## Plan général

## Ill-conditioned problems

- In some areas of applications, linear systems $Ax = b$ ($A \in \mathbb{K}^{m\times n}$, most often $m \geq n$) with "unpleasant" properties, e.g.
  - $\rightarrow$ $A$ has, in theory, full column rank;
  - $\rightarrow$ However, $A$ ill-conditioned with very fast decay of singular values, i.e. numerically rank-deficient:

$$\boxed{\|A - A_r\| \ll \|A\|} \quad \text{for some rank-}r\text{ matrix } A_r,\ r \ll n$$

  - $\rightarrow$ imperfect data $b$ (e.g. measurement errors)
- Such cases occur e.g. for
  - $\rightarrow$ Inverse and identification problems (infer "hidden" physical properties from indirect measurements)
  - $\rightarrow$ Image processing and image restoration
  - $\rightarrow$ Data analysis

- In what follows: least-squares solutions of $Ax = b$ ($A \in \mathbb{K}^{m\times n}$, $m \geq n$, $\text{Rank}(A) = n$).

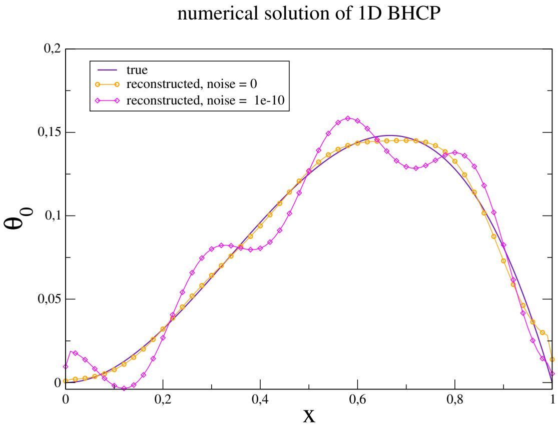
Recall matrix SVD (see lecture 2):

$$A = USV^{\mathsf{H}} = \sum_{i=1}^{n} \sigma_i u_i v_i^{\mathsf{H}} \qquad \begin{cases} U = [u_1, \ldots, u_n] \in \mathbb{K}^{m\times n} \\ V = [v_1, \ldots, v_n] \in \mathbb{K}^{n\times n} \\ S = \text{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{n\times n} \end{cases}$$

## Example: backward heat equation

Physical problem: find the temperature distribution in a system before thermal measurements are made (example: space shuttle re-entry).

$$\underbrace{\Theta(\cdot, T)}_{\text{measurement}} = \underbrace{\mathcal{A}([0,T])}_{\text{heat eq.}} \underbrace{\Theta(\cdot, 0)}_{\text{unknown}}$$



## Example: backward heat equation

Physical problem: find the temperature distribution in a system before thermal measurements are made (example: space shuttle re-entry).

$$\underbrace{\Theta(\cdot,T)}_{\text{measurement}} = \underbrace{\mathcal{A}([0,T])}_{\text{heat eq.}}\underbrace{\Theta(\cdot,0)}_{\text{unknown}}$$

$$\begin{aligned} \kappa\partial_{xx}\Theta - \partial_t\Theta &= 0 && (0\le t\le T,\ 0\le x\le \ell) && \kappa := k/(\rho c) \\ \Theta(0,t) = \Theta(\ell,t) &= 0 && (0\le t\le T) \\ \Theta(x,0) &= \Theta_0(x) && (0\le x\le \ell) \end{aligned}$$

- General solution (Fourier series): $\Theta(x,t) = \sum_{n\ge 0} a_n \sin\frac{n\pi x}{\ell}\, e^{-(n\pi)^2\kappa t/\ell^2}$
- Initial temperature: $\Theta(x,0) = \sum_{n\ge 0} a_n \sin\frac{n\pi x}{\ell}, \quad a_n = \frac{2}{\ell}\int_0^\ell \Theta_0(x)\sin\frac{n\pi x}{\ell}\,\text{d}x$
- Final temperature: $\Theta(x,T) = \sum_{n\ge 0} b_n \sin\frac{n\pi x}{\ell}, \quad b_n = a_n \underbrace{e^{-(n\pi)^2\kappa T/\ell^2}}_{\lambda_n}$

$$\{\, b_0, b_1, b_2 \ldots \,\}^{\mathsf{T}} = \text{diag}\big[\lambda_0, \lambda_1, \lambda_2 \ldots\big]\{\, a_0, a_1, a_2 \ldots \,\}^{\mathsf{T}}$$

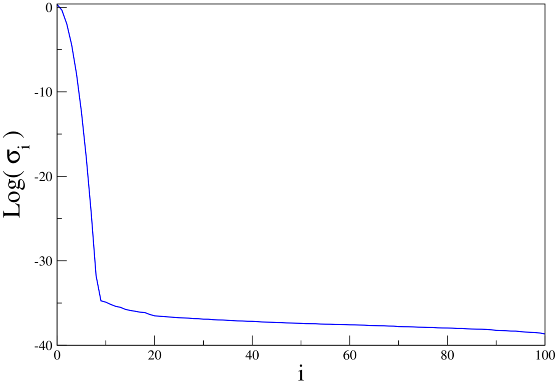
- Reconstruction de $\Theta(\cdot,0)$ given $\Theta(\cdot,T)$ (explicit inversion):

$$\Theta(x,0) = \sum_{n\ge 0} \lambda_n^{-1} b_n \sin\frac{n\pi x}{\ell} \qquad \lambda_n^{-1} = O(e^{Cn^2})\,!$$

## Example: backward heat equation

Physical problem: find the temperature distribution in a system before thermal measurements are made (example: space shuttle re-entry).

$$\underbrace{\Theta(\cdot,T)}_{\text{measurement}} = \underbrace{Acal(T)}_{\text{heat eq.}}\underbrace{\Theta(\cdot,0)}_{\text{unknown}} \quad \Longrightarrow \boxed{A(T)\Theta_0 = \Theta_T} \text{ after space discretization of } \Theta$$




Singular values of $A(T)$ ($x \in [0,1]$, $\Delta x = 1/100$)

- Matrix $A$: exact rank 100, numerical rank $< 10$.

## Sensitivity of least squares solutions to data errors

Goal: solve $\boxed{\min_{x\in\mathbb{K}^n} \|Ax-b\|^2}$ with $A$ "bad" (ill-conditioned, numerically rank-deficient).

Recall (again!) SVD of A: $A = USV^{\mathrm{H}} = \sum_{i=1}^{n} \sigma_i u_i v_i^{\mathrm{H}}$ $\quad \begin{cases} U = [u_1, \ldots, u_n] \in \mathbb{K}^{m\times n} \\ V = [v_1, \ldots, v_n] \in \mathbb{K}^{n\times n} \\ S = \mathrm{diag}(\sigma_1, \ldots, \sigma_n) \in \mathbb{R}^{n\times n} \end{cases}$

- Unique solution ($\mathrm{rank}(A) = n$ assumed), given by
  $$x = \sum_{i=1}^{n} \frac{u_i^{\mathrm{H}} b}{\sigma_i} v_i.$$
- Noisy data $b_\delta = b + w$ with $\|b - b_\delta\|_2 = \|w\|_2 = \delta$ ($\delta$: size of data error).

  Solution error:
  $$x_\delta = \sum_{i=1}^{n} \frac{b_\delta^{\mathrm{H}} u_i}{\sigma_i} v_i, \qquad x_\delta - x = \sum_{i=1}^{n} \frac{w^{\mathrm{H}} u_i}{\sigma_i} v_i, \qquad \frac{\|x_\delta - x\|_2}{\delta} = \frac{1}{\delta}\Big( \sum_{i=1}^{n} \frac{|w^{\mathrm{H}} u_i|^2}{\sigma_i^2} \Big)^{1/2}.$$
- If $A$ numerically rank deficient, may have $\boxed{\dfrac{w^{\mathrm{H}} u_1}{\sigma_1} \text{ small, but } \dfrac{w^{\mathrm{H}} u_i}{\sigma_i} \text{ large for some } i.}$

  In some cases, exponential decay of $\sigma_i$: $|w^{\mathrm{H}} u_i|/\sigma_i$ very large even if $\delta$ small.

## Low-rank approximations

- Often useful to replace $A$ with low-rank approximation $A_r$ with $\|A - A_r\|$ "small enough"
- Natural choice of $A_r$: truncated SVD (TSVD) of $A$

$$\boxed{\widehat{A}_r := U_r S_r V_r = \sum_{i=1}^{r} \sigma_i u_i v_i^{\mathrm{H}}} \qquad \begin{cases} U_r = [u_1, \ldots, u_r] \in \mathbb{K}^{m\times r} \\ V_r = [v_1, \ldots, v_r] \in \mathbb{K}^{n\times r} \\ S_r = \mathrm{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{r\times r} \end{cases}$$

### Eckart-Young-Mirsky theorem

Let $A \in \mathbb{K}^{m\times n}$, $r \leq n$. $\widehat{A}_r$ is best rank-$r$ approximation of $A$ (spectral and Frobenius norms):

$$\widehat{A}_r = \underset{\substack{B\in\mathbb{K}^{m\times n} \\ \mathrm{rank}(B)=r}}{\arg\min} \left\{ \|A - B\|_2 \text{ or } \|A - B\|_{\mathrm{F}}. \right\}; \qquad \|A - \widehat{A}_r\|_2 \leq \sigma_{r+1}, \quad \|A - \widehat{A}_r\|_{\mathrm{F}}^2 \leq \sum_{i=r+1}^{n} \sigma_i^2.$$

- Corresponding estimates for relative truncation errors:

$$\frac{\|A - \widehat{A}_r\|_2}{\|A\|_2} \leq \frac{\sigma_{r+1}}{\sigma_1}, \qquad \frac{\|A - \widehat{A}_r\|_{\mathrm{F}}}{\|A\|_{\mathrm{F}}} \leq \frac{\sqrt{\sum_{i=r+1}^{n} \sigma_i^2}}{\sqrt{\sum_{i=1}^{n} \sigma_i^2}}.$$

Smallest rank $r$ such that $\|A - \widehat{A}_r\| \leq \varepsilon$ can be found knowing $\sigma_1, \ldots, \sigma_n$.

Practical computation of $\widehat{A}_r$ given $A$ potentially expensive (needs SVD of $A$, $O(m^2 n)$ cost).

## Regularized least squares

- Alternative to low-rank truncation by SVD: modified minimization problem (Tikhonov(-Phillips) regularization)

$$\min_{x\in\mathbb{K}^n} J_\alpha(x;b), \quad J_\alpha(x;b) := \|Ax-b\|_2^2 + \alpha\|x\|_2^2 \qquad (\alpha \geq 0 \text{ “small”})$$

- Heuristic idea: $\alpha\|x\|_2^2$ “penalizes” solutions $x$ with $\|x\|$ large.
  Supplementary *prior information*: prefer solutions with smaller $\|x\|$.
- Analysis: use SVD of $A$ (note that $\|x\|_2^2 = \|V^{\mathsf{H}}x\|_2^2$):

$$J_\alpha(x;b) = \sum_{i=1}^{n} \Big\{ |\sigma_i y_i - z_i|^2 + \alpha|y_i|^2 \Big\} + \sum_{i=n+1}^{m} |z_i|^2 \qquad (y_i := v_i^{\mathsf{H}}x,\ z_i := u_i^{\mathsf{H}}b)$$

  Minimization uncouples into $n$ univariate quadratic minimizations, hence

$$y_i = \frac{\sigma_i z_i}{\sigma_i^2+\alpha}, \qquad x_\alpha = \sum_{i=1}^{n} \frac{\sigma_i z_i}{\sigma_i^2+\alpha} v_i$$

- Properties of regularized least squares solution $x_\alpha$:
  - → $x_\alpha$ is unique for any $\alpha > 0$ (even if $\text{Rank}(A) < n$);
  - → For $\alpha > 0$, $x_\alpha$ does not minimize $\|Ax-b\|_2^2$;
  - → Limit of $x_\alpha$ as $\alpha \to 0$ is minimum-norm least-squares solution of $Ax = b$.

## Regularized least squares, noisy data

- Solve $Ax = b_\delta$ with noisy data $b_\delta = b + w$ (with $\|w\|_2 = \delta$).
- Regularized solution for data $b_\delta$:

$$x_{\alpha,\delta} := \arg\min_x J_\alpha(x; b_\delta) = \sum_{i=1}^n \frac{\sigma_i z_i^\delta}{\sigma_i^2 + \alpha} v_i = x_\alpha + \sum_{i=1}^n \frac{\sigma_i (w^H u_i)}{\sigma_i^2 + \alpha} v_i$$

- Regularized solution error $e_{\alpha,\delta} := x_{\alpha,\delta} - x$:

$$e_{\alpha,\delta} = e_{\alpha,\delta}^{\text{reg}} + e_{\alpha,\delta}^{\text{noise}}, \qquad e_{\alpha,\delta}^{\text{reg}} = -\alpha \sum_{i=1}^n \frac{z_i}{\sigma_i(\sigma_i^2 + \alpha)} v_i, \quad e_{\alpha,\delta}^{\text{noise}} = \sum_{i=1}^n \frac{\sigma_i (w^H u_i)}{\sigma_i^2 + \alpha} v_i.$$

Optimal choice method for $\alpha$?

- Define

$$J_\alpha(x_\alpha; b) = D(\alpha) + \alpha R(\alpha), \qquad D(\alpha) := \|Ax_\alpha - b\|_2^2, \quad R(\alpha) := \|x_\alpha\|_2^2$$

  and study behavior of $\alpha \mapsto \{ D(\alpha),\, R(\alpha) \}$

- We find

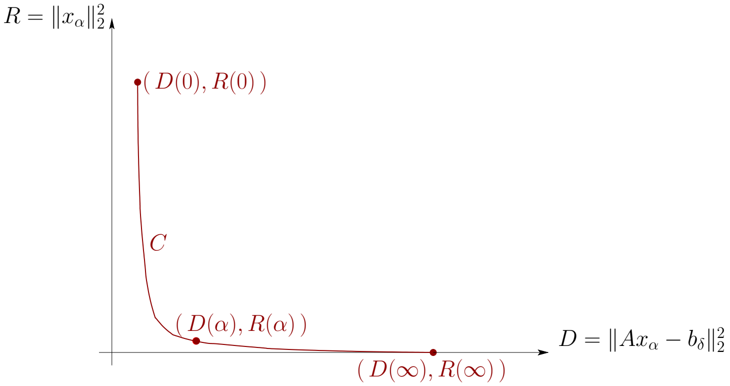
$$D(\alpha) = \sum_{i=1}^{n} \frac{\alpha^2}{(\sigma_i^2 + \alpha)^2} |z_i|^2 + \sum_{i=n+1}^{m} |z_i|^2, \qquad R(\alpha) = \sum_{i=1}^{n} \frac{\sigma_i^2}{(\sigma_i^2 + \alpha)^2} |z_i|^2$$

$$D'(\alpha) = 2\alpha \sum_{i=1}^{n} \frac{\sigma_i^2}{(\sigma_i^2 + \alpha)^2} |z_i|^2 > 0, \qquad R'(\alpha) = -2 \sum_{i=1}^{n} \frac{\sigma_i^2}{(\sigma_i^2 + \alpha)^3} |z_i|^2 < 0$$

- Outcome: $\alpha \mapsto D(\alpha)$ increasing and $\alpha \mapsto R(\alpha)$ decreasing, i.e.:

The L-curve $\alpha \in [0, \infty[ \mapsto \big( D(\alpha), R(\alpha) \big)$ is convex

## Regularized least squares: choice of $\alpha$ using $L$-curve




### L-curve properties

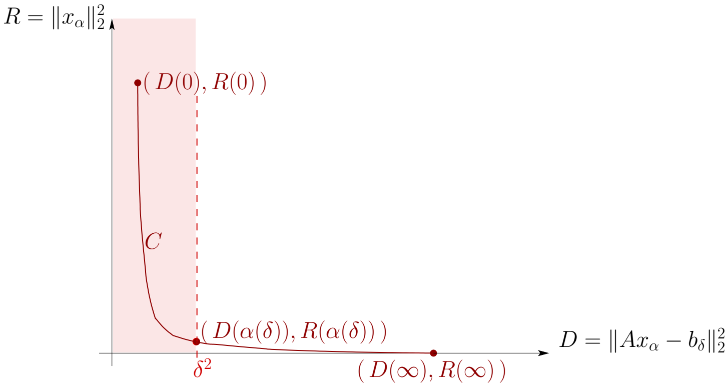
- $C = (D, R)$ is monotonically decreasing ($R$ a decreasing function of $D$) and convex;
- Extremities $A = \big(D(0), R(0)\big)$ and $A = \big(D(\infty), R(\infty)\big)$ of $C$ given by

$$D(0) = \|Ax - b\|_2^2, \quad R(0) = \|x\|_2^2, \qquad D(\infty) = \|b\|_2^2, \quad R(\infty) = 0$$

($x = x_0$: basic least-squares solution).

## Regularized least squares: choice of $\alpha$ using $L$-curve

- Assume data noise level $\delta$ is known (realistic in some cases, e.g. mechanical testing using digital image correlation).
- Use that L-curve is convex, reformulate regularized least-squares:
  $$\min_{x \in \mathbb{K}^n} \|x\|_2^2, \qquad \text{subject to} \quad \|Ax - b\|_2^2 \leq \delta^2$$
- Select $\alpha$ such that $D(\alpha) = \delta^2$ (i.e. set LS residual equal to data noise)
- Unique solution provided $\delta < \|b_\delta\|_2$



## Regularized solution using truncated SVD

- Matrices with fast decay of $\sigma_i$: truncated SVD as alternative to Tikhonov regularization:

$$x_r := \arg\min_{x\in\mathbb{K}^n} \|\widehat{A}_r x - b\|^2 = \sum_{i=1}^{r} \frac{u_i^{\mathrm{H}} b}{\sigma_i} v_i$$

- By analogy to regularized least squares, define

$$D_r := \|\widehat{A}_r x_r - b\|_2^2 = \sum_{i=r+1}^{m} |u_i^{\mathrm{H}} b|^2, \qquad R_r := \|x_r\|_2^2 = \sum_{i=1}^{r} \frac{|u_i^{\mathrm{H}} b|^2}{\sigma_i^2}$$

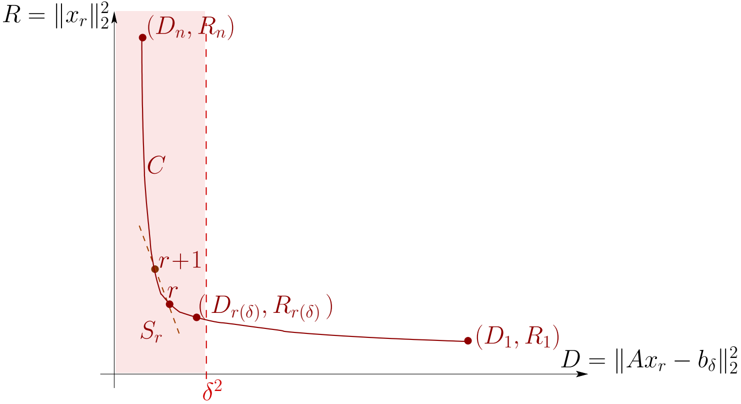
- Clearly $r \mapsto D_r$ decreasing and $r \mapsto R_r$ increasing.
- L-curve $C_n$: interpolates points $(D_r, R_r)$ $(1 \le r \le n)$. $C_n$ is convex:

$$S_r := \frac{R_r - R_{r+1}}{D_r - D_{r+1}} = -\frac{|z_{r+1}|^2}{\sigma_{r+1}^2}\frac{1}{|z_{r+1}|^2} = -\frac{1}{\sigma_{r+1}^2}, \qquad r \mapsto S_r \text{ increasing}$$

- Discrete parameter $1/r$ plays role of regularization parameter $\alpha$. Optimal value:

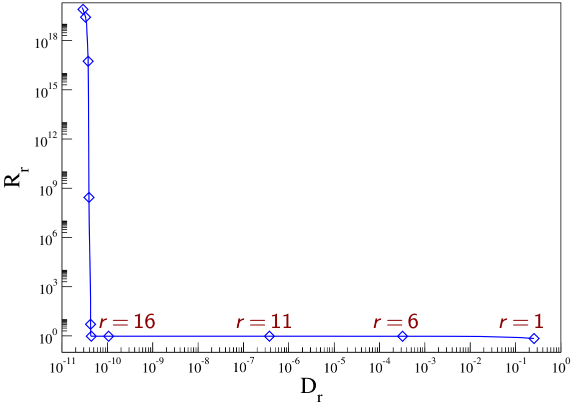
$$r(\delta) = \min_{1\le r\le n} R_r \quad \text{subject to} \quad D_r \le \delta^2$$

## Discrete L-curve



## Example: backward heat equation

Discrete L-curve, simulated data with $\delta = 10^{-5}$,




- Optimal choice of $r$ (L-curve for noise level $\delta = 10^{-5}$);
- Lowest actual temperature reconstruction error: $\approx 10^{-2}$ (in relative $L^2$ norm) for $r = 19$.

## Plan général

Sparsity of (approximate) solutions of linear systems important for some applications:

- Image deblurring:

$$Ax = b + w, \qquad A = BW\ , \quad \begin{cases} W \in \mathbb{R}^{m\times n}: & \text{wavelet basis} \\ B \in \mathbb{R}^{m\times m}: & \text{models blurring} \\ x \in \mathbb{R}^{n}: & \text{restored image} \\ b \in \mathbb{R}^{m}: & \text{blurred image} \\ w \in \mathbb{R}^{m}: & \text{unknown noise} \end{cases}$$

  Goal: find a sparse representation $Wx = \sum_{i=1}^{n} w_i x_i$ of restored image ($x_i = 0$ for many $i$)

- Reflexive idea: regularized least squares (see previous lecture)
  $\min_{x\in\mathbb{R}^n} \|BWx - b\|_2^2 + \alpha\|x\|_2^2$

  However, 2-norm regularizer $\|x\|_2^2$ allows many entries with *small* magnitude.

- Better idea: use 1-norm regularizer instead:

$$\min_{x\in\mathbb{R}^n} J_\alpha(x), \qquad J_\alpha(x) := \|BWx - b\|_2^2 + \alpha\|x\|_1 \quad (\text{“}L^2\text{-}L^1\text{ functional”})$$

- More-difficult minimization problem: $J_\alpha$ not quadratic and not differentiable

## Minimization of functionals with a nonsmooth part

- Recall steepest descent update step for smooth (e.g. quadratic) functionals:

  $$\boxed{x^{(k+1)} = x^{(k)} - t^{(k)}\nabla J(x^{(k)})}\,, \qquad \text{step length } t^{(k)} \text{ found by line search}$$

  Method of explicit type ($\nabla J$ evaluated at initial point).

- $L^2$-$L^1$ functional $J_\alpha$ not everywhere differentiable:
  - → $\nabla J(x^{(k)})$ potentially not defined
  - → $t \mapsto J_\alpha(x(t))$ potentially not differentiable at some $x(t) := x^{(k)} - t\nabla J(x^{(k)})$
- Update step however generalizable to

  $$J(x) = f(x) + g(x) \qquad f, g \text{ convex and } f \text{ differentiable}$$

  Idea: modified update step (explicit for $f$ but implicit for $g$):

  $$\text{(a)}\;\; \hat{x}^{(k)} = x^{(k)} - t\nabla f(x^{(k)}), \qquad \text{(b)}\;\; x^{(k+1)} = \hat{x}^{(k)} - t\nabla g(x^{(k+1)}).$$

  Fix step length $t$, solve (b) for $x^{(k+1)}$.
  - → Trivial (closed-form) if $g$ quadratic
  - → Newton's method if $g$ twice-differentiable

## Minimization of functionals with a nonsmooth part

(a) $\hat{x}^{(k)} = x^{(k)} - t\nabla f(x^{(k)})$, (b) $x^{(k+1)} = \hat{x}^{(k)} - t\nabla g(x^{(k+1)})$.

- If $g$ convex and differentiable, (b) can be reformulated as

  $$\big(x^{(k+1)} - \hat{x}^{(k)}\big) + t\nabla g(x^{(k+1)}) = 0,$$

  equivalent (as necessary and sufficient optimality condition) to

  $$x^{(k+1)} = \arg\min_{x\in\mathbb{R}^n} \Big( \tfrac{1}{2}\|x - \hat{x}^{(k)}\|_2^2 + tg(x) \Big)$$

  Still (uniquely) solvable if $g$ convex and **not** differentiable.

- Define proximal operator $\text{prox}_h : \mathbb{R}^n \to \mathbb{R}$:

  $$\text{prox}_h(y) = \arg\min_{x\in\mathbb{R}^n} \big( \tfrac{1}{2}\|x - y\|_2^2 + h(x) \big) \quad h: \text{ any } \textit{lowersemicontinuous}(\textit{lsc}) \text{ convex function}$$

Update rule for $J(x) = f(x) + g(x)$, g$convexbutpossiblynon-smooth$:

$$\boxed{x^{(k+1)} = \text{prox}_{tg}\big(x^{(k)} - t\nabla f(x^{(k)})\big)}$$

$h$ lsc: Epigraph $(x,t) \in \mathbb{R}^n \times \mathbb{R},\ h(x) \leq t$ closed in $\mathbb{R}^n \times \mathbb{R}$ (among several equivalent definitions).

## Minimization of $L^2$-$L^1$ functionals

Specialize update step to $L^2$-$L^1$ functional $J_\alpha$:

- Major simplification of non-smooth part:

  $$g(x) = \alpha\|x\|_1 = \alpha\big(|x_1| + \ldots + |x_n|\big)$$

- Proximal-operator minimization

  $$\text{prox}_{tg}(y) := \underset{x\in\mathbb{R}^n}{\text{arg min}}\,\big(\tfrac{1}{2}\|x-y\|_2^2 + g(x)\big)$$

  uncouples into $n$ univariate problems

  $$\min_{x_1\in\mathbb{R}^n} \tfrac{1}{2}|x_1 - y_1|^2 + \alpha t|x_1|, \quad \ldots \quad \min_{x_n\in\mathbb{R}^n} \tfrac{1}{2}|x_n - y_n|^2 + \alpha t|x_n|$$

- Univariate proximal operators found in closed form:

  $$\text{prox}_{u\mapsto\alpha t|u|}(y) = \underset{x\in\mathbb{R}}{\text{arg min}}\,\big(\tfrac{1}{2}(x-y)^2 + \alpha t|x|\big) = \begin{cases} 0 & |y| \le \alpha t \\ y(1-\alpha t/|y|) & |y| \ge \alpha t \end{cases}$$

**$L^2$-$L^1$ update step (given step length $t$):**

(a) $\hat{x}^{(k)} = x^{(k)} - tA^{\mathsf{T}}(Ax^{(k)} - b)$, (b) $x_i^{(k+1)} = \begin{cases} 0 & |\hat{x}_i^{(k)}| \le \alpha t \\ \hat{x}_i^{(k)} - \alpha t & \hat{x}_i^{(k)} \ge \alpha t \\ \hat{x}_i^{(k)} + \alpha t & \hat{x}_i^{(k)} \le -\alpha t \end{cases}$.

## Minimization of $L^2$-$L^1$ functionals

$L^2$-$L^1$ update step (given step length $t$):

$$\text{(a)}\quad \hat{x}^{(k)} = x^{(k)} - tA^{\mathsf{T}}(Ax^{(k)} - b), \qquad \text{(b)}\quad x_i^{(k+1)} = \begin{cases} 0 & |\hat{x}_i^{(k)}| \le \alpha t \\ \hat{x}_i^{(k)} - \alpha t & \hat{x}_i^{(k)} \ge \alpha t \\ \hat{x}_i^{(k)} + \alpha t & \hat{x}_i^{(k)} \le -\alpha t \end{cases}.$$

- $\text{prox}_{u\mapsto\alpha t|u|}(y) = 0$ whenever $|y| < \alpha t$: sparsity-promoting mechanism of $L^2$-$L^1$ minimization.
- Reduce $\alpha \implies$ weaker sparsity promotion.

# Minimization of $L^2$-$L^1$ functionals

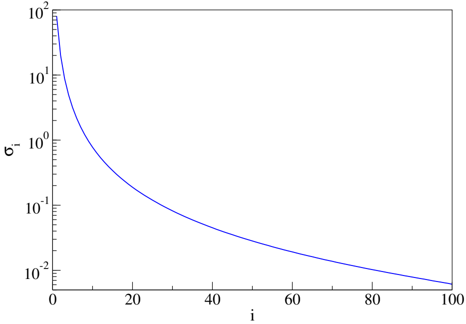
**Algorithm 16** FISTA iterations for the $L^2$-$L^1$ minimization problem (Beck, Teboulle 2009)

$A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $\alpha > 0$, $x^{(0)} \in \mathbb{R}^n$ (data and initial guess)
$L = \|A^\mathsf{T} A\|_2$ (spectral radius of $A^\mathsf{T} A$: find largest eigenvalue of $A^\mathsf{T} A$)
$t = \alpha/L$ (step length, maximum permissible value)
$y^{(1)} = x^{(0)}$, $s^{(1)} = 1$ (first iteration)
**for** $k = 1, 2, \ldots$ **do**
    $\hat{x}^{(k)} = x^{(k-1)} - tA^\mathsf{T}(Ax^{(k-1)} - b)$ (explicit step)
    $x^{(k)} = \text{prox}_{u\mapsto t\|u\|}(\hat{x}^{(k)})$ (apply proximal operator)
    $s^{(k+1)} = \frac{1}{2}\big(1 + \sqrt{1+4s^{(k)2}}\big)$ (update algorcolor parameter $s^{(k)}$)
    $y^{(k+1)} = x^{(k)} + \frac{s^{(k)}-1}{s^{(k+1)}}\big(x^{(k)} - x^{(k-1)}\big)$
    **If convergence test satisfied:** return $x = x^{(k)}$
**end for**

## “random” example

$$J_\alpha(x) = \frac{1}{2}\|Ax - b\|_2^2 + \alpha\|x\|_1, \quad A \in \mathbb{R}^{m\times n} \ (m = 5000,\ n = 100).$$

$$\boxed{\min_{x\in\mathbb{R}^n} \|Ax - b\|_2 \approx 77.51}$$




Singular values of $A$

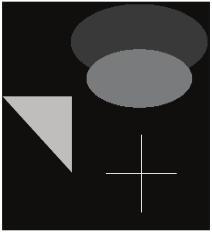
## “random” example

$$J_\alpha(x) = \frac{1}{2}\|Ax - b\|_2^2 + \alpha\|x\|_1, \quad A \in \mathbb{R}^{m\times n} \ (m = 5000,\ n = 100). \tag{7}$$

$$\boxed{\min_{x\in\mathbb{R}^n} \|Ax - b\|_2 \approx 77.51} \tag{8}$$

| $\alpha$ | nombre $x_i^\alpha \neq 0$ | $\|Ax^\alpha - b\|_2$ | $\|x^\alpha\|_1$ | nombre iters. |
|---|---|---|---|---|
| 0 | 100 | 77.51 | 2496 | N/A |
| 0.1 | 19 | 78.05 | 62.97 | 67 181 |
| 0.25 | 11 | 78.14 | 13.54 | 30 640 |
| 0.5 | 7 | 78.19 | 3.559 | 23 875 |
| 1 | 3 | 78.21 | 0.452 | 3 439 |
| 5 | 1 | 78.22 | 0.11 | 464 |
| 10 | 1 | 78.23 | 0.0136 | 198 |

## Image restoration example (Beck, Teboulle 2009)

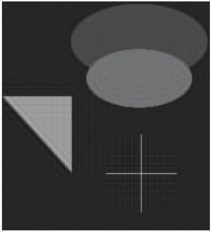


Original

Blurred

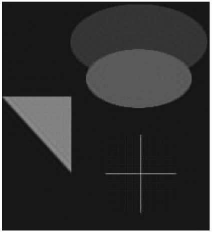
MTWIST: $F_{100} = 3.83\text{e-}1$

MTWIST: $F_{200} = 3.41\text{e-}1$

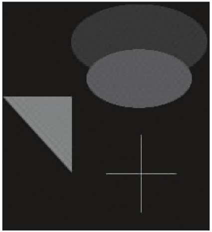


with alternative algorithm

FISTA: $F_{100} = 3.21\text{e-}1$



FISTA: $F_{200} = 3.09\text{e-}1$



with FISTA

## Plan général

**Partie 1:** Généralités
- Séance 1: Motivations et exemples. Généralités sur le calcul matriciel numérique
- Séance 1b: TP numérique 0 (prise en main de Julia) **(LF)**

**Partie 2:** Méthodes directes
- Séance 1: Généralités, factorisation LU
- Séance 2a: Factorisations LU et $LDL^T$
- Séance 2b: Factorisation QR et problèmes de moindres carrés
- Séance 3b: TP numérique 1 **(LF)**

**Partie 3:** Méthodes itératives
- Séance 3a: Méthodes de type point fixe; gradient conjugué
- Séances 4a, 4b: Gradient conjugué, GMRES
- Séance 5b: TP numérique 2 **(LF)**

**Partie 4:** Problèmes aux valeurs et vecteurs propres
- Séance 5a: Généralités, Puissances itérées, puissances inverses
- Séance 6a: Itérations orthogonales et algorithme QR

**Partie 5:** Systèmes linéaires mal conditionnés
- Séance 6b Compression et approximation de systèmes mal conditionnés.
- Séance 7a: Recherche de solutions parcimonieuses .