A^{\dagger} b$ $\quad$ $x = (A^T A + \lambda I)^{-1} A^T b$ $\quad$ $x = A_k^{\dagger} b$



Nonnegative constraints!

$$\min_{x \geq 0} \|b - Ax\|_2$$



Outline

1 Nonnegative LS problems
- Brief survey of the available methods
- Krylov methods for unconstrained problems

2 Flexible Krylov subspaces
- The need of introducing "flexibility"
- Deriving FCGLS
- MFCGLS for nonnegative LS problems

3 Numerical Experiments
- Restoration of Astronomical Images
- CT Reconstruction

4 Looking at Poisson noise
- Modeling Poisson Noise
- CP-MFCGLS

5 Final Remarks

Popular approaches

Popular approaches

- Projected gradient

Popular approaches

- Projected gradient

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m)) \, .$$

Popular approaches

- Projected gradient
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m)) .$$

Popular approaches

- Projected gradient
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m)) .$$

- Modified Residual-Norm Steepest Descent (MRNSD)
  [Nagy and Strakos. Enforcing nonnegativity in image reconstruction algorithms, *Proc. SPIE*, 2000]

Popular approaches

- Projected gradient
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m)) .$$

- Modified Residual-Norm Steepest Descent (MRNSD)
  [Nagy and Strakos. Enforcing nonnegativity in image reconstruction algorithms, *Proc. SPIE*, 2000]

  From the KKT conditions

$$XA^T(b - Ax) = 0 , \quad \text{where} \quad X = \mathrm{diag}(x) , \quad x \geq 0 .$$

Popular approaches

- Projected gradient
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_{+}(x_m + \alpha_m A^T (b - Ax_m)) \,.$$

- Modified Residual-Norm Steepest Descent (MRNSD)
  [Nagy and Strakos. Enforcing nonnegativity in image reconstruction algorithms, *Proc. SPIE*, 2000]

  From the KKT conditions

$$XA^T(b - Ax) = 0\,, \quad \text{where} \quad X = \text{diag}(x)\,, \quad x \geq 0\,.$$

Fixed point iterations:

Popular approaches

- **Projected gradient**
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m))\,.$$

- **Modified Residual-Norm Steepest Descent (MRNSD)**
  [Nagy and Strakos. Enforcing nonnegativity in image reconstruction algorithms, *Proc. SPIE*, 2000]

  From the KKT conditions

$$XA^T(b - Ax) = 0\,, \quad \text{where} \quad X = \text{diag}(x)\,, \quad x \geq 0\,.$$

  Fixed point iterations:

$$x_{m+1} = x_m + \alpha_m \underbrace{X_m A^T(b - Ax_m)}_{=:\, d_m}\,, \text{ where } X_m = \text{diag}(x_m)\,,\ x_m \geq 0\,.$$

Popular approaches

- Projected gradient
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m)) .$$

- Modified Residual-Norm Steepest Descent (MRNSD)
  [Nagy and Strakos. Enforcing nonnegativity in image reconstruction algorithms, *Proc. SPIE*, 2000]
  From the KKT conditions

$$XA^T(b - Ax) = 0 , \quad \text{where} \quad X = \text{diag}(x) , \quad x \geq 0 .$$

Fixed point (or gradient descent) iterations:

$$x_{m+1} = x_m + \alpha_m \underbrace{X_m A^T(b - Ax_m)}_{=:\, d_m} , \text{ where } X_m = \text{diag}(x_m) , \ x_m \geq 0 .$$

Popular approaches

- Projected gradient
  [Beck and Teboulle. FISTA, *SIIMS*, 2009]

$$x_{m+1} = P_+(x_m + \alpha_m A^T(b - Ax_m)) .$$

- Modified Residual-Norm Steepest Descent (MRNSD)
  [Nagy and Strakos. Enforcing nonnegativity in image reconstruction algorithms, *Proc. SPIE*, 2000]

  From the KKT conditions

$$XA^T(b - Ax) = 0 , \quad \text{where} \quad X = \text{diag}(x) , \quad x \geq 0 .$$

  Fixed point (or gradient descent) iterations:

$$x_{m+1} = x_m + \alpha_m \underbrace{X_m A^T(b - Ax_m)}_{=: d_m} , \text{ where } X_m = \text{diag}(x_m) , \ x_m \geq 0 .$$

- Active set methods
  [Morigi, Plemmons, Reichel, Sgallari. Hybrid multilevel-active set method for box-constr. Calcolo, 2011]

Krylov methods for unconstrained problems

Unconstrained LS problem:

$$\min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2 \,.$$

Krylov methods for unconstrained problems

Unconstrained LS problem:

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 \,.$$

Iterative methods such that:

- $x_m \in \mathcal{K}_m(C,d)$, where

$$\mathcal{K}_m(C,d) = \text{span}\{d, Cd, \dots, C^{m-1}d\}\,,$$

  and $C = A,\ A^TA,\ AA^T,\ d = b,\ A^Tb,\ A^\ell b\,(\ell \geq 1)$.

- $r_m := b - Ax_m$ satisfies some conditions, e.g.,

$$\min_{\hat{x}\in\mathcal{K}_m(C,d)} \|b - A\hat{x}\|_2^2 \,.$$

Krylov methods for unconstrained problems

Unconstrained LS problem:

$$\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 .$$

Iterative methods such that:

- $x_m \in \mathcal{K}_m(C, d)$, where

$$\mathcal{K}_m(C,d) = \text{span}\{d, Cd, \dots, C^{m-1}d\} ,$$

  and $C = A,\ A^T A,\ AA^T,\ d = b,\ A^T b,\ A^\ell b\,(\ell \geq 1)$.

- $r_m := b - Ax_m$ satisfies some conditions, e.g.,

$$\min_{\hat{x}\in\mathcal{K}_m(C,d)} \|b - A\hat{x}\|_2^2 .$$

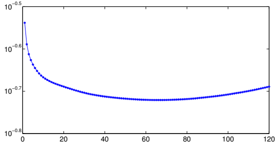
Extremely efficient!

Reichel & CO., Hansen & CO., Nagy & CO., Strakos & CO., Novati & CO [...]

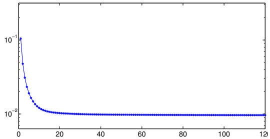
Lewis, Sgallari, Morigi, Lanza, Calvetti, Hanke, Donatelli, Chung, Elden, Simoncini, Jensen, Rodriguez, Russo, O'Leary, Plemmons, Jorgensen, Kilmer, Hnetynkova, Hochstenbach, Noschese, Dykes, Hayami, Ye, Saunders, Palmer, Huang, Jia [...]

Krylov methods (CGLS) in action

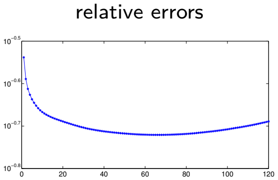
relative errors

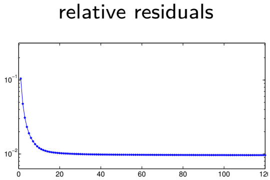


relative residuals

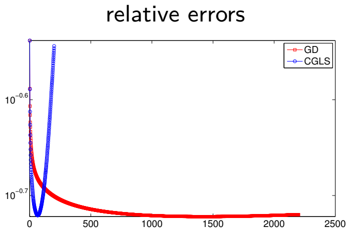


Krylov methods (CGLS) in action



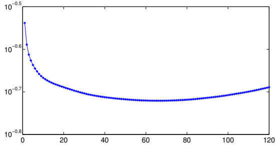


Gradient Descent approach VS. Krylov Subspaces approach

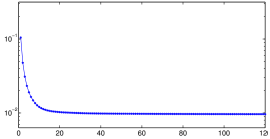


Krylov methods (CGLS) in action

relative errors

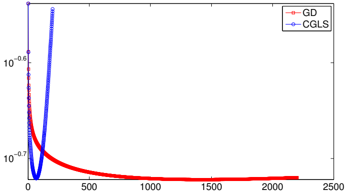


relative residuals



Gradient Descent approach VS. Krylov Subspaces approach

relative errors



They work efficiently as:

- they mimic the TSVD;
- $\mathcal{K}_m(C, d) \simeq \mathcal{K}_{m+1}(C, d)$ for small m.

Defining Krylov methods

[Axelsson. Iterative Solution Methods, *Cambridge*, 1994]

Defining Krylov methods

[Axelsson. Iterative Solution Methods, *Cambridge*, 1994]

Explicitly computing the solution along descent directions:

$$x_{m+1} = x_m + \sum_{j=0}^{m} \alpha_j^{(m)} d_j$$

$$d_{m+1} = \bar{z}_{m+1} + \sum_{j=0}^{m} \beta_j^{(m)} d_j \,.$$

Defining Krylov methods

[Axelsson. Iterative Solution Methods, *Cambridge*, 1994]

Explicitly computing the solution along descent directions:

$$x_{m+1} = x_m + \sum_{j=0}^{m} \alpha_j^{(m)} d_j$$

$$d_{m+1} = \bar{z}_{m+1} + \sum_{j=0}^{m} \beta_j^{(m)} d_j \,.$$

Requiring Ad_j orthogonal and minimal residual:

Defining Krylov methods

[Axelsson. Iterative Solution Methods, *Cambridge*, 1994]

Explicitly computing the solution along descent directions:

$$x_{m+1} = x_m + \sum_{j=0}^{m} \alpha_j^{(m)} d_j$$

$$d_{m+1} = \bar{z}_{m+1} + \sum_{j=0}^{m} \beta_j^{(m)} d_j \,.$$

Requiring Ad_j orthogonal and minimal residual:

- $x_{m+1} = x_m + \alpha_m d_m$, where $\alpha_m = \frac{(r_m, Ad_m)}{(Ad_m, Ad_m)}$;
- $\beta_j^{(m)} = -\frac{(A\bar{z}_{m+1}, Ad_j)}{(Ad_j, Ad_j)}\,, \quad j = 0, \ldots, m\,.$

Special case: CGLS.

The need of introducing "flexibility"

The need of introducing "flexibility"

Back the KKT conditions, most general case:

The need of introducing “flexibility”

Back the KKT conditions, most general case:

$$XA^T(b - Ax) = 0\,, \quad x \geq 0$$

The need of introducing "flexibility"

Back the KKT conditions, most general case:

$$XA^T(b - Ax_m) = 0\,, \quad x_m \geq 0$$

- make sure that $x_m \geq 0$ for each m.

The need of introducing "flexibility"

Back the KKT conditions, most general case:

$$X^{(m)}A^T(b - Ax_m) = 0\,, \quad x_m \geq 0$$

where, at the mth step, $X^{(m)} = \mathrm{diag}(x_{m-1})$.

- variable "preconditioners";
- make sure that $x_m \geq 0$ for each m.

The need of introducing "flexibility"

Back the KKT conditions, most general case:

$$X^{(m)} A^T (b - A x_m) = 0\,, \quad x_m \geq 0$$

where, at the mth step, $X^{(m)} = \text{diag}(x_{m-1})$.

We need a Krylov method (CGLS) that handles:

- variable "preconditioners";
- make sure that $x_m \geq 0$ for each m.

[Simoncini and Szyld. Recent computational developments in Krylov meth, *NLAA*, 2007]

Flexible Krylov methods for regularization

"Preconditioners" that enforce "regularity"

[Saad. FGMRES. SISC, 1993]

Flexible Krylov methods for regularization

"Preconditioners" that enforce "regularity"

[Saad. FGMRES. SISC, 1993]

- $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 + \lambda\|x\|_1 \quad \equiv \quad \min_{x\in\mathbb{R}^N} \|b - AX^{(m)}x\|_2^2 + \lambda\|x\|_2^2$

  [G. and Nagy. GAT for sparse reconstruction. SISC, 2014]

- $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 \quad \longrightarrow \quad x_m = Z_m y_m$

  [Morikuni, Reichel, Hayami. FGMRES for linear ill-posed pb. Appl.Numer.Math., 2014]

- $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 + \lambda\|Lx\|_2^2$
  $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 + \sum_{i=1}^{\ell} \lambda^{(i)}\|L^{(i)}x\|_2^2$

  [Reichel, Yu. Tikhonov regularization via flexible Arnoldi. BIT, 2015]

Flexible Krylov methods for regularization

"Preconditioners" that enforce "regularity"

[Saad. FGMRES. SISC, 1993]

- $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 + \lambda\|x\|_1 \quad \equiv \quad \min_{x\in\mathbb{R}^N} \|b - AX^{(m)}x\|_2^2 + \lambda\|x\|_2^2$

  [G. and Nagy. GAT for sparse reconstruction. SISC, 2014]

- $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 \quad \longrightarrow \quad x_m = Z_m y_m$

  [Morikuni, Reichel, Hayami. FGMRES for linear ill-posed pb. Appl.Numer.Math., 2014]

- $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 + \lambda\|Lx\|_2^2$
  $\min_{x\in\mathbb{R}^N} \|b - Ax\|_2^2 + \sum_{i=1}^{\ell} \lambda^{(i)}\|L^{(i)}x\|_2^2$

  [Reichel, Yu. Tikhonov regularization via flexible Arnoldi. BIT, 2015]

now... deriving FCGLS

If no preconditioners: CGLS

Input: A, b, x_0.
Initialize: $r_0 = b - Ax_0$, $\bar{z}_0 = A^T r_0$.

Take $d_0 = \bar{z}_0$.
For $m = 0, \dots,$ till a stopping criterion is satisfied

- Set $\alpha_m = \frac{(r_m, Ad_m)}{(Ad_m, Ad_m)}$.
- Update $x_{m+1} = x_m + \alpha_m d_m$.
- Update $r_{m+1} = r_m - \alpha_m Ad_m$.
- Compute $\bar{z}_{m+1} = A^T r_{m+1}$.

- Set $\beta_m = -\frac{(A\bar{z}_{m+1}, Ad_m)}{(Ad_m, Ad_m)}$.
- Update $d_{m+1} = \bar{z}_{m+1} + \beta_m d_m$.

If the preconditioner L is fixed: PCGLS

Input: A, L, b, x_0.
Initialize: $r_0 = b - Ax_0$, $z_0 = A^T r_0$.
Compute $\bar{z}_0 = Lz_0$.
Take $d_0 = \bar{z}_0$.
For $m = 0, \dots,$ till a stopping criterion is satisfied

- Set $\alpha_m = \frac{(r_m, Ad_m)}{(Ad_m, Ad_m)}$.
- Update $x_{m+1} = x_m + \alpha_m d_m$.
- Update $r_{m+1} = r_m - \alpha_m Ad_m$.
- Compute $z_{m+1} = A^T r_{m+1}$.

- Compute $\bar{z}_{m+1} = Lz_{m+1}$.
- Set $\beta_m = -\frac{(A\bar{z}_{m+1}, Ad_m)}{(Ad_m, Ad_m)}$.
- Update $d_{m+1} = \bar{z}_{m+1} + \beta_m d_m$.

If the "preconditioners" $L^{(m)}$ are variable: FCGLS

Input: A, $L^{(0)}$, b, x_0.
Initialize: $r_0 = b - Ax_0$, $z_0 = A^T r_0$.
Compute $\bar{z}_0 = L^{(0)} z_0$.
Take $d_0 = \bar{z}_0$.
For $m = 0, \ldots,$ till a stopping criterion is satisfied

- Set $\alpha_m = \frac{(r_m, Ad_m)}{(Ad_m, Ad_m)}$.
- Update $x_{m+1} = x_m + \alpha_m d_m$.
- Update $r_{m+1} = r_m - \alpha_m A d_m$.
- Compute $z_{m+1} = A^T r_{m+1}$.
- Compute $L^{(m+1)}$.
- Compute $\bar{z}_{m+1} = L^{(m+1)} z_{m+1}$.
- Set $\beta_j^{(m)} = -\frac{(A\bar{z}_{m+1}, Ad_j)}{(Ad_j, Ad_j)}$, $j = 0, \ldots, m$.
- Update $d_{m+1} = \bar{z}_{m+1} + \sum_{j=0}^{m} \beta_j^{(m)} d_j$.

[Notay. Flexible CG, *SISC*, 2000]

If the "preconditioners" $L^{(m)}$ are variable: FCGLS

Input: A, $L^{(0)}$, b, x_0.
Initialize: $r_0 = b - Ax_0$, $z_0 = A^T r_0$.
Compute $\bar{z}_0 = L^{(0)} z_0$.
Take $d_0 = \bar{z}_0$.
For $m = 0, \ldots,$ till a stopping criterion is satisfied

- Set $\alpha_m = \frac{(r_m, Ad_m)}{(Ad_m, Ad_m)}$.
- Update $x_{m+1} = x_m + \alpha_m d_m$.
- Update $r_{m+1} = r_m - \alpha_m A d_m$.
- Compute $z_{m+1} = A^T r_{m+1}$.
- Compute $L^{(m+1)}$.
- Compute $\bar{z}_{m+1} = L^{(m+1)} z_{m+1}$.
- Set $\beta_j^{(m)} = -\frac{(A\bar{z}_{m+1}, Ad_j)}{(Ad_j, Ad_j)}$, $j = \hat{m}, \ldots, m$.
- Update $d_{m+1} = \bar{z}_{m+1} + \sum_{j=\hat{m}}^{m} \beta_j^{(m)} d_j$.

[Notay. Flexible CG, *SISC*, 2000]

MFCGLS for nonnegative LS problems

MFCGLS for nonnegative LS problems

Solving:

$$X^{(m)}A^T(b - Ax) = 0\,, \quad x \geq 0\,, \quad \text{with} \quad X^{(m)} = \text{diag}(x_m)\,,$$

by FCGLS:

$$x_{m+1} = x_m + \alpha_m d_m\,.$$

MFCGLS for nonnegative LS problems

Solving:

$$X^{(m)}A^T(b - Ax) = 0\,, \quad x \geq 0\,, \quad \text{with} \quad X^{(m)} = \text{diag}(x_m)\,,$$

by FCGLS:

$$x_{m+1} = x_m + \alpha_m d_m\,.$$

To guarantee nonnegativity at each step:

$$\alpha_m = \min\left(\alpha_m, \min\left(-\frac{x_m(d_m < 0)}{d_m(d_m < 0)}\right)\right).$$

MFCGLS for nonnegative LS problems

Solving:

$$X^{(m)}A^T(b - Ax) = 0\,, \quad x \geq 0\,, \quad \text{with} \quad X^{(m)} = \text{diag}(x_m)\,,$$

by FCGLS:

$$x_{m+1} = x_m + \alpha_m d_m\,.$$

To guarantee nonnegativity at each step:

$$\alpha_m = \min\left(\alpha_m, \min\left(-\frac{x_m(d_m < 0)}{d_m(d_m < 0)}\right)\right).$$

Prone to stagnation: $\alpha_m = 0$!!!

MFCGLS for nonnegative LS problems

Solving:

$$X^{(m)}A^T(b - Ax) = 0\,, \quad x \geq 0\,, \quad \text{with} \quad X^{(m)} = \text{diag}(x_m)\,,$$

by FCGLS:

$$x_{m+1} = x_m + \alpha_m d_m\,.$$

To guarantee nonnegativity at each step:

$$\alpha_m = \min\left(\alpha_m, \min\left(-\frac{x_m(d_m < 0)}{d_m(d_m < 0)}\right)\right).$$

Prone to stagnation: $\alpha_m = 0$!!!

MFCGLS (Modified FCGLS)

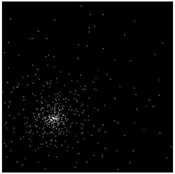
Input: A, b, $x_0^0 \geq 0$.

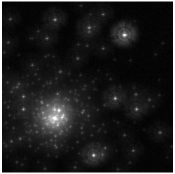
- For $k = 0, \dots,$ till a stopping criterion is satisfied
  - For $m = 0, \dots,$ till $m_{\max}$ or $\alpha_m = 0$
    - Run a FCGLS with initial guess x_0^k, preconditioner $X^{(m)}$, and NN:
      $x_{m+1} = x_m + \alpha_m d_m$;
      $d_{m+1} = \bar{z}_{m+1} + \sum_{j=\hat{m}}^{m} \beta_j^{(m)} d_j$.
    - Restart with the last approximation.

Numerical experiments

Numerical experiments

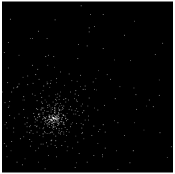
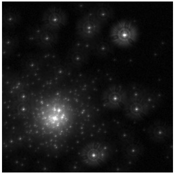
- star_cluster test problem [Nagy et al. *Restore Tools*, 2012]



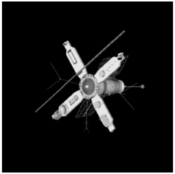
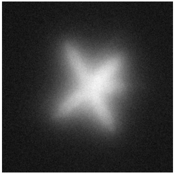


Numerical experiments

- star_cluster test problem [Nagy et al. *Restore Tools*, 2012]

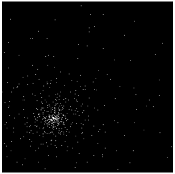



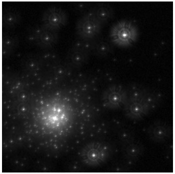
- satellite test problem [Nagy et al. *Restore Tools*, 2012]

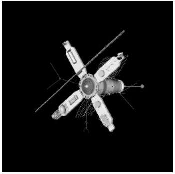
Numerical experiments

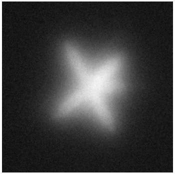
- star_cluster test problem [Nagy et al. *Restore Tools*, 2012]





- satellite test problem [Nagy et al. *Restore Tools*, 2012]





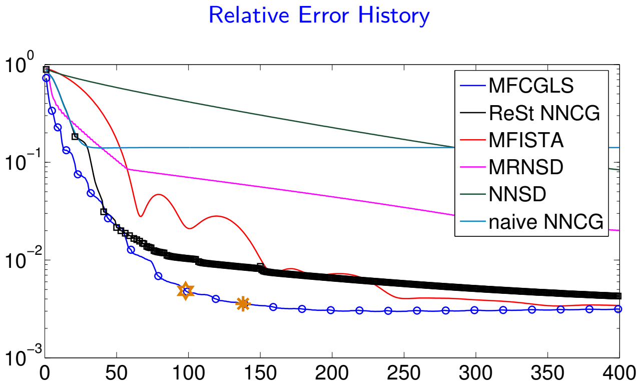
- paralleltomo test problem [Hansen et al. *AIR Tools*, 2012]

star_cluster test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-2}$

star_cluster test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-2}$



ReSt NNCG by

[Calvetti et al. Non-neg. and iterative methods for ill-posed pb, *Inv. Problems*, 2004]

star_cluster test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-2}$

	rel.error	iterations	tot.time	av.time
MFCGLS	2.8132e-03	248.67	62.56	0.25
ReSt NNCG	5.3699e-03	261.00	113.51	0.43
FISTA	9.1283e-02	72.00	42.06	0.58
MFISTA	3.2803e-03	400.00	216.11	0.54
MFISTA(0.2)	3.2445e-03	400.00	194.78	0.49
MFISTA(5)	4.2834e-03	400.00	185.22	0.46
MRNSD	1.9889e-02	400.00	91.11	0.23
NNSD	8.3206e-02	400.00	91.59	0.23
naive NNCG	1.4028e-01	400.00	105.02	0.26

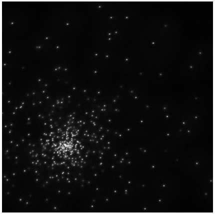
star_cluster test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-2}$, 200th iteration

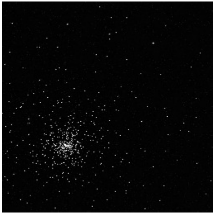
star_cluster test problem

$A \in \mathbb{R}^{65536\times 65536}$, $\widetilde{\varepsilon} = 10^{-2}$, 200th iteration

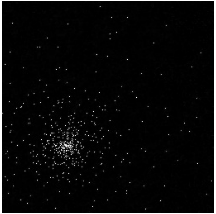
MRNSD



ReSt NNCG



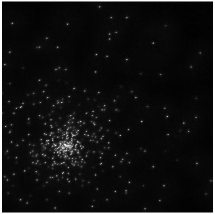
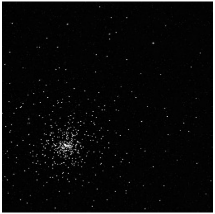
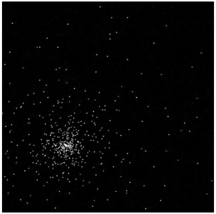
MFCGLS

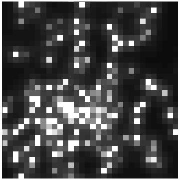
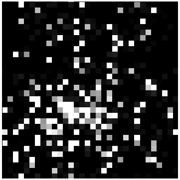
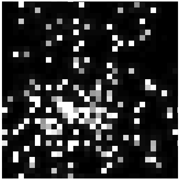


star_cluster test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-2}$, 200th iteration

MRNSD **ReSt NNCG** **MFCGLS**

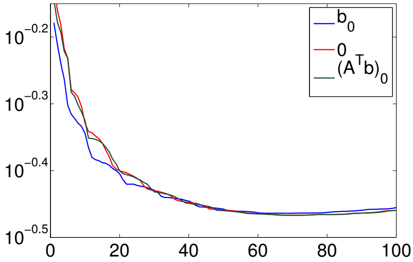




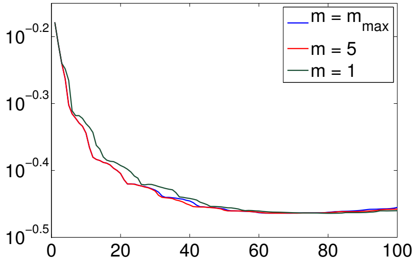
satellite test problem

satellite test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-1}$

Relative Error Histories, varying some "side" parameters

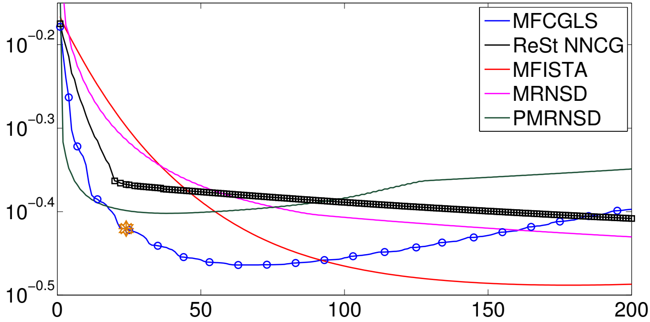




satellite test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-1}$

Relative Error History



ReSt NNCG by

[Hanke, Nagy, and Plemmons. Preconditioned iterative regularization, *Num.Lin.Alg.*, 1993]

satellite test problem

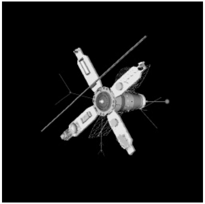
$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-1}$

	rel.error	iterations	tot.time	av.time
MFCGLS	3.5098e-01	70.33	5.49	0.08
ReSt NNCG	4.0957e-01	106.67	9.38	0.08
FISTA	3.2969e-01	164.33	21.22	0.12
MFISTA	3.2583e-01	177.00	23.10	0.13
MFISTA(0.2)	3.3318e-01	137.00	20.58	0.15
MFISTA(5)	3.3397e-01	200.00	26.86	0.13
MRNSD	3.7720e-01	200.00	12.55	0.06
PMRNSD	4.0032e-01	37.33	2.62	0.07
NNSD	4.3095e-01	200.00	13.82	0.07

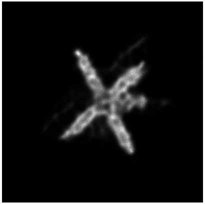
satellite test problem

$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} = 10^{-1}$

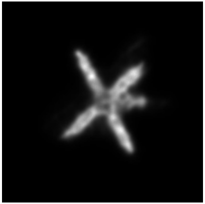
exact



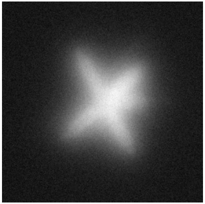
MFCGLS (# 84)



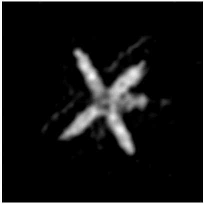
MRNSD (# 200)



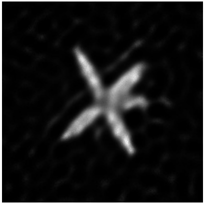
blurred & noisy



MFISTA (# 200)



PMRNSD (# 30)



`paralleltomo` test problem

paralleltomo test problem

overdetermined (81088×65536 coefficient matrix, $\widetilde{\varepsilon} = 5 \cdot 10^{-2}$)

	rel.error	iterations	tot.time	av.time
MFCGLS	1.8268e-01	17.33	0.92	0.09
ReSt NNCG	2.0133e-01	56.00	16.31	0.11
MFISTA	2.0029e-01	37.00	53.13	1.44
MRNSD	1.8506e-01	45.00	4.10	0.09
Cimmino	1.9982e-01	100.00	33.47	0.33

paralleltomo test problem

overdetermined (81088×65536 coefficient matrix, $\widetilde{\varepsilon} = 5 \cdot 10^{-2}$)

	rel.error	iterations	tot.time	av.time
MFCGLS	1.8268e-01	17.33	0.92	0.09
ReSt NNCG	2.0133e-01	56.00	16.31	0.11
MFISTA	2.0029e-01	37.00	53.13	1.44
MRNSD	1.8506e-01	45.00	4.10	0.09
Cimmino	1.9982e-01	100.00	33.47	0.33

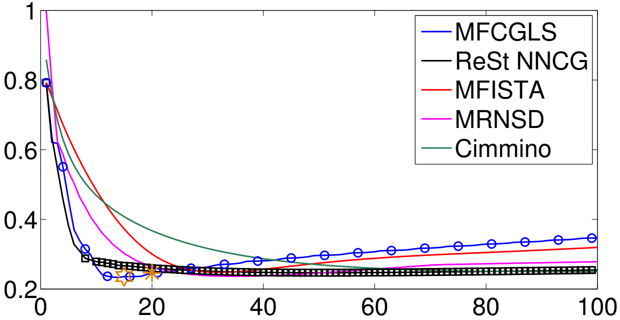
underdetermined (32580×65536 coefficient matrix, $\widetilde{\varepsilon} = 5 \cdot 10^{-2}$)

	rel.error	iterations	tot.time	av.time
MFCGLS	2.3145e-01	13.00	0.15	0.07
ReSt NNCG	2.4572e-01	51.00	0.59	0.05
MFISTA	2.4634e-01	32.00	12.41	0.39
MRNSD	2.3485e-01	35.00	3.43	0.09
Cimmino	2.4715e-01	94.33	8.84	0.09

underdetermined `paralleltomo` test problem

$A \in \mathbb{R}^{65160\times 65536}$, $\widetilde{\varepsilon} = 5 \cdot 10^{-2}$.

Relative Error History

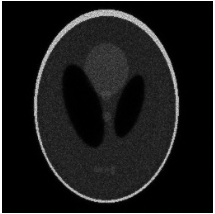


underdetermined `paralleltomo` test problem

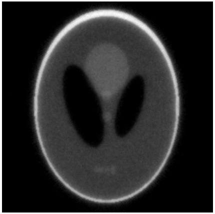
$A \in \mathbb{R}^{65160 \times 65536}$, $\widetilde{\varepsilon} = 5 \cdot 10^{-2}$, 17th iteration

exact

MFCGLS



MFISTA



Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta \mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta \mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta\mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]

$$\underbrace{b - \beta\mathbf{1}}_{=:\, b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta\mathbf{1} + \sigma^2\mathbf{1})}_{=:\, C_\eta}) \,.$$

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta\mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]
[Bardsley and Nagy. Covariance-prec. meth. for NN, *SIMAX*, 2006]

$$\underbrace{b - \beta\mathbf{1}}_{=:\, b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta\mathbf{1} + \sigma^2\mathbf{1})}_{=:\, C_\eta}) \,.$$

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta\mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]
[Bardsley and Nagy. Covariance-prec. meth. for NN, *SIMAX*, 2006]

$$\underbrace{b - \beta\mathbf{1}}_{=: b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta\mathbf{1} + \sigma^2\mathbf{1})}_{=: C_\eta}).$$

Problem to solve:

$$\min_{x \geq 0} \| C_\eta^{-1/2}(b_\beta - Ax)\|_2^2,$$

by the class of MRNSD methods:

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta\mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]
[Bardsley and Nagy. Covariance-prec. meth. for NN, *SIMAX*, 2006]

$$\underbrace{b - \beta\mathbf{1}}_{=:\, b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta\mathbf{1} + \sigma^2\mathbf{1})}_{=:\, C_\eta}) \,.$$

Problem to solve:

$$\min_{x\geq 0} \| C_\eta^{-1/2}(b_\beta - Ax)\|_2^2,$$

by the class of MRNSD methods:

- $$C_\eta = \text{diag}(b + \sigma^2\mathbf{1});$$

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta\mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]
[Bardsley and Nagy. Covariance-prec. meth. for NN, *SIMAX*, 2006]

$$\underbrace{b - \beta\mathbf{1}}_{=:\, b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta\mathbf{1} + \sigma^2\mathbf{1})}_{=:\, C_\eta})\,.$$

Problem to solve:

$$\min_{x \geq 0} \|C_\eta^{-1/2}(b_\beta - Ax)\|_2^2,$$

by the class of MRNSD methods:

- WMRNSD : $\quad C_\eta = \text{diag}(b + \sigma^2\mathbf{1});$

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta \mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]
[Bardsley and Nagy. Covariance-prec. meth. for NN, *SIMAX*, 2006]

$$\underbrace{b - \beta \mathbf{1}}_{=: b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta \mathbf{1} + \sigma^2 \mathbf{1})}_{=: C_\eta}).$$

Problem to solve:

$$\min_{x \geq 0} \| C_\eta^{-1/2}(b_\beta - Ax)\|_2^2,$$

by the class of MRNSD methods:

- WMRNSD : $C_\eta = \text{diag}(b + \sigma^2 \mathbf{1});$
- $C_\eta^{(k)} = \text{diag}(Ax_k + \beta \mathbf{1} + \sigma^2 \mathbf{1}).$

Incorporating Poisson noise

$$b = \text{Poisson}(Ax^{ex}) + \text{Poisson}(\beta \mathbf{1}) + \text{Normal}(0, \sigma^2 I)$$

[Snyder et al.. Image recovery from CCD, *J. Opt. Soc. Amer. A*, 1993]
[Bardsley and Nagy. Covariance-prec. meth. for NN, *SIMAX*, 2006]

$$\underbrace{b - \beta\mathbf{1}}_{=: b_\beta} = Ax^{ex} + \text{Normal}(0, \underbrace{\text{diag}(Ax^{ex} + \beta\mathbf{1} + \sigma^2\mathbf{1})}_{=: C_\eta}).$$

Problem to solve:

$$\min_{x \geq 0} \| C_\eta^{-1/2}(b_\beta - Ax)\|_2^2,$$

by the class of MRNSD methods:

- WMRNSD : $C_\eta = \text{diag}(b + \sigma^2\mathbf{1})$;
- KWMRNSD : $C_\eta^{(k)} = \text{diag}(Ax_k + \beta\mathbf{1} + \sigma^2\mathbf{1})$.

Introducing CP-MFCGLS

Input: A, b, $x_0^0 \geq 0$, $X^{(0)}$, $C_\eta^{(0)}$.

Initialize: $r_0^k = b - Ax_0^0$, $z_0^0 = A^T(C_\eta^{(0)})^{-1} r_0^0$, $\bar{z}_0^0 = X^{(0)} z_0^0$, $d_0^0 = \bar{z}_0^0$.

- For $k = 0, \ldots,$ till a stopping criterion is satisfied
  - For $m = 0, \ldots,$ till $m_{\max}$ or $\alpha_m = 0$ or a stopping criterion is satisfied
    - Run a FCGLS with x_0^k, preconditioners $X^{(m)}$, $C_\eta^{(k)}$, and NN:
      - Compute $\alpha_m = \frac{(\bar{r}_m^k, (C_\eta^{(k)})^{-1/2} A d_m^k)}{((C_\eta^{(k)})^{-1/2} A d_m^k, (C_\eta^{(k)})^{-1/2} A d_m^k)}$.
      - Update $x_{m+1}^k = x_m^k + \alpha_m d_m^k$.
      - Update $r_{m+1}^k = r_m^k - \alpha_m A d_m^k$.
      - Compute $X^{(m+1)}$.
      - Compute $z_{m+1}^k = A^T (C_\eta^{(k)})^{-1} r_{m+1}^k$.
      - Compute $\bar{z}_{m+1}^k = X^{(m+1)} z_{m+1}^k$.
      - Set $\beta_j = -\frac{((C_\eta^{(k)})^{-1/2} A z_{m+1}^k, (C_\eta^{(k)})^{-1/2} A d_m^k)}{((C_\eta^{(k)})^{-1/2} A d_m^k, (C_\eta^{(k)})^{-1/2} A d_m^k)}$, $j = \hat{m}, \ldots, m$.
      - Update $d_{m+1}^k = \bar{z}_{m+1}^k + \sum_{j=\hat{m}}^{m} \beta_j d_j^k$.
    - Restart with the last approximation, and update $C_\eta^{(k+1)}$.

satellite test problem, Gaussian and Poisson noise

satellite test problem, Gaussian and Poisson noise

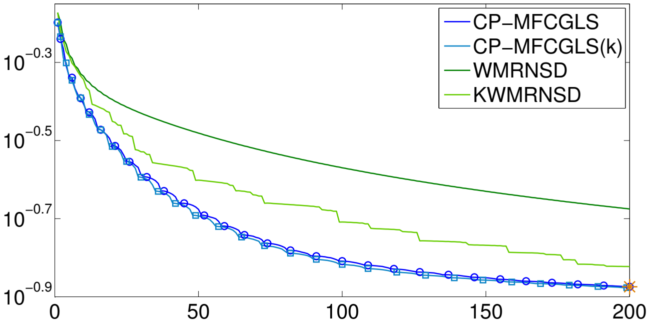
$A \in \mathbb{R}^{65536 \times 65536}$
Gauss: $\sigma = 20$, Poisson: $\beta = 60$, $\widetilde{\varepsilon} \simeq 1.5 \cdot 10^{-2}$

satellite test problem, Gaussian and Poisson noise

$A \in \mathbb{R}^{65536 \times 65536}$

Gauss: $\sigma = 20$, Poisson: $\beta = 60$, $\widetilde{\varepsilon} \simeq 1.5 \cdot 10^{-2}$

Relative Error History



satellite test problem, Gaussian and Poisson noise

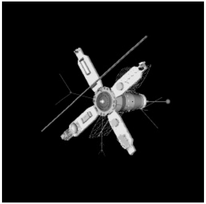
$A \in \mathbb{R}^{65536 \times 65536}$, $\widetilde{\varepsilon} \simeq 1.5 \cdot 10^{-2}$

	rel.error	iterations	tot.time	av.time
CP-MFCGLS	1.2785e-01	300.00	31.65	0.08
CP-MFCGLS(k)	1.2778e-01	300.00	32.17	0.08
WMRNSD	1.8201e-01	300.00	28.34	0.09
KWMRNSD	1.3590e-01	300.00	37.19	0.12

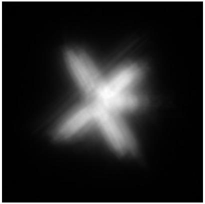
satellite test problem, Gaussian and Poisson noise

$A \in \mathbb{R}^{65536 \times 65536}$, $\tilde{\varepsilon} \simeq 1.5 \cdot 10^{-2}$, 100th iteration

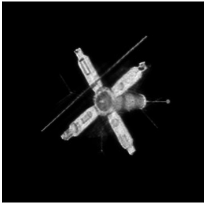
exact



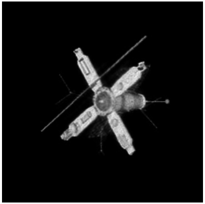
blurred & noisy



KWMRNSD



VP-MFCGLS



Final Remarks

Final Remarks

- Systematic and efficient way
  to enforce nonnegativity within Krylov subspace methods.

Final Remarks

- Systematic and efficient way
  to enforce nonnegativity within Krylov subspace methods.
- Embraces and improves many methods already available.

Final Remarks

- Systematic and efficient way
  to enforce nonnegativity within Krylov subspace methods.
- Embraces and improves many methods already available.
- Possible generalizations:
  - box constraints;
  - sparsity (?);
  - other Krylov methods (FGCR...);
  - ...

Final Remarks

- Systematic and efficient way
  to enforce nonnegativity within Krylov subspace methods.
- Embraces and improves many methods already available.
- Possible generalizations:
  - box constraints;
  - sparsity (?);
  - other Krylov methods (FGCR...);
  - ...

Paper available on arXiv:
S. Gazzola and Y. Wiaux.
Fast nonnegative least squares through flexible Krylov subspaces.
arXiv:1511.06269

Final Remarks

- Systematic and efficient way
  to enforce nonnegativity within Krylov subspace methods.
- Embraces and improves many methods already available.
- Possible generalizations:
  - box constraints;
  - sparsity (?);
  - other Krylov methods (FGCR...);
  - ...

Paper available on arXiv:
S. Gazzola and Y. Wiaux.
Fast nonnegative least squares through flexible Krylov subspaces.
arXiv:1511.06269

Thanks for your attention!