Numerical Solution of Large-Scale Inverse Problems

Silvia Gazzola

Department of Mathematical Sciences

UNIVERSITY OF BATH

SAMBa SLS
December 7, 2016

Introducing Inverse Problems

Introducing Inverse Problems

When solving a

DIRECT PROBLEM

we are interested in a system's behavior, given the internal structure of the system.

Introducing Inverse Problems

When solving a

DIRECT PROBLEM

we are interested in a system's behavior, given the internal structure of the system.

When solving an

INVERSE PROBLEM

we are interested in determining the internal structure of a system, given the system's observed behavior.

Introducing Inverse Problems

When solving a

DIRECT PROBLEM

we are interested in a system's behavior, given the internal structure of the system.

When solving an

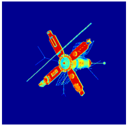
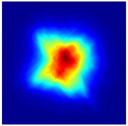
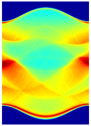
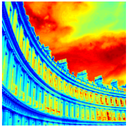
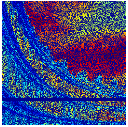
INVERSE PROBLEM

we are interested in determining the internal structure of a system, given the system's observed behavior.

- **Direct problems**: from the cause of a observed phenomenon, to its effect.
- **Inverse problems**: from the effect of an observed phenomenon, to its cause.

Some examples

Some examples

model	input	output
image deblurring (astronomical imaging)		
PSF, convolution		
computerised tomography (industrial, medical)		
X-rays, line integrals		
matrix completion		
		

Inverse problems are ill-posed

According to Hadamard (1923), a problem is **ill-posed** if

- (the) solution is not unique

or

- (the) solution does not depend continuously on the given data

Inverse problems are ill-posed

According to Hadamard (1923), a problem is **ill-posed** if

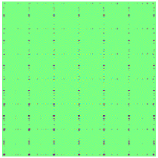
- (the) solution is not unique

or

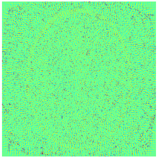
- (the) solution does not depend continuously on the given data

Indeed, this is what happens:

deblurring



CT



Understanding what goes wrong

Understanding what goes wrong

Consider the SVD of $A \in \mathbb{R}^{N \times N}$, for these examples $N = 65536$:

$$A = U \Sigma V^T$$

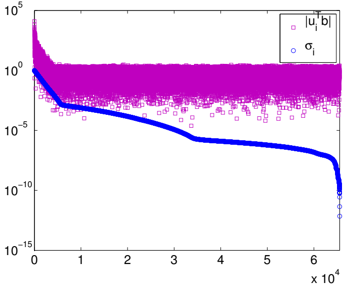
When there is noise, $b = b^{ex} + e$ (for this example $\|e\|_2 / \|b\|_2 = 10^{-2}$):

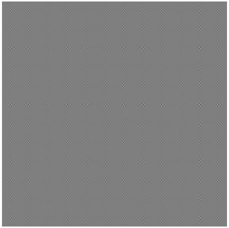
Understanding what goes wrong

Consider the SVD of $A \in \mathbb{R}^{N\times N}$, for these examples $N = 65536$:

$$A = U\Sigma V^T \quad \Longrightarrow \quad x = \sum_{i=1}^{N} \frac{u_i^T b}{\sigma_i} v_i$$

When there is noise, $b = b^{ex} + e$ (for this example $\|e\|_2/\|b\|_2 = 10^{-2}$):



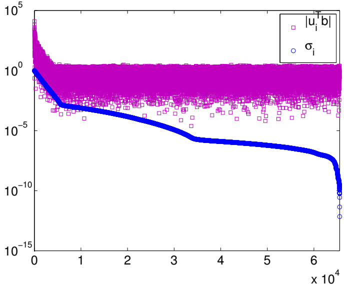


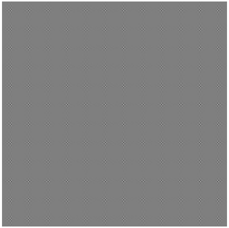
Understanding what goes wrong

Consider the SVD of $A \in \mathbb{R}^{N\times N}$, for these examples $N = 65536$:

$$A = U\Sigma V^T \quad \Longrightarrow \quad x = \sum_{i=1}^{N} \frac{u_i^T b}{\sigma_i} v_i = \sum_{i=1}^{N} \frac{u_i^T b^{ex}}{\sigma_i} v_i + \sum_{i=1}^{N} \frac{u_i^T e}{\sigma_i} v_i\,.$$

When there is noise, $b = b^{ex} + e$ (for this example $\|e\|_2/\|b\|_2 = 10^{-2}$):





Applying direct regularisation

Applying direct regularisation

Truncated SVD (TSVD):

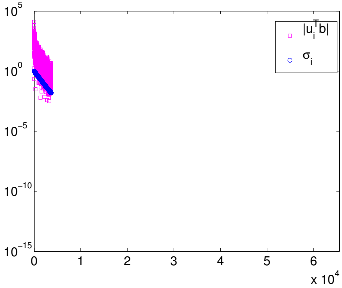
$$x_k = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i\,, \quad k \ll N\,.$$

Applying direct regularisation

Truncated SVD (TSVD):

$$x_k = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i \,, \quad k \ll N \,.$$

For this example: $k = 3653$.



Applying direct regularisation

Tikhonov regularization:

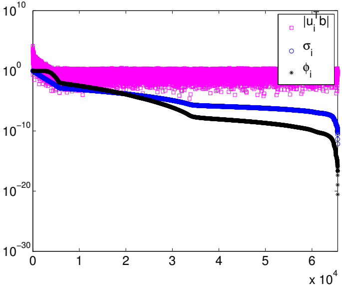
$$x_\lambda = \arg\min_{x\in\mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|Lx\|_2^2 , \quad \lambda > 0$$

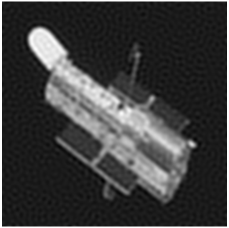
Applying direct regularisation

Tikhonov regularization:

$$x_\lambda = \arg\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|Lx\|_2^2 \,, \quad \lambda > 0$$

For this example: $\lambda = 1.76 \cdot 10^{-4}$, $L = I$.





Iterative Regularization Methods (no SVD!)

Iterative Regularization Methods (no SVD!)

- Gradient Descent Methods

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

Iterative Regularization Methods (no SVD!)

- Gradient Descent Methods

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

- Krylov methods: CGLS (LSQR), GMRES, ...

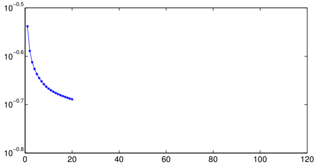
Iterative Regularization Methods (no SVD!)

- Gradient Descent Methods

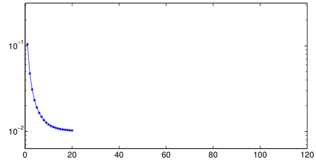
$$x_{m+1} = x_m + \alpha_m A^T(b - Ax_m)$$

- Krylov methods: CGLS (LSQR), GMRES, ...

relative errors



relative residuals



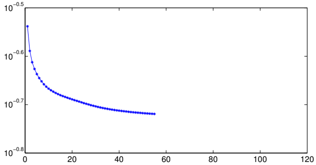
Iterative Regularization Methods (no SVD!)

- Gradient Descent Methods

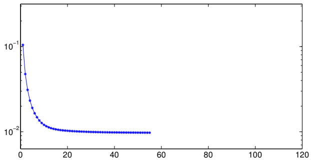
$$x_{m+1} = x_m + \alpha_m A^T (b - Ax_m)$$

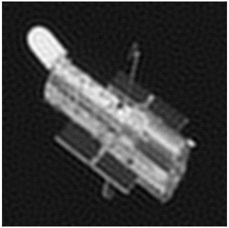
- Krylov methods: CGLS (LSQR), GMRES, ...

relative errors



relative residuals



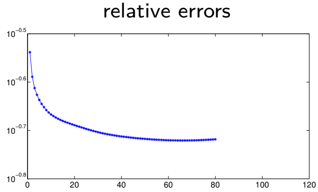


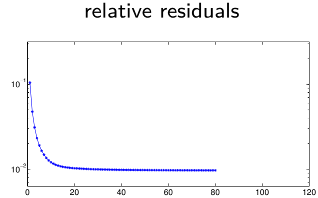
Iterative Regularization Methods (no SVD!)

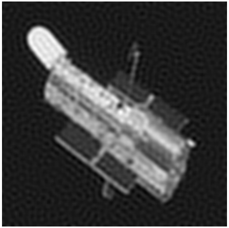
- Gradient Descent Methods

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

- Krylov methods: CGLS (LSQR), GMRES, ...







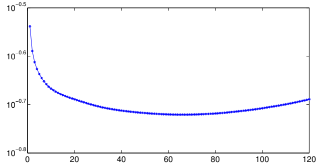
Iterative Regularization Methods (no SVD!)

- Gradient Descent Methods

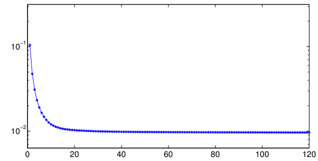
$$x_{m+1} = x_m + \alpha_m A^T (b - A x_m)$$

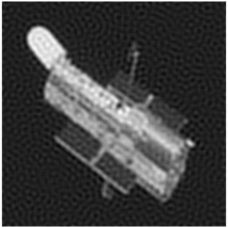
- Krylov methods: CGLS (LSQR), GMRES, ...

relative errors

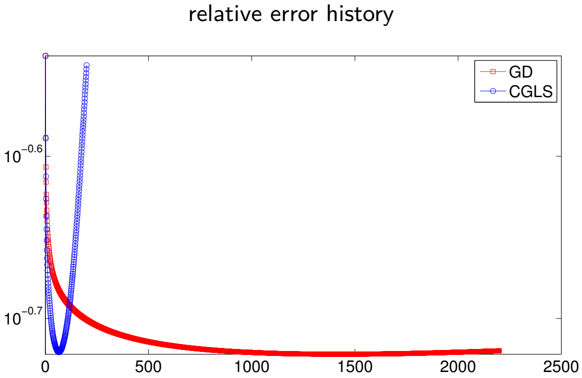


relative residuals





Gradient Descent approach VS. Krylov Subspaces approach



Enforcing sparsity

- By penalisation:

$$x_\lambda = \arg\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|x\|_1$$

$$x_\lambda = \arg\min_{x \in \mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|\Psi x\|_1$$

Enforcing sparsity

- By penalisation:

$$
\begin{aligned}
x_\lambda &= \arg\min_{x\in\mathbb{R}^N} \|Ax-b\|_2^2 + \lambda\,\|x\|_1 \\
x_\lambda &= \arg\min_{x\in\mathbb{R}^N} \|Ax-b\|_2^2 + \lambda\,\|\Psi x\|_1
\end{aligned}
$$

- By imposing constraints:

$$
\begin{aligned}
x_B &= \arg\min_{x\in B} \|Ax-b\|_2\,, \quad \text{e.g., } B=\{x \text{ s.t. } \|\Psi x\|_1 \leq \zeta\} \\
x_C &= \arg\min_{x\in C} \|\Psi x\|_1\,, \quad \text{e.g., } C=\{x \text{ s.t. } \|Ax-b\|_2 \leq \varepsilon\}
\end{aligned}
$$

Enforcing sparsity

- By penalisation:

$$
\begin{aligned}
x_\lambda &= \arg\min_{x\in\mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|x\|_1 \\
x_\lambda &= \arg\min_{x\in\mathbb{R}^N} \|Ax - b\|_2^2 + \lambda \|\Psi x\|_1
\end{aligned}
$$

- By imposing constraints:

$$
\begin{aligned}
x_B &= \arg\min_{x\in B} \|Ax - b\|_2 \,, \quad \text{e.g., } B = \{x \text{ s.t. } \|\Psi x\|_1 \leq \zeta\} \\
x_C &= \arg\min_{x\in C} \|\Psi x\|_1 \,, \quad \text{e.g., } C = \{x \text{ s.t. } \|Ax - b\|_2 \leq \varepsilon\} \\
x_+ &= \arg\min_{x\geq 0} \|Ax - b\|_2
\end{aligned}
$$

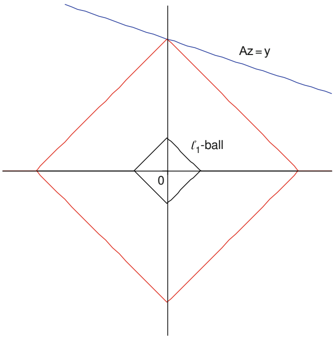
Why $\|\cdot\|_1$ is successful in enforcing sparsity?

Why $\|\cdot\|_1$ is successful in enforcing sparsity?

$$\min \|z\|_1 \quad \text{s.t.} \quad Az = y\,, \quad \text{with } A \in \mathbb{R}^{1\times 2},\ \|z\|_1 = |z_1| + |z_2|$$

Why $\|\cdot\|_1$ is successful in enforcing sparsity?

$$\min \|z\|_1 \quad \text{s.t.} \quad Az = y\,, \quad \text{with } A \in \mathbb{R}^{1\times 2},\ \|z\|_1 = |z_1| + |z_2|$$

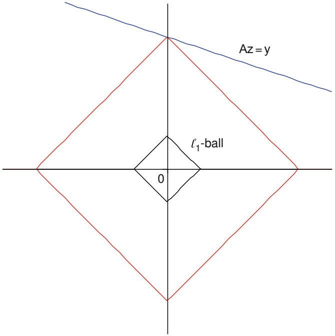


!

[Image courtesy: Fornasier and Rauhut. *Compressive Sensing*, 2011]

Why $\|\cdot\|_1$ is successful in enforcing sparsity?

$$\min \|z\|_1 \quad \text{s.t.} \quad Az = y\,, \quad \text{with } A \in \mathbb{R}^{1\times 2},\ \|z\|_1 = |z_1| + |z_2|$$

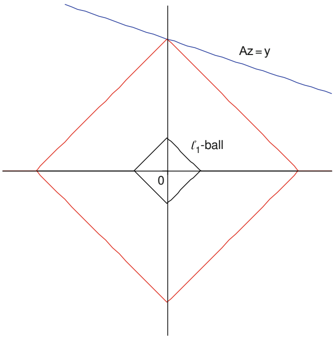


!

[Image courtesy: Fornasier and Rauhut. *Compressive Sensing*, 2011]

Why $\|\cdot\|_1$ is successful in enforcing sparsity?

$$\min \|z\|_1 \quad \text{s.t.} \quad Az = y\,, \quad \text{with } A \in \mathbb{R}^{1\times 2},\ \|z\|_1 = |z_1| + |z_2|$$

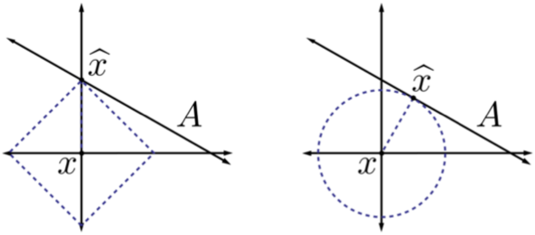


!

[Image courtesy: Fornasier and Rauhut. *Compressive Sensing*, 2011]

the sparsest solution is recovered!

$\|\cdot\|_1$ vs. $\|\cdot\|_2$ minimisation



[Image courtesy: Baraniuk et al. *An Introduction to Compressive Sensing*, 2013]

Examples of sparsity transforms Ψ

Examples of sparsity transforms Ψ

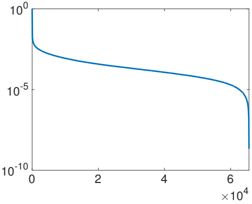
original

Examples of sparsity transforms Ψ

original

2D dct

dct coefficients

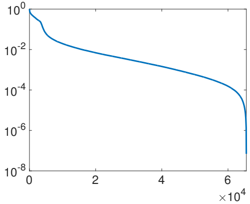


Examples of sparsity transforms Ψ

original

2D wavelets

wavelet coefficients

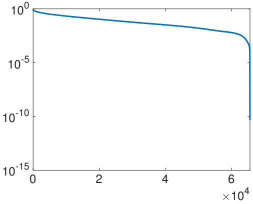


Examples of sparsity transforms Ψ

original

gradient

gradient coefficients



Incomplete information & compressive sensing theory

Assume we wish recover $x \in \mathbb{R}^N$ from

$$Ax + n = b, \quad A \in \mathbb{R}^{M \times N}, \quad M \ll N.$$

Incomplete information & compressive sensing theory

Assume we wish recover $x \in \mathbb{R}^N$ from

$$Ax + n = b, \quad A \in \mathbb{R}^{M \times N}, \quad M \ll N.$$

Then, provided that we have

- **sparsity** (Ψx is k-sparse)
- **randomness** (the rows of A are chosen uniformly at random)
- "**incoherence**" ((A, Ψ) with "coherence" $\mu \ll \sqrt{N}$)

and

$$M \geq C \cdot \mu^2 \cdot (k \cdot \log(N)),$$

Incomplete information & compressive sensing theory

Assume we wish recover $x \in \mathbb{R}^N$ from

$$Ax + n = b, \quad A \in \mathbb{R}^{M \times N}, \quad M \ll N.$$

Then, provided that we have

- **sparsity** (Ψx is k-sparse)
- **randomness** (the rows of A are chosen uniformly at random)
- "**incoherence**" ((A, Ψ) with "coherence" $\mu \ll \sqrt{N}$)

and

$$M \geq C \cdot \mu^2 \cdot (k \cdot \log(N)),$$

then the **compressive sensing** theory guarantees that we can recover $x \in \mathbb{R}^N$ with overwhelming probability by solving

$$\min_{x \in \mathbb{R}^N} \|\Psi x\|_1 \quad \text{s.t.} \quad \|Ax - b\|_2 \leq \varepsilon.$$

Making sense of "(in)coherence"

Let a^i, $i = 1, \ldots, N$, ψ^j, $j = 1, \ldots, N$ be two basis of $\mathbb{R}^N$. Then, the **coherence** μ between a^i and ψ^j is defined as

$$\mu = \sqrt{N} \max_{1 \leq i, j \leq N} \frac{|\langle a^i, \psi^j \rangle|}{\|a^i\|_2 \|\psi^j\|_2} .$$

Making sense of "(in)coherence"

Let a^i, $i = 1, \ldots, N$, ψ^j, $j = 1, \ldots, N$ be two basis of $\mathbb{R}^N$. Then, the **coherence** μ between a^i and ψ^j is defined as

$$\mu = \sqrt{N} \max_{1 \leq i,j \leq N} \frac{|\langle a^i, \psi^j \rangle|}{\|a^i\|_2 \|\psi^j\|_2} .$$

Note that

$$1 \leq \mu \leq \sqrt{N} .$$

Making sense of "(in)coherence"

Let a^i, $i = 1, \ldots, N$, ψ^j, $j = 1, \ldots, N$ be two basis of $\mathbb{R}^N$. Then, the **coherence** μ between a^i and ψ^j is defined as

$$\mu = \sqrt{N} \max_{1 \leq i,j \leq N} \frac{|\langle a^i, \psi^j \rangle|}{\|a^i\|_2 \|\psi^j\|_2} .$$

Note that

$$1 \leq \mu \leq \sqrt{N} .$$

[time for a movie...]

On the whiteboard...

J. F. Cai, E. J. Candes, and Z. Shen

A SINGULAR VALUE THRESHOLDING ALGORITHM FOR MATRIX COMPLETION

SIAM J. Optim., Vol 20, No 4, pp. 1956–1982

http://epubs.siam.org/doi/ref/10.1137/080738970