Conjugate Gradient algorithm

- **Need:** A symmetric positive definite;
- **Cost:** 1 matrix-vector product per step;
- **Storage:** fixed, independent of number of steps.

The CG method minimizes the A norm of the error,

$$x_k = \arg \min_{x \in \mathcal{K}_k(A,b)} \|x - x_*\|_A^2.$$

$$x_* = \text{true solution}, \quad \|z\|_A^2 = z^{\mathrm{T}} A z.$$

KRYLOV SUBSPACES

The kth Krylov subspace associated with the matrix A and the vector b is

$$\mathcal{K}_k(A, b) = \text{span}\{b, Ab, \ldots, A^{k-1}b\}.$$

Iterative methods which seek the solution in a Krylov subspace are called Krylov subspace iterative methods.

At each step, minimize

$$\alpha \mapsto \|x_{k-1} + \alpha p_{k-1} - x_*\|_A^2$$

Solution:

$$\alpha_k = \frac{\|r_{k-1}\|^2}{p_{k-1}^{\mathrm{T}} A p_{k-1}}$$

New update

$$x_k = x_{k-1} + \alpha_{k-1} p_{k-1}.$$

Search directions:

$$p_0 = r_0 = b - Ax_0,$$

Iteratively, *A-conjugate* to the previous ones:

$$p_k^{\mathrm{T}} A p_j = 0, \quad 0 \le j \le k-1.$$

Found by writing

$$p_k = r_k + \beta_k p_{k-1}, \quad r_k = b - A x_k,$$

$$\beta_k = \frac{\|r_k\|^2}{\|r_{k-1}\|^2},$$

ALGORITHM (CG)

Initialize: $x_0 = 0$; $r_0 = b - Ax_0$; $p_0 = r_0$;

for $k = 1, 2, \ldots$ until stopping criterion is satisfied

$$\alpha_k = \frac{\|r_{k-1}\|^2}{p_{k-1}^T A p_{k-1}};$$

$$x_k = x_{k-1} + \alpha_k p_{k-1};$$

$$r_k = r_{k-1} - \alpha_k A p_{k-1};$$

$$\beta_k = \frac{\|r_k\|^2}{\|r_{k-1}\|^2};$$

$$p_k = r_k + \beta_k p_{k-1};$$

end

CGLS Method

Conjugate Gradient method for Least Squares (CGLS)

- **Need:** A can be rectangular (non-square);
- **Cost:** 2 matrix-vector products (one with A, one with A^{T}) per step;
- **Storage:** fixed, independent of number of steps.

Mathematically equivalent to applying CG to normal equations

$$A^{\mathrm{T}}Ax = A^{\mathrm{T}}b$$

without actually forming them.

CGLS minimization problem

The kth iterate solves the minimization problem

$$x_k = \arg \min_{x \in \mathcal{K}_k(A^{\mathrm{T}}A, A^{\mathrm{T}}b)} \|b - Ax\|.$$

The kth iterate x_k of CGLS method ($x_0 = 0$) is characterized by

$$\Phi(x_k) = \min_{x \in \mathcal{K}_k(A^{\mathrm{T}}A, A^{\mathrm{T}}b)} \Phi(x)$$

where

$$\Phi(x) := \frac{1}{2} x^{\mathrm{T}} A^{\mathrm{T}} A x - x^{\mathrm{T}} A^{\mathrm{T}} b.$$

DETERMINATION OF THE MINIMIZER

Perform sequential linear searches along $A^{\mathrm{T}}A$-conjugate directions

$$p_0, p_1, \ldots, p_{k-1}$$

that span $\mathcal{K}_k(A^{\mathrm{T}}A, A^{\mathrm{T}}b)$.

Determine x_k from x_{k-1} and p_{k-1} according to

$$x_k := x_{k-1} + \alpha_{k-1} p_{k-1}$$

where $\alpha_{k-1} \in \mathbb{R}$ solves

$$\min_{\alpha \in \mathbb{R}} \Phi(x_{k-1} + \alpha p_{k-1}).$$

RESIDUAL ERROR

Introduce the residual error associated with x_k:

$$r_k := A^{\mathrm{T}}b - A^{\mathrm{T}}Ax_k.$$

Then

$$p_k := r_k + \beta_{k-1}p_{k-1}$$

Choose β_{k-1} so that p_k is $A^{\mathrm{T}}A$-conjugate to the previous search directions:

$$p_k^{\mathrm{T}}A^{\mathrm{T}}Ap_j = 0, \quad 1 \leq j \leq k-1.$$

Discrepancy

The discrepancy associated with x is

$$d_k = b - Ax_k.$$

Then

$$r_k = A^{\mathrm{T}} d_k$$

It was shown by Hestenes and Stiefel that

$$\|d_{k+1}\| \leq \|d_k\|; \qquad \|x_{k+1}\| \geq \|x_k\|.$$

Algorithm (CGLS)

$x_0 := 0; \quad d_0 = b; \quad r_0 = A^{\mathrm{T}} b;$
$p_0 = r_0; \quad t_0 = A p_0;$

for $k = 1, 2, \ldots$ until stopping criterion is satisfied

$\alpha_k = \|r_{k-1}\|^2 / \|t_{k-1}\|^2$

$x_k = x_{k-1} + \alpha_k p_{k-1};$

$d_k = d_{k-1} - \alpha_k t_{k-1};$

$r_k = A^{\mathrm{T}} d_k;$

$\beta_k = \|r_k\|^2 / \|r_{k-1}\|^2;$

$p_k = r_k + \beta_k p_{k-1};$

$t_k = A p_k;$

end

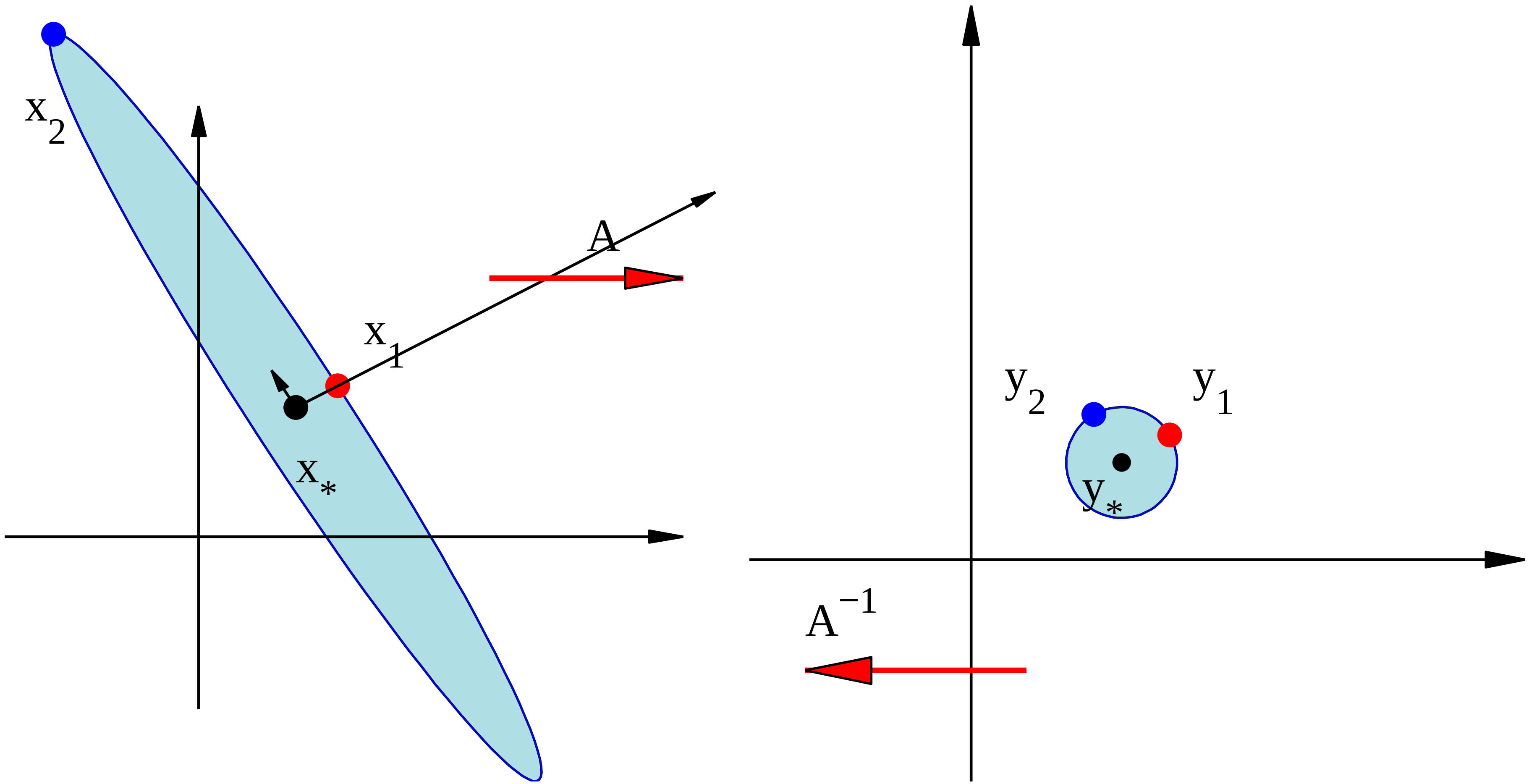
Example: a Toy Problem

An invertible 2×2-matrix A,

$$y_j = Ax_* + \varepsilon_j, \quad j = 1, 2.$$

Preimages,

$$x_j = A^{-1} y_j, \quad j = 1, 2,$$


x
2
A
x
1
x
*
y
2
y
1
y
*
A
−1

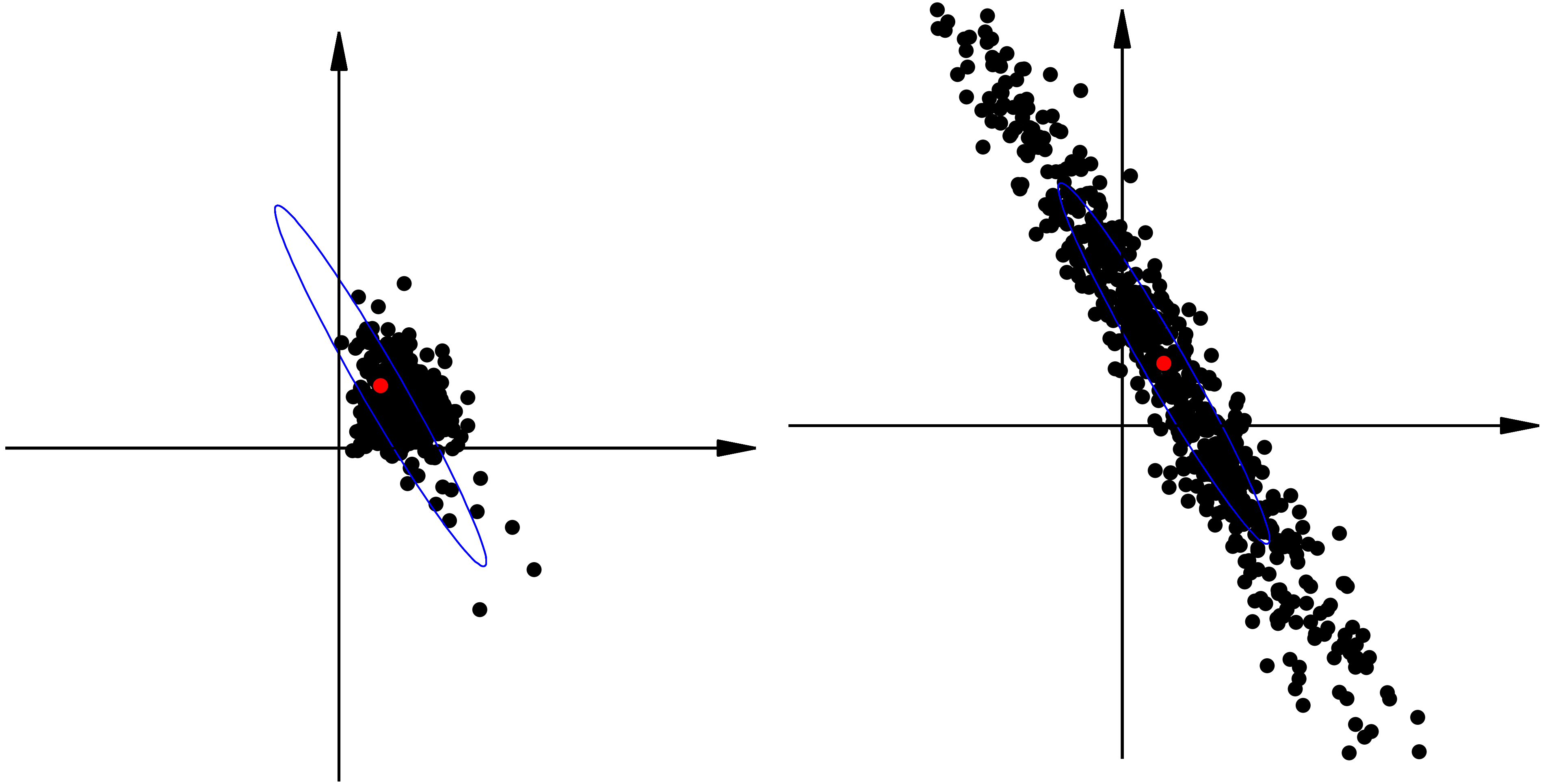
solution by iterative methods: *semiconvergence.*

Write

$$B = Ax_* + E, \quad E \sim \mathcal{N}(0, \sigma^2 I),$$

and generate a sample of data vectors, $b_1, b_2, \ldots, b_n$

Solve with CG



When should one stop iterating?

Let

$$Ax = b_* + \varepsilon = Ax_* + \varepsilon = b,$$

Approximate information

$$\|\varepsilon\| \approx \eta,$$

where $\eta > 0$ is known. Write

$$\|A(x - x_*)\| = \|\varepsilon\| \approx \eta$$

Any solution satisfying

$$\|Ax - b\| \leq \tau\eta$$

is reasonable.

Morozov discrepancy principle

Example: numerical differentiation

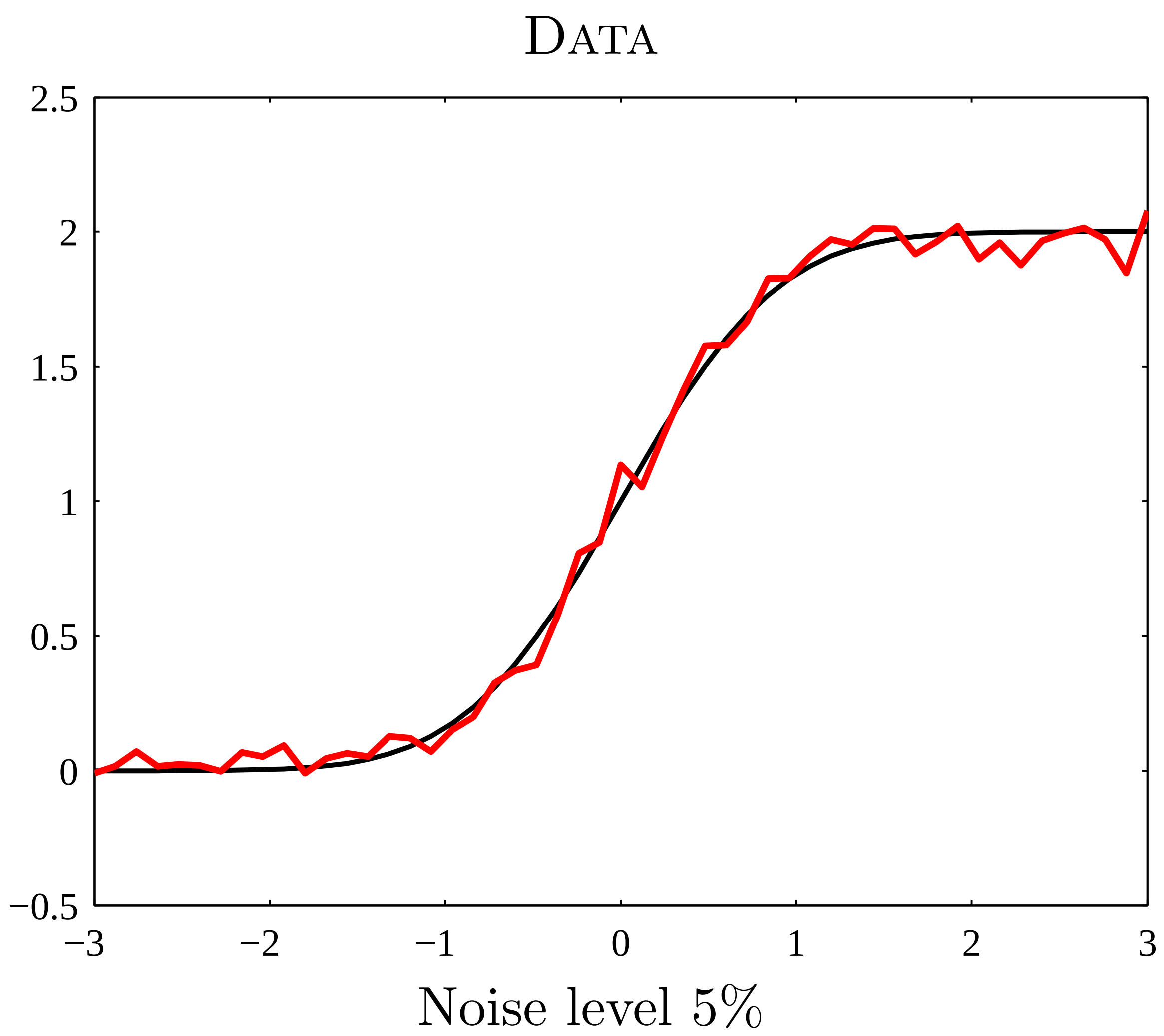
Let $f : [0, 1] \to \mathbb{R}$ be a differentiable function, $f(0) = 0$.

$$\text{Data} = f(t_j) + \text{ noise}, \quad t_j = \frac{j}{n}, \quad j = 1, 2, \ldots n.$$

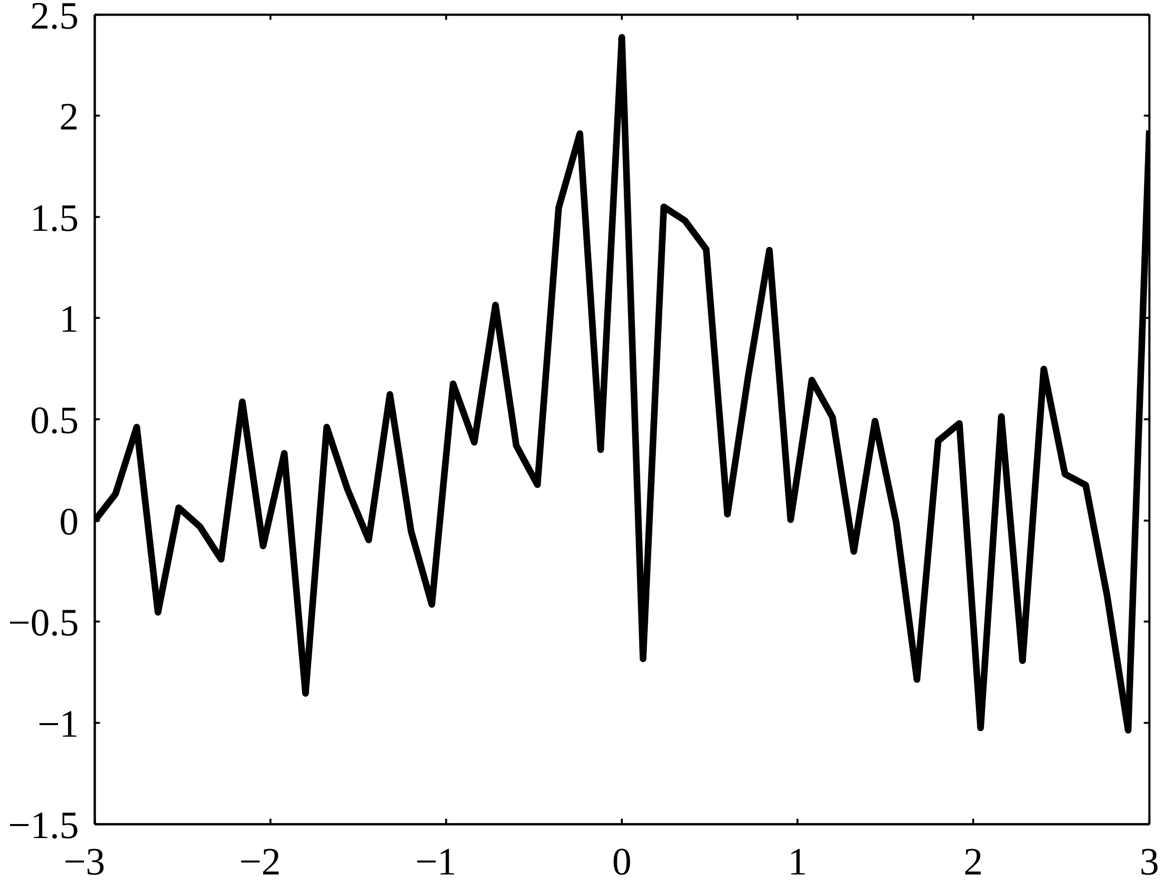
Problem: Estimate $f'(t_j)$.

Direct numerical differentiation by, e.g., finite difference formula does not work: the noise takes over.

Where is the inverse problem?


Data
2.5
2
1.5
1
0.5
0
−0.5
−3
−2
−1
0
1
2
3
Noise level 5%

Solution by finite difference



Formulation as an inverse problem

Denote $g(t) = f'(t)$. Then,

$$f(t) = \int_0^t g(\tau) d\tau.$$

Linear model:

$$\text{Data} = y_j = f(t_j) + e_j = \int_0^{t_j} g(\tau) d\tau + e_j,$$

where e_j is the noise.

DISCRETIZATION

Write

$$\int_0^{t_j} g(\tau)d\tau \approx \frac{1}{n}\sum_{k=1}^{j} g(t_k).$$

By denoting $g(t_k) = x_k$,

$$y = Ax + e,$$

where

$$A = \frac{1}{n}\begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & & \ddots & \\ 1 & & & 1 \end{bmatrix}.$$