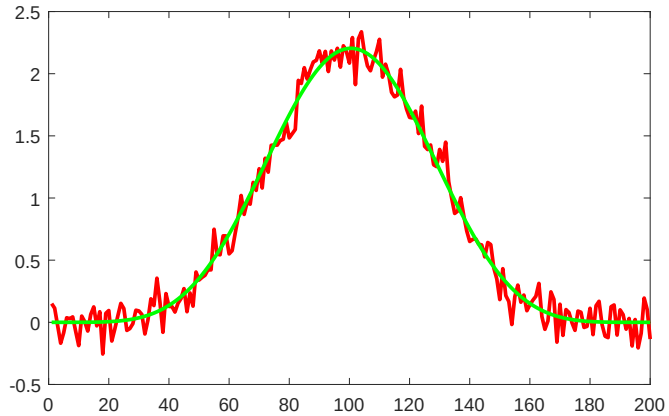
Projected Newton method for the regularization of ill-posed linear inverse problems

Sparse Days Meeting 2020, November 23 - 24



Applied Mathematics

Jeffrey Cornelis[†], N. Schenkels, W. Vanroose

That's me!

[†]mail: jeffrey.cornelis@uantwerpen.be

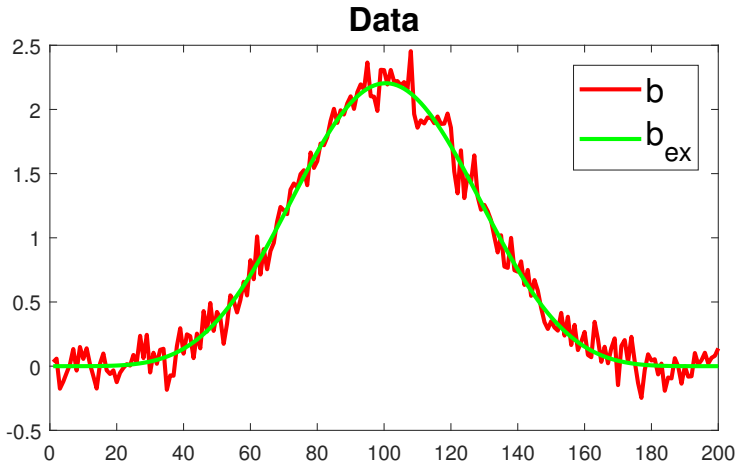
Universiteit Antwerpen

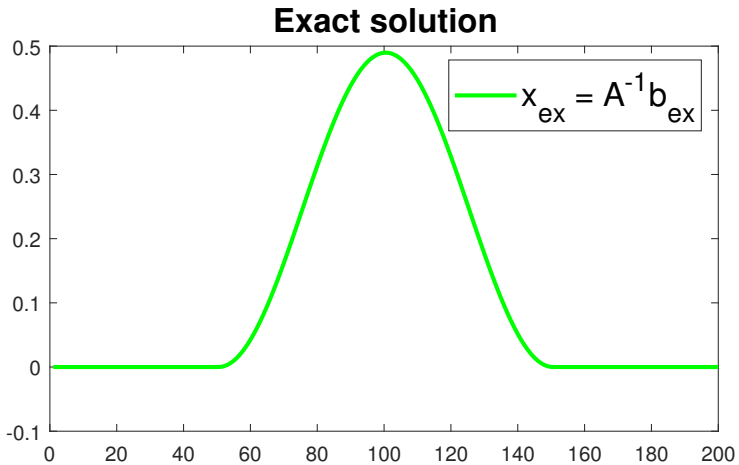
Need for regularization

Consider a linear system of equations

$$Ax = b = \underbrace{b_{ex}}_{=Ax_{ex}} + \underbrace{e}_{\sim\mathcal{N}(0,\sigma^2 I)}$$

with matrix $A \in \mathbb{R}^{m\times n}$ and data $b \in \mathbb{R}^m$.





`phillips` from `RegTools` with $\kappa(A) \approx 10^7$.

Need for regularization

Consider a linear system of equations

$$Ax = b = \underbrace{b_{ex}}_{=Ax_{ex}} + \underbrace{e}_{\sim\mathcal{N}(0,\sigma^2 I)}$$

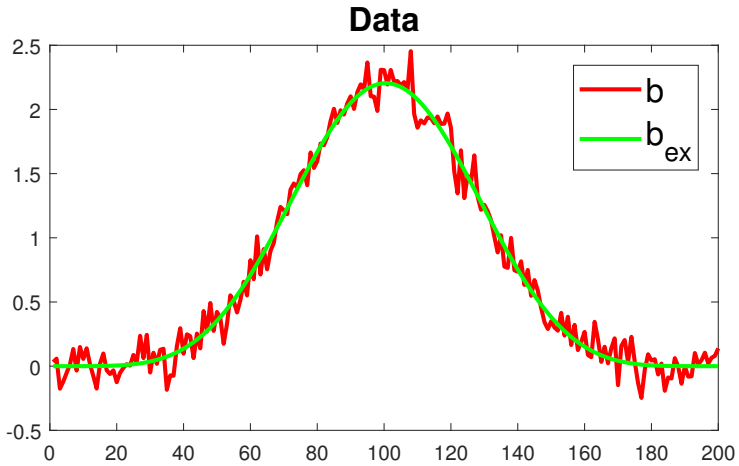
with matrix $A \in \mathbb{R}^{m\times n}$ and data $b \in \mathbb{R}^m$.

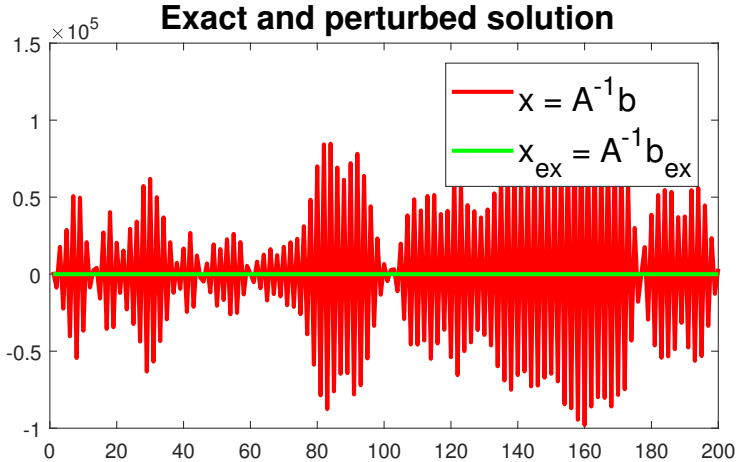
For square nonsingular matrices we have

$$\frac{||\delta x||_2}{||x||_2} \leq \underbrace{||A^{-1}||_2||A||_2}_{=\kappa(A)} \frac{||e||_2}{||b||_2}$$

with $\delta x := x - x_{ex}$.

Small changes in b can lead to huge changes in x for ill-conditioned matrix A.





`phillips` from `RegTools` with $\kappa(A) \approx 10^7$.

Tikhonov regularization

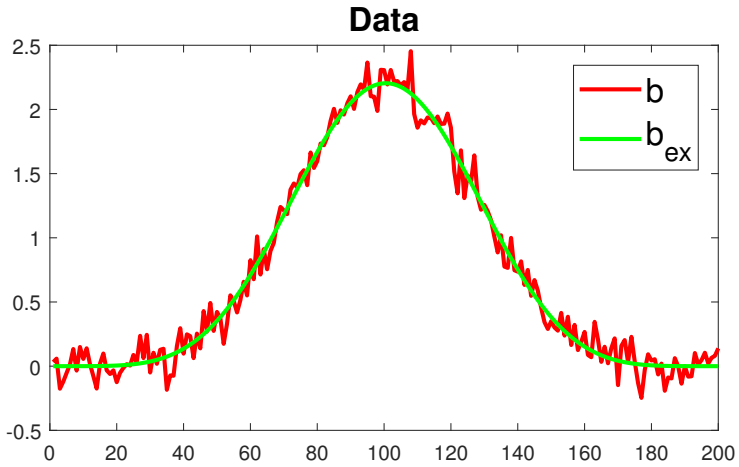
For rectangular matrices ($m \geq n$) we solve the linear least squares problem

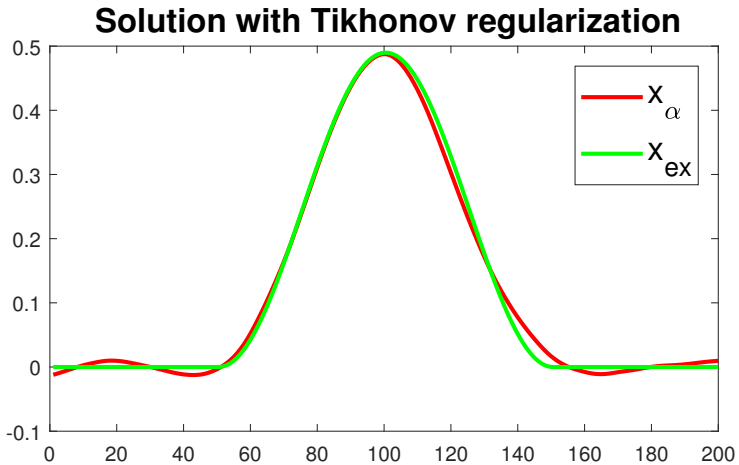
$$x = \arg\min_{x \in \mathbb{R}^n} ||Ax - b||_2^2 = (A^T A)^{-1} A^T b.$$

Sensitivity now related to $\kappa(A^T A)$.

Tikhonov regularization is a popular approach to obtain a meaningful solution:

$$\begin{aligned} x_\alpha &= \arg\min_{x \in \mathbb{R}^n} ||Ax - b||_2^2 + \alpha ||x||_2^2 \\ &= \left(A^T A + \alpha I\right)^{-1} A^T b. \end{aligned}$$





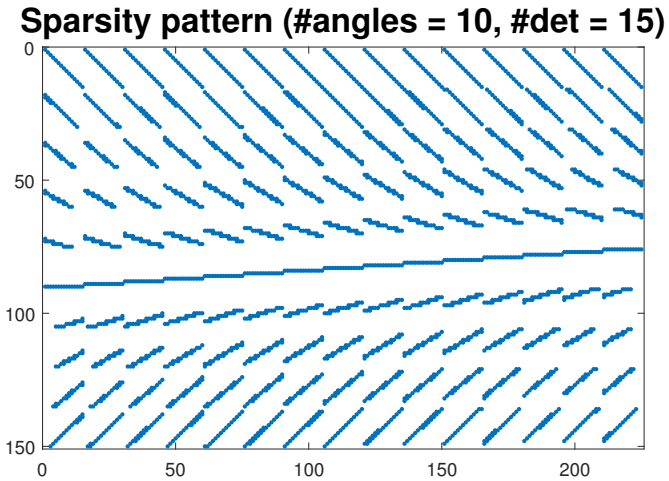
`phillips` from `RegTools` with $\kappa(A) \approx 10^7$.

An example from Computed Tomography

Consider a **Computed Tomography** example with 270 projection angles and 256 detectors (ASTRA toolbox):

$$Ax = b = b_{ex} + e = Ax_{ex} + e$$

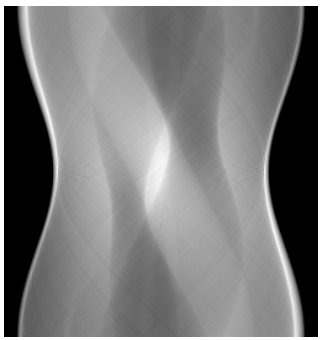
with $m = 69,120$ and $n = 65,536$.



x_{ex}

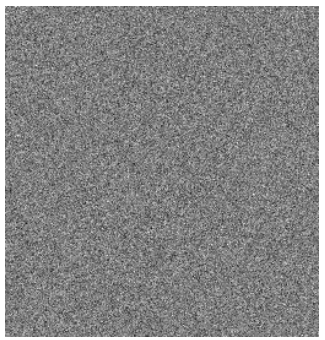
Exact image

$Ax_{ex} = b_{ex}$



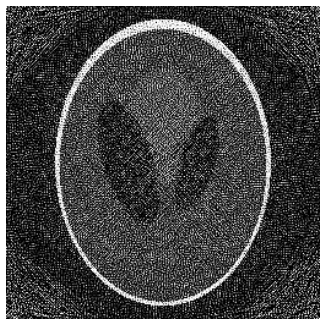
Exact sinogram

e



1% Gaussian noise

$\arg\min ||Ax - b||$



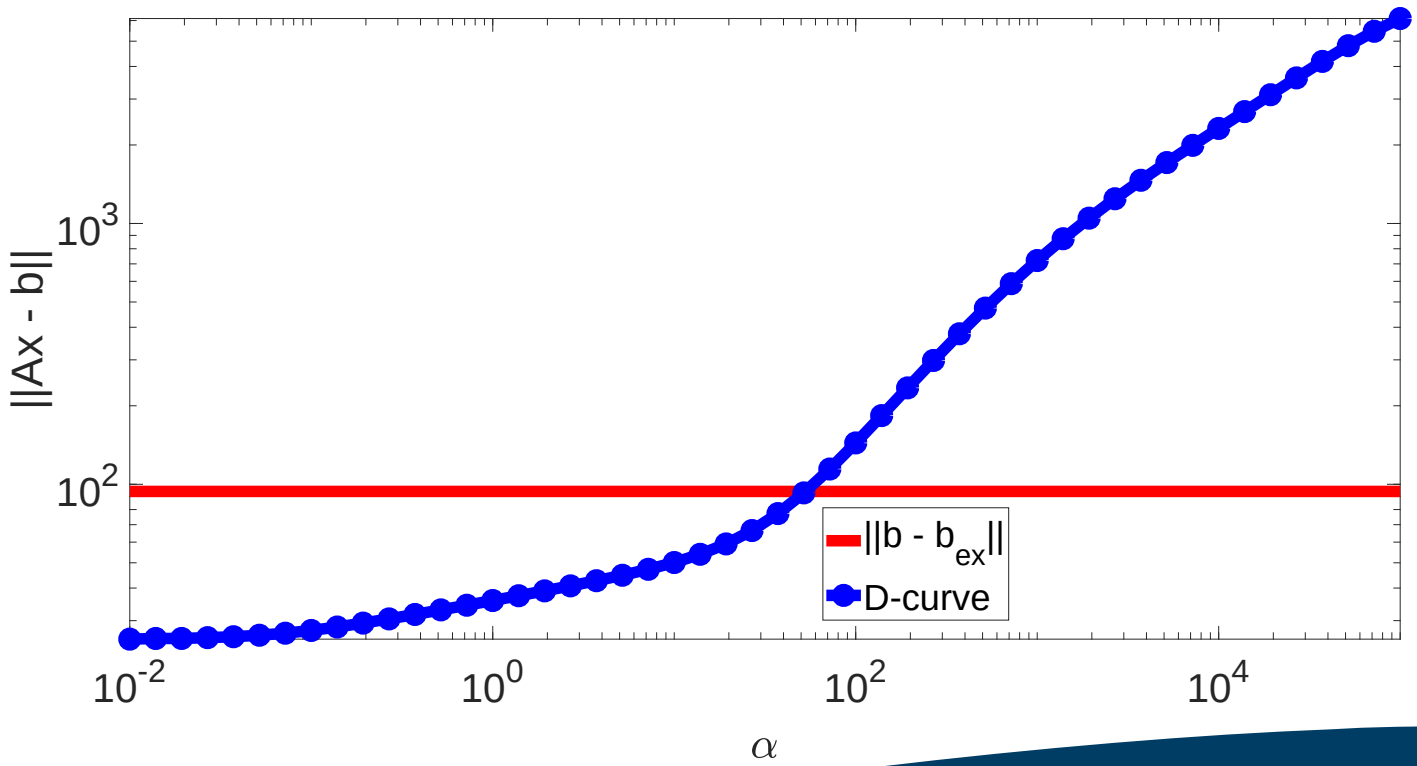
Reconstruction

Tikhonov regularization & Discrepancy principle

Compute $x_\alpha = \underset{x \in \mathbb{R}^n}{\arg\min} \, ||Ax - b||^2 + \alpha ||x||^2$ for many values of α.

How do we choose the parameter $\alpha \in \mathbb{R}^+$?

Discrepancy principle

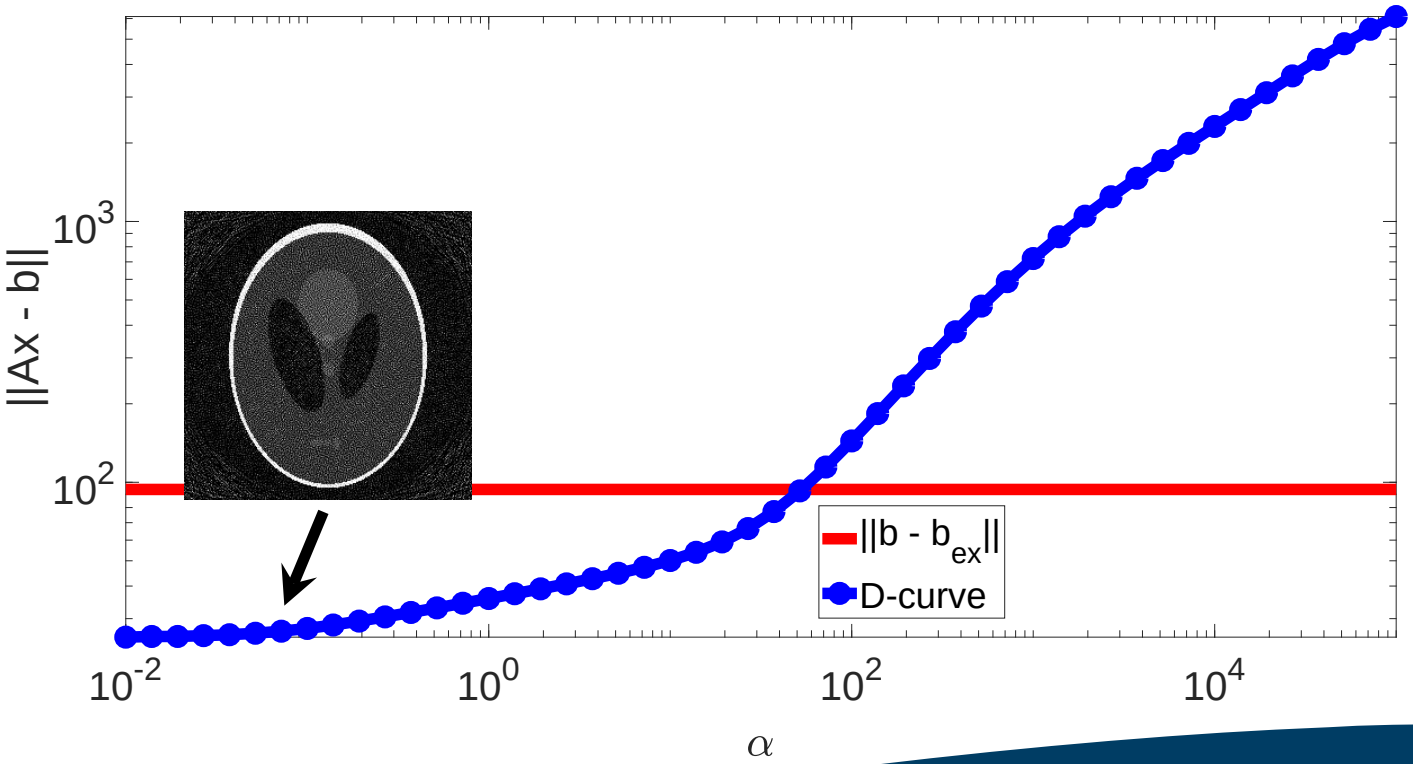


Tikhonov regularization & Discrepancy principle

Compute $x_\alpha = \arg\min_{x \in \mathbb{R}^n} ||Ax - b||^2 + \alpha||x||^2$ for many values of α.

How do we choose the parameter $\alpha \in \mathbb{R}^+$?

Discrepancy principle

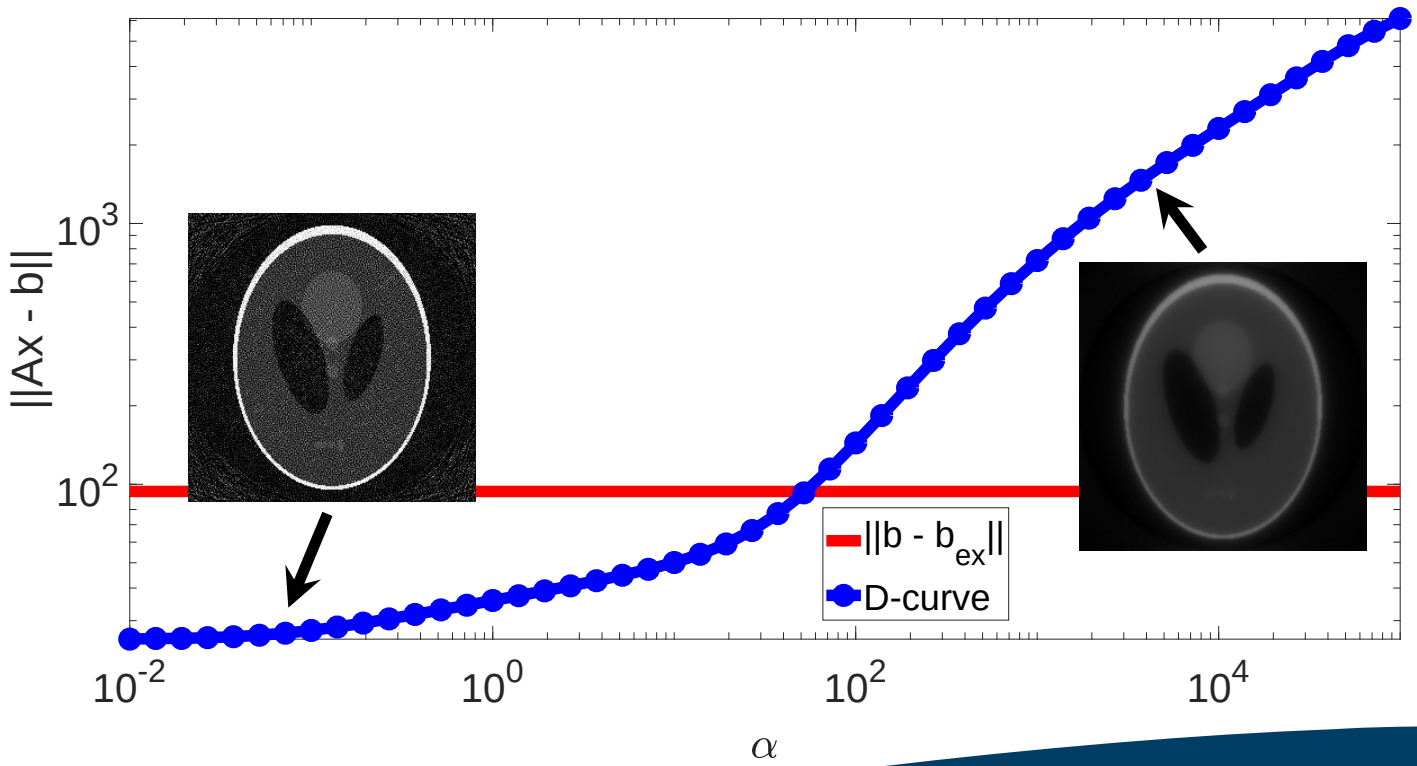


Tikhonov regularization & Discrepancy principle

Compute $x_\alpha = \arg\min_{x \in \mathbb{R}^n} ||Ax - b||^2 + \alpha||x||^2$ for many values of α.

How do we choose the parameter $\alpha \in \mathbb{R}^+$?

Discrepancy principle

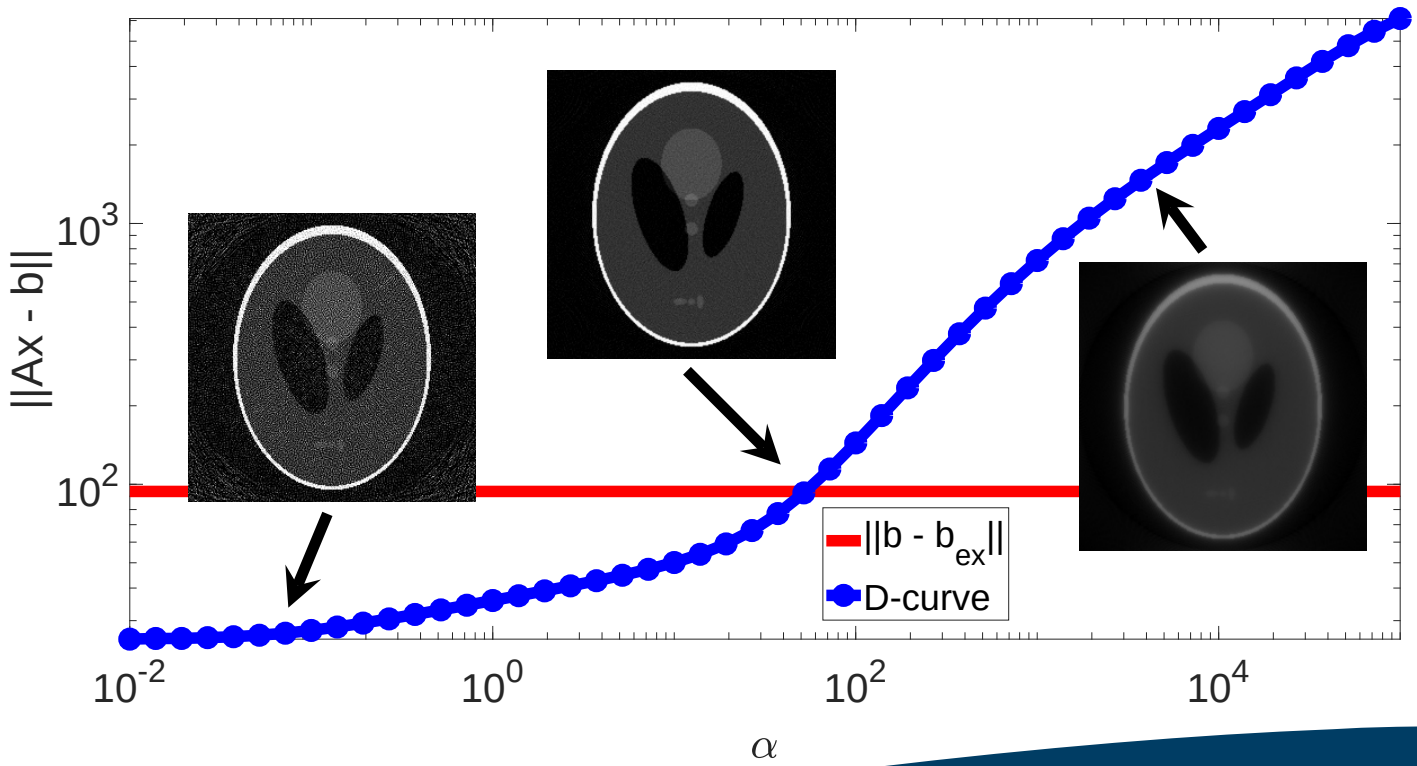


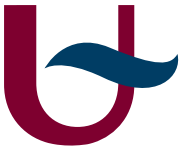
Tikhonov regularization & Discrepancy principle

Compute $x_\alpha = \arg\min_{x \in \mathbb{R}^n} ||Ax - b||^2 + \alpha||x||^2$ for many values of α.

How do we choose the parameter $\alpha \in \mathbb{R}^+$? $\rightarrow ||Ax_\alpha - b|| = ||b - b_{ex}||$.

Discrepancy principle





Tikhonov regularization & Discrepancy principle

The constrained optimization problem

$$\min_{x \in \mathbb{R}^n} \quad ||x||^2 \quad \text{subject to} \quad ||Ax - b||^2 = ||e||^2$$

has a unique solution x^*.

$\exists!$ Lagrange multiplier $\lambda^* > 0 \in \mathbb{R}$, such that $F(x^*, \lambda^*) = 0$ with

$$F(x, \lambda) = \begin{pmatrix} \lambda A^T (Ax - b) + x \\ \frac{1}{2}||Ax - b||^2 - \frac{||e||^2}{2} \end{pmatrix}.$$

Any point (x, λ) with $\lambda > 0$ that is a root of the first component of F is a Tikhonov solution with $\alpha = 1/\lambda$:

$$\lambda A^T (Ax - b) + x = 0 \Leftrightarrow (A^T A + \alpha I)x = A^T b.$$

Tikhonov regularization & Discrepancy principle

We can use **Newton's method** solve $F(x, \lambda) = 0$.

Let (x_0, λ_0) be an initial point. We have a sequence of iterates

$$(x_k, \lambda_k) = (x_{k-1}, \lambda_{k-1}) + \gamma_k(\Delta x_k, \Delta \lambda_k), \qquad k = 1, 2, 3, \ldots$$

with step-length $\gamma_k \in]0, 1]$ and Newton step $(\Delta x_k, \Delta \lambda_k)$ defined by

$$J(x_{k-1}, \lambda_{k-1}) \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = -F(x_{k-1}, \lambda_{k-1})$$

where $J(x, \lambda)$ is the Jacobian matrix of $F(x, \lambda)$.

We can solve this linear system with a **Krylov subspace method**, e.g. MINRES.

Projected Newton method

The bidiagonalization algorithm (Bidiag1) generates matrices $V_k = [v_0, \ldots, v_{k-1}]$ and $U_{k+1} = [u_0, \ldots, u_k]$ with orthonormal columns that satisfy

$$AV_k = U_{k+1}B_{k+1}$$

with a lower bidiagonal matrix $B_{k+1} \in \mathbb{R}^{(k+1)\times k}$ and such that

$$\mathcal{R}(V_k) = \mathcal{K}_k(A^TA, A^Tb) = \text{span}\left\{A^Tb, (A^TA)A^Tb, \ldots, (A^TA)^{k-1}A^Tb\right\}.$$

By writing $x_k = V_k y_k$ for $y_k \in \mathbb{R}^k$ we have $||x_k|| = ||V_k y_k|| = ||y_k||$ and

$$\begin{aligned} ||Ax_k - b|| &= ||AV_k y_k - U_{k+1}c_k|| \\ &= ||U_{k+1}(B_{k+1}y_k - c_k)|| \\ &= ||B_{k+1}y_k - c_k||. \end{aligned}$$

with $c_k = ||b||e_1 \in \mathbb{R}^{k+1}$.

Projected Newton method

The bidiagonalization algorithm (Bidiag1) generates matrices $V_k = [v_0, \ldots, v_{k-1}]$ and $U_{k+1} = [u_0, \ldots, u_k]$ with orthonormal columns that satisfy

$$AV_k = U_{k+1}B_{k+1}$$

with a lower bidiagonal matrix $B_{k+1} \in \mathbb{R}^{(k+1)\times k}$ and such that

$$\mathcal{R}(V_k) = \mathcal{K}_k(A^TA, A^Tb) = \text{span}\left\{A^Tb, (A^TA)A^Tb, \ldots, (A^TA)^{k-1}A^Tb\right\}.$$

By writing $x_k = V_k y_k$ for $y_k \in \mathbb{R}^k$ we have $||x_k|| = ||V_k y_k|| = ||y_k||$ and

$$\begin{aligned} ||Ax_k - b|| &= ||AV_k y_k - U_{k+1}c_k|| \\ &= ||U_{k+1}(B_{k+1}y_k - c_k)|| \\ &= ||B_{k+1}y_k - c_k||. \end{aligned}$$

with $c_k = ||b||e_1 \in \mathbb{R}^{k+1}$.

Projected Newton method

This shows that the **projected minimization problem**

$$\min_{x_k \in \mathcal{R}(V_k)} \quad ||x_k||^2 \quad \text{s.t.} \quad ||Ax_k - b||^2 = ||e||^2$$

is equivalent to

$$\min_{y_k \in \mathbb{R}^k} \quad ||y_k||^2 \quad \text{s.t.} \quad ||B_{k+1}y_k - c_k||^2 = ||e||^2.$$

The solution y_k^* with Lagrange multiplier λ_k^* satisfies $F^{(k)}(y_k^*, \lambda_k^*) = 0$ with

$$F^{(k)}(y, \lambda) = \begin{pmatrix} \lambda B_{k+1}^T (B_{k+1}y - c_k) + y \\ \frac{1}{2}||B_{k+1}y - c_k||^2 - \frac{||e||^2}{2} \end{pmatrix}.$$

A solution of $F^{(k)}(y, \lambda) = 0$ gives an approximation for the solution of $F(x, \lambda) = 0$ and is much cheaper to compute if $k \ll n$.

Projected Newton method

Suppose we have $x_{k-1} = V_{k-1} y_{k-1}$ and let us denote $\bar{y}_{k-1} = (y_{k-1}^T, 0) \in \mathbb{R}^k$. We can calculate **a single Newton step** for the projected function $F^{(k)}(y, \lambda)$:

$$J^{(k)}(\bar{y}_{k-1}, \lambda_{k-1}) \begin{pmatrix} \Delta y_k \\ \Delta \lambda_k \end{pmatrix} = -F^{(k)}(\bar{y}_{k-1}, \lambda_{k-1})$$

with $J^{(k)}(y, \lambda)$ the Jacobian matrix of $F^k(y, \lambda)$ and we update

$$(y_k, \lambda_k) = (\bar{y}_{k-1}, \lambda_{k-1}) + \gamma_k (\Delta y_k, \Delta \lambda_k).$$

This gives us a corresponding update for x_{k-1}:

$$x_k = V_k y_k = V_k \bar{y}_{k-1} + \gamma_k \underbrace{V_k \Delta y_k}_{:=\Delta x_k} = x_{k-1} + \gamma_k \Delta x_k.$$

Projected Newton method

Suppose we have $x_{k-1} = V_{k-1} y_{k-1}$ and let us denote $\bar{y}_{k-1} = (y_{k-1}^T, 0) \in \mathbb{R}^k$. We can calculate **a single Newton step** for the projected function $F^{(k)}(y, \lambda)$:

$$J^{(k)}(\bar{y}_{k-1}, \lambda_{k-1}) \begin{pmatrix} \Delta y_k \\ \Delta \lambda_k \end{pmatrix} = -F^{(k)}(\bar{y}_{k-1}, \lambda_{k-1})$$

with $J^{(k)}(y, \lambda)$ the Jacobian matrix of $F^k(y, \lambda)$ and we update

$$(y_k, \lambda_k) = (\bar{y}_{k-1}, \lambda_{k-1}) + \gamma_k (\Delta y_k, \Delta \lambda_k).$$

This gives us a corresponding update for x_{k-1}:

$$x_k = V_k y_k = V_k \bar{y}_{k-1} + \gamma_k \underbrace{V_k \Delta y_k}_{:= \Delta x_k} = x_{k-1} + \gamma_k \Delta x_k.$$

Projected Newton method

Theorem

Let $(\Delta y_k, \Delta \lambda_k)$ *be the Newton step for* $F^{(k)}(\bar{y}_{k-1}, \lambda_{k-1})$ *and* $\Delta x_k = V_k \Delta y_k$ *then*

$$\nabla f(x_{k-1}, \lambda_{k-1})^T \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = -||F(x_{k-1}, \lambda_{k-1})||^2 \leq 0$$

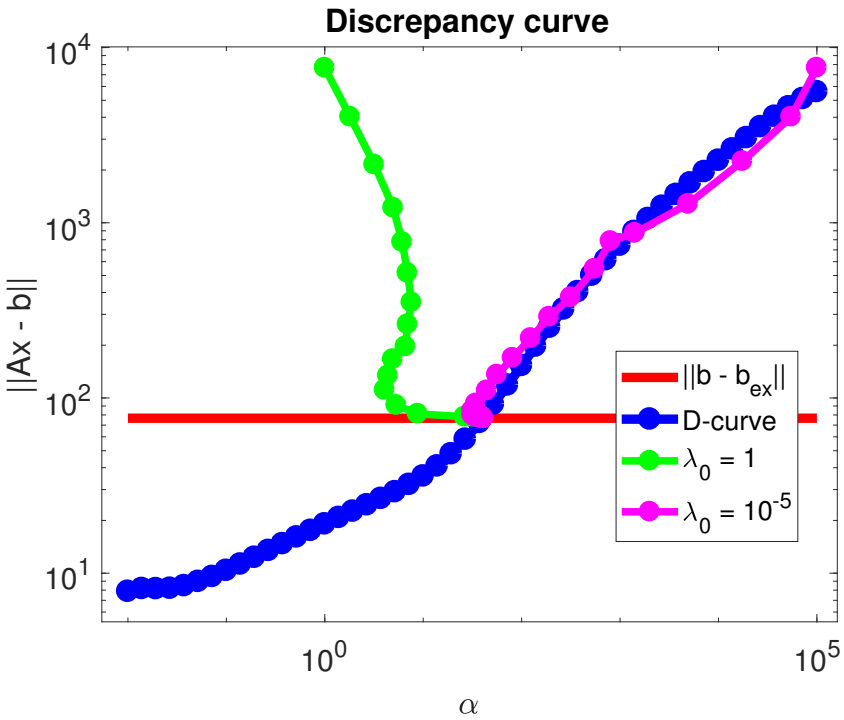
with $f(x, \lambda) = \frac{1}{2}||F(x, \lambda)||^2$.

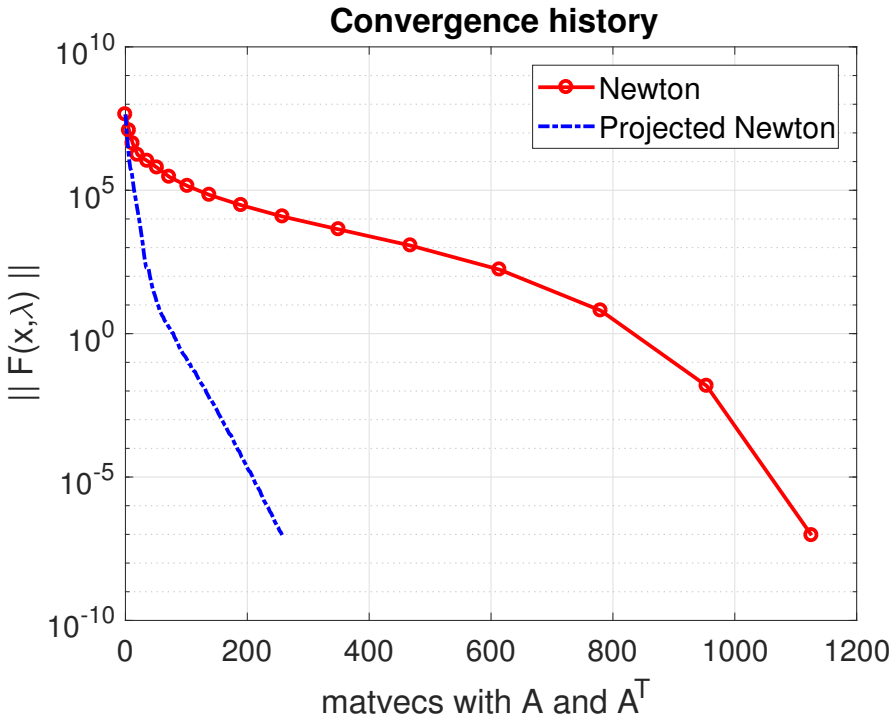
Use this descent direction in a backtracking line-search to obtain step-length γ_k such that the **sufficient decrease condition** is satisfied

$$\frac{1}{2}||F(x_k, \lambda_k)||^2 \leq \left(\frac{1}{2} - \rho\gamma_k\right) ||F(x_{k-1}, \lambda_{k-1})||^2$$

with some constant $\rho \in (0, 1)$.

Projected Newton method





Extension of the Projected Newton method

The Projected Newton method can be extended to solve

$$\min_{x \in \mathbb{R}^n} \quad \Psi(x) \quad \text{s.t.} \quad ||Ax - b||^2 = ||e||^2$$

with convex twice differentiable regularization function $\Psi(x) : \mathbb{R}^n \to \mathbb{R}$.

First order optimality conditions are given by $F(x, \lambda) = 0$ with

$$F(x, \lambda) = \begin{pmatrix} \lambda A^T (Ax - b) + \nabla \Psi(x) \\ \frac{1}{2}||Ax - b||^2 - \frac{||e||^2}{2} \end{pmatrix}.$$

We now project the optimization problem on a **Generalized Krylov subspace**

$$\mathcal{R}(V_k) = \text{span}\{r_0, r_1, \ldots, r_{k-1}\}$$

with orthonormal columns computed using Gram–Schmidt and

$$r_{k-1} = \lambda_{k-1} A^T (Ax_{k-1} - b) + \nabla \Psi(x_{k-1}).$$

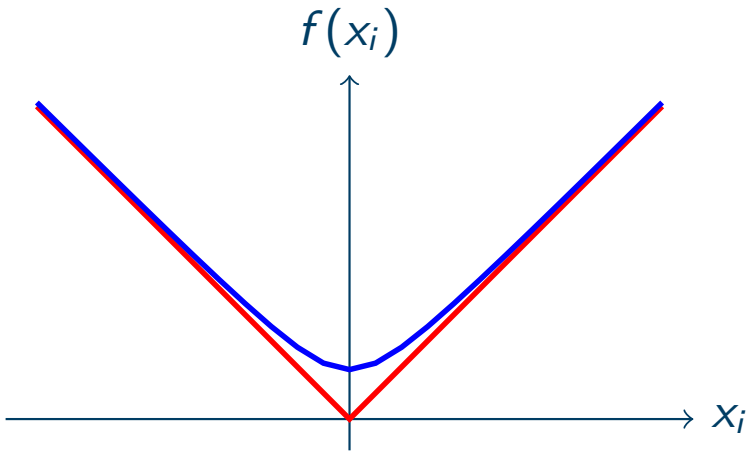
Extension of the Projected Newton method

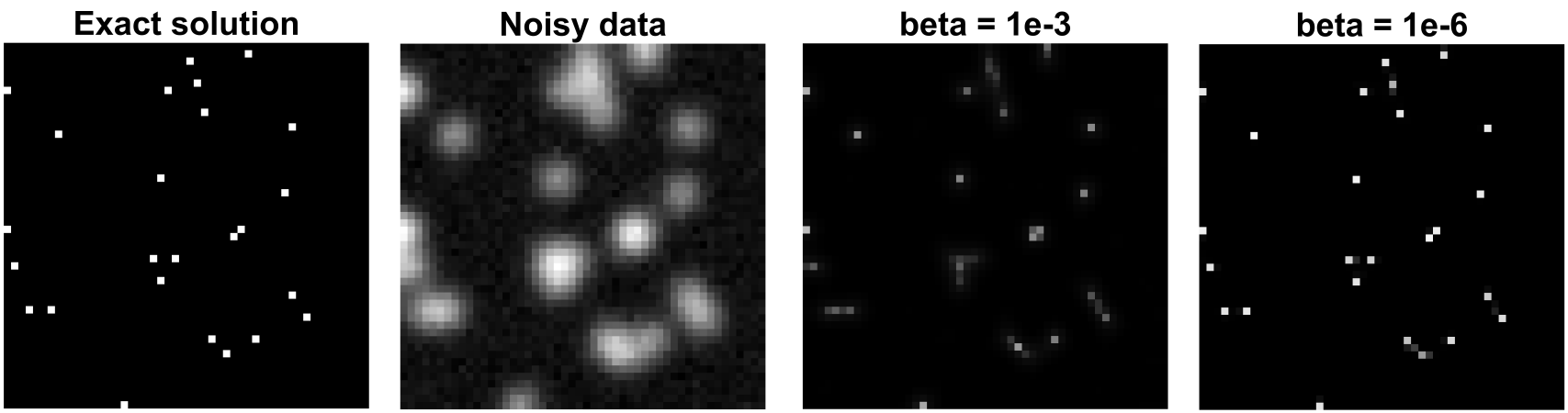
A popular regularization technique is given by

$$\min_{x \in \mathbb{R}^n} \quad ||x||_1 \quad \text{s.t.} \quad ||Ax - b||^2 = ||e||^2$$

with $||x||_1 = \sum_{i=1}^{n} |x_i|$, since it induces sparsity in the solution.

Replace non-differentiable $|x_i|$ with $\sqrt{x_i^2 + \beta}$.





Conclusions & ongoing research

Conclusions

- Ill-conditioned linear systems need regularization.
- Tihkonov regularization is a good basic regularization technique.
- Discrepancy principle is an easy way to choose regularization parameter.
- We can develop an efficient algorithm by projection on a Krylov subspace.
- Extend Projected Newton method with Generalized Krylov subspace.

Ongoing research

- Nonlinear inverse problems

$$\min \Psi(x) \quad \text{s.t.} \quad ||s(x) - b||^2 = \epsilon^2.$$

References

Some interesting references

- Per Christian Hansen. *Discrete inverse problems: insight and algorithms.* Vol. 7. Siam, 2010.
- Per Christian Hansen. *Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems.* Numerical algorithms 6.1, 1994.

The most interesting references*

- Jeffrey Cornelis, Nick Schenkels and Wim Vanroose. *Projected Newton Method for noise constrained Tikhonov regularization.* IOP Inverse Problems 36.5 (2020): 055002.
- Jeffrey Cornelis and Wim Vanroose. *Projected Newton Method for noise constrained ℓ_p regularization.* IOP Inverse Problems in press, `https://doi.org/10.1088/1361-6420/abb2fc`, 2020.

∗ Disclaimer: I might be a bit biased.