Lecture 2
Regularization

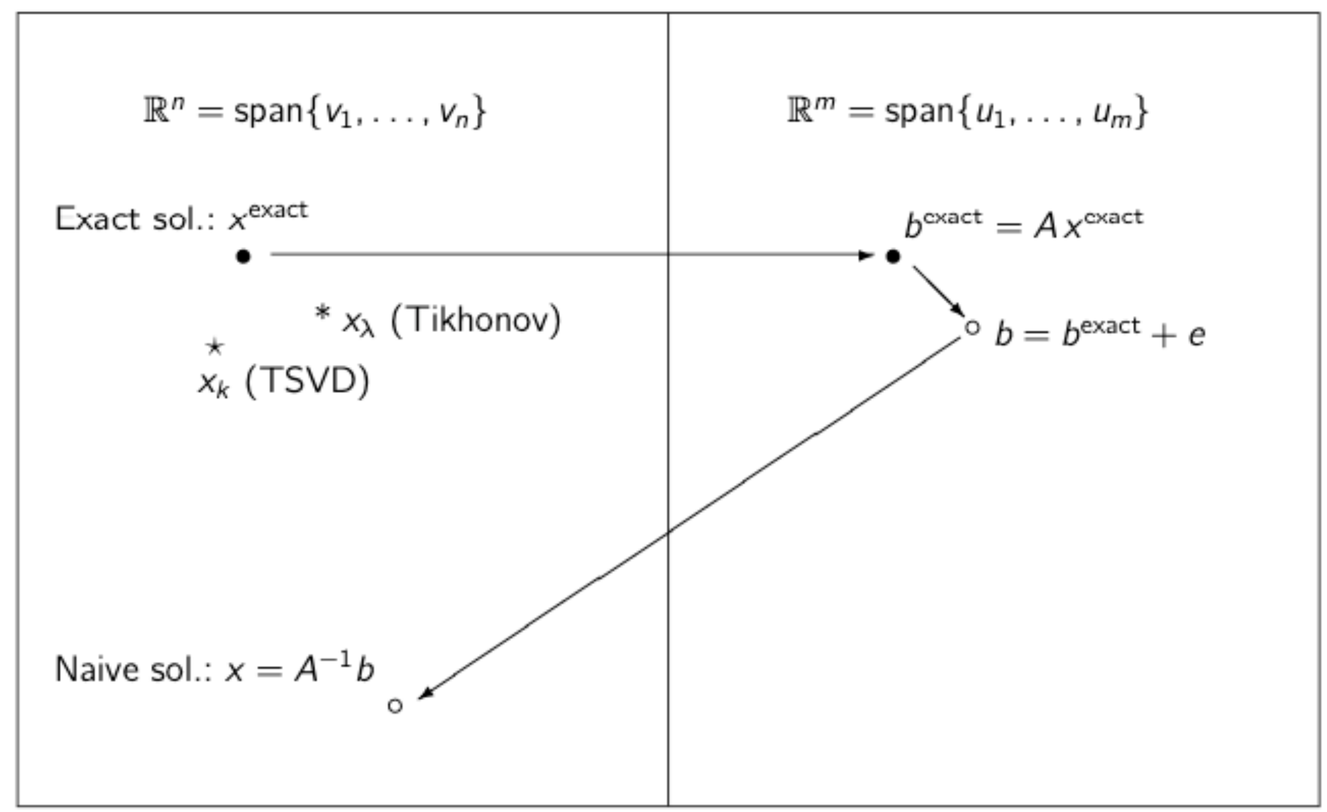
Topics in Inverse Problems
Spring 2020

Last time

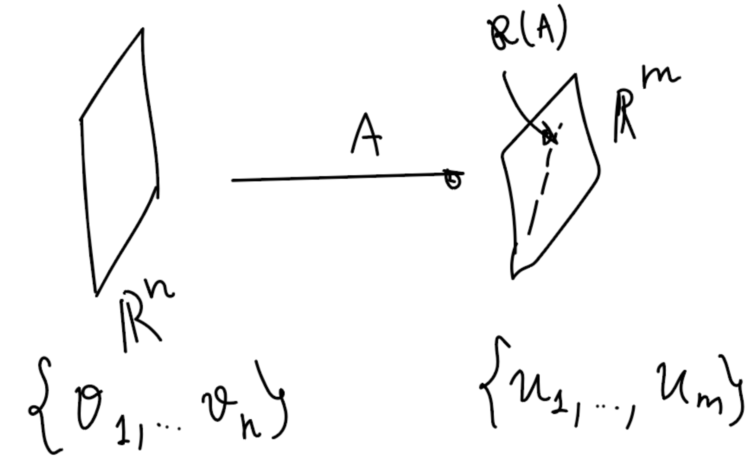
- Well-posed problems: Hadamard's criterion
- Integral equations: $g(t) = (Af)(t) = \int K(s,t)f(s)ds$
- $A : X \to Y$ is a **compact operator** (X,Y Hilbert spaces)
    - A^{-1} is always discontinuous (unbounded), unless $\dim \mathscr{R}(\mathscr{A}) < \infty$.
    - The problem is always ill-posed
- Singular Value Expansion ("operator SVD")

Today's lecture

1. Picard's theorem - solvability of $Ax = y$.
   - Least squares solution: $x = A^{\dagger}y$ (pseudo-inverse) is **unstable**
2. Regularisation: restoring stability in practice
3. Convergence and optimality
4. Choosing the regularisation parameter



Operator SVD



Finite-dimensional setting: $A \in \mathbb{R}^{m \times n} (m \geq n)$

$A = U \Sigma V^T$, U, V **orthogonal,** Σ **diagonal**

$$Ax = \sum_{i=1}^{n} \mu_i \left\langle x, v_i \right\rangle u_i \, .$$

Let $A : X \to Y$ **be compact. Let** $\{\mu_i\}_{i=1}^{\infty}$ **be the nonzero singular values of** A**, repeated according to multiplicity:** $\mathrm{mult}(\mu_i) = \dim \mathcal{N}(\mu_i^2 - A^*A) < \infty$**. Then**

1. **There exist orthonormal sequences** $\{\varphi_i\} \subset X, \{\psi_i\} \subset Y$ **(right and left singular functions, resp.) such that**

$$A\varphi_i = \mu_i \psi_i, \quad A^*\psi_i = \mu_i \varphi_i.$$

2. **For each** $\varphi \in X$ **there holds** $\varphi = \sum \left\langle \varphi, \varphi_i \right\rangle \varphi_i + Q\varphi$**, where** Q **is the orthogonal projection onto** $\mathcal{N}(A)$**, and**

$$A\varphi = \sum_{i=1}^{\infty} \mu_i \left\langle \varphi, \varphi_i \right\rangle \psi_i \, .$$

Picard's theorem

$A : X \to Y$ **compact, with SVE** $\{\mu_i, \varphi_i, \psi_i\}$**. The equation** $Ax = y$ **is solvable iff**

1. $y \in \mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$**;**
2. $\sum_{i=1}^{\infty} \frac{1}{\mu_i^2} \left| \langle y, \psi_i \rangle \right|^2 < \infty$. **(Picard's criterion)**

In this case, a solution is given by

$$x = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \langle y, \psi_i \rangle \varphi_i \,.$$

=>: $Ax = y$ **solvable then** $y \in \mathcal{R}(A) \subset \mathcal{N}(A^*)^\perp$**.**

Let φ **be a solution, then**

$$\langle \varphi, \varphi_i \rangle = \frac{1}{\mu_i} \langle \varphi, A^*\psi_i \rangle = \frac{1}{\mu_i} \langle A\varphi, \psi_i \rangle = \frac{1}{\mu_i} \langle y, \psi_i \rangle,$$

and so $\sum_{i=1}^{\infty} \left| \langle \varphi, \varphi_i \rangle \right|^2 = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \left| \langle y, \psi_i \rangle \right|^2 \leq \|\varphi\|^2 < \infty$**.**

(Parseval: $\|\varphi\|^2 = \sum_{i=1}^{\infty} \left| \langle \varphi, \varphi_i \rangle \right|^2 + \|Q\varphi\|^2$**)**

Picard's theorem

$A : X \to Y$ **compact, with SVE** $\{\mu_i, \varphi_i, \psi_i\}$**. The equation** $Ax = y$ **is solvable iff**

1. $y \in \mathcal{N}(A^*)^\perp = \overline{\mathcal{R}(A)}$**;**

2. $\sum_{i=1}^{\infty} \frac{1}{\mu_i^2} \left| \langle y, \psi_i \rangle \right|^2 < \infty$. **(Picard's criterion)**

In this case, a solution is given by

$$x = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \langle y, \psi_i \rangle \varphi_i .$$

<u><=:</u> Suppose $y \in \mathcal{N}(A^*)^\perp$**.**

Then $y = \sum_{i=1}^{\infty} \langle y, \psi_i \rangle \psi_i$ *(why?),*

and so

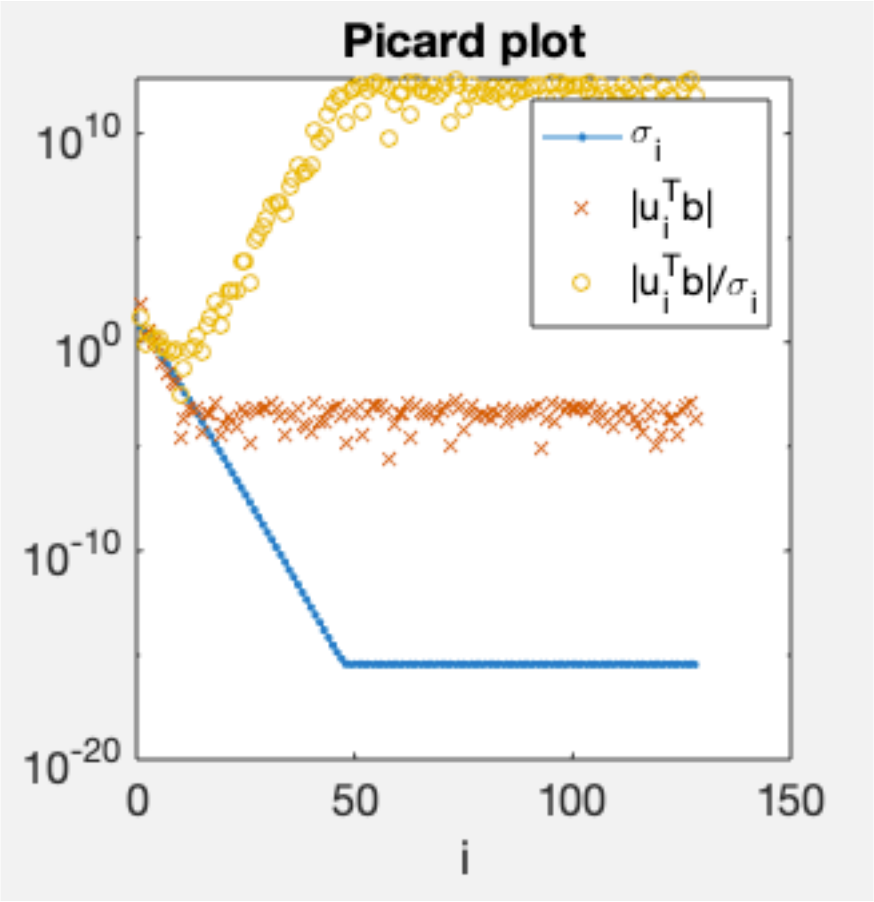
$$Ax = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \langle y, \psi_i \rangle A\varphi_i = y.$$

Picard's criterion

$$\sum_{i=1}^{\infty} \frac{1}{\mu_i^2} \left| \langle y, \psi_i \rangle \right|^2 < \infty \, .$$

Coefficients must decay very fast => data must be very smooth.

In practice: $y^\delta = y + \delta y$, and so PC is violated



Stability in finite dimensions

$Ax = y, A \in \mathbb{R}^{m \times n}$ $\qquad \|y_\delta - y\| \leq \delta$ $\qquad Ax = \sum_{i=1}^{n} \mu_i \left\langle x, v_i \right\rangle u_i \, .$

Direct solution: $x^\delta = A^\dagger y^\delta, \quad A^\dagger = (A^T A)^{-1} A^T$ **(Moore-Penrose pseudo-inverse)**

Problem: this can be very unstable!

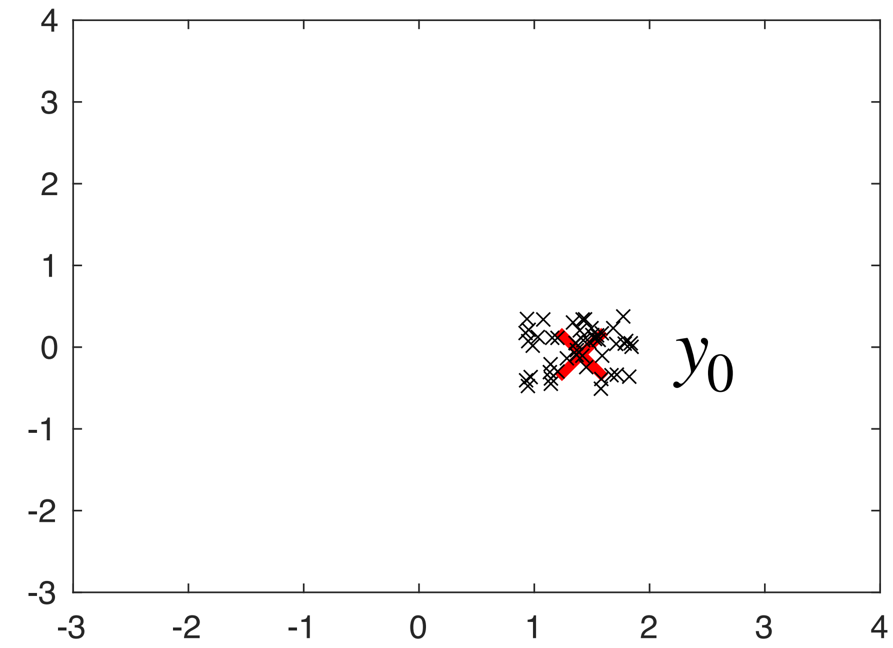
$$x^\delta = A^\dagger y^\delta = \sum_{i=1}^{n} \frac{1}{\mu_i} \left\langle y^\delta, u_i \right\rangle v_i$$

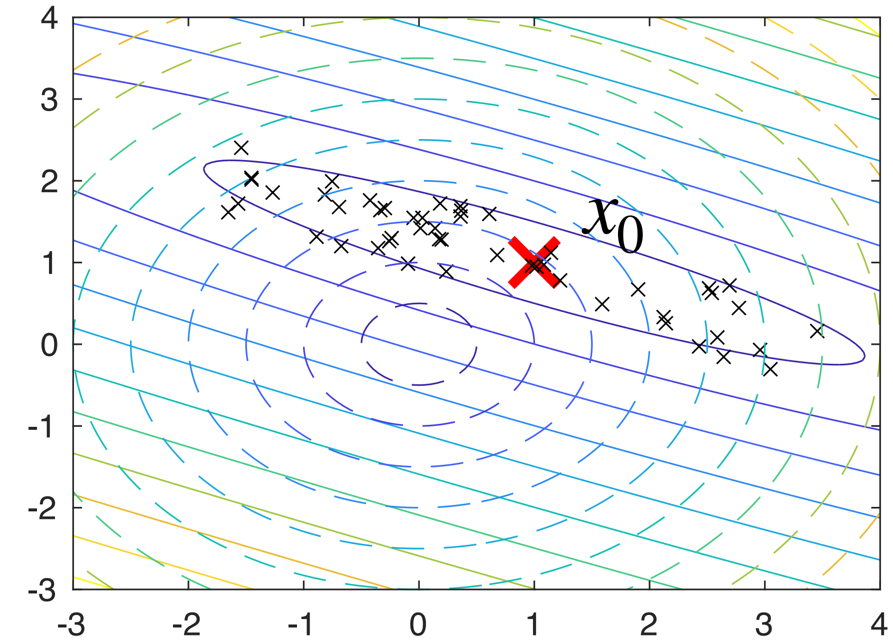
$$x^\delta - x = \sum_{i=1}^{n} \frac{1}{\mu_i} \left\langle y^\delta - y, u_i \right\rangle v_i \quad \Longrightarrow \quad \|x^\delta - x\|^2 = \sum_{i=1}^{n} \frac{1}{\mu_i^2} \left| \langle y^\delta - y, u_i \rangle \right|^2 \leq \frac{1}{{\mu_n}^2} \|y^\delta - y\|^2$$

As $n \to \infty$**,** $\mu_n \to 0$ **and so this is no good!**

Need for prior information

- Problem: $A^\dagger$ can amplify noise
- Possible solutions:
  1. ignore small singular values
  2. restrict $\|x\| \leq \rho$
  3. restrict $\|Ax - y\| \leq \delta$



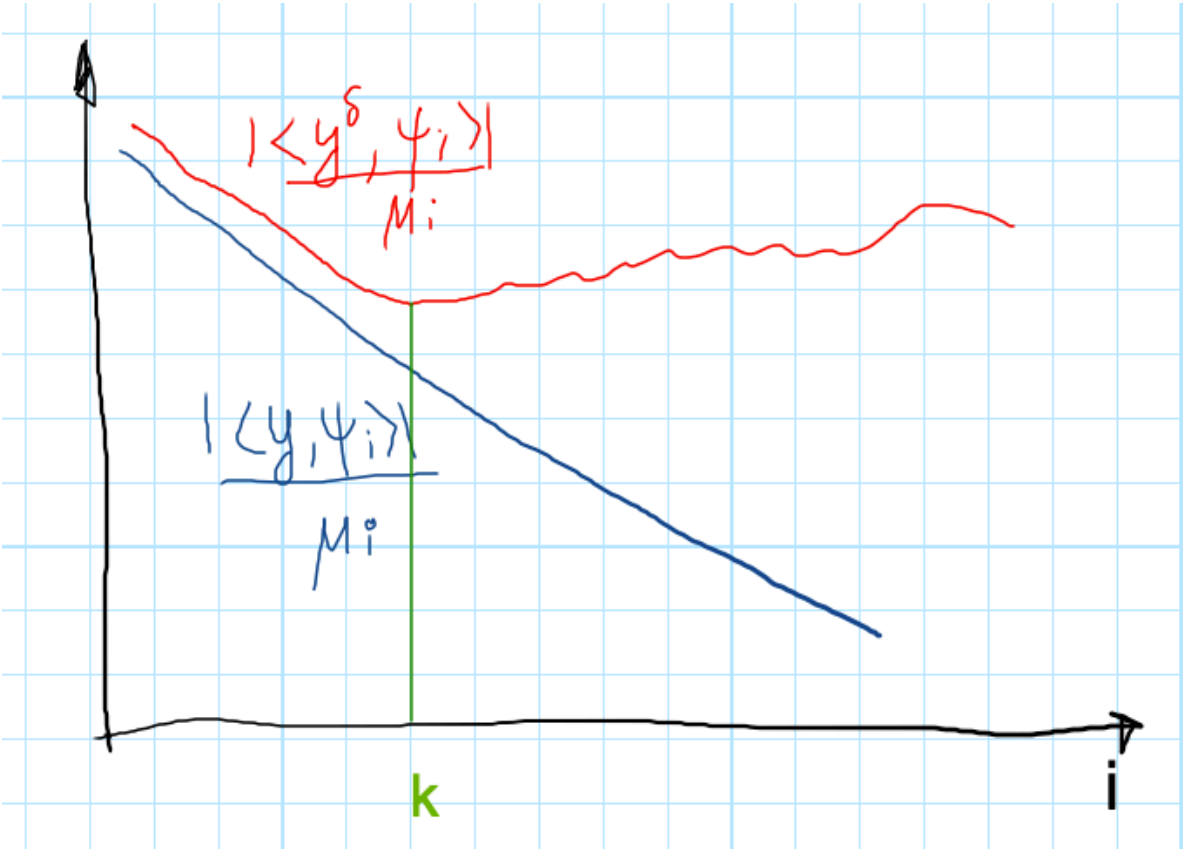


Truncated SVD

$$A^{-1}y = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \langle y, \psi_i \rangle \varphi_i \qquad y^\delta = y + \delta y, \quad \|\delta y\| \leq \delta$$

We assume $y \in \mathcal{R}(A)$

Picard's plot



- Choose truncation index k
- Chop off components with $\{\mu_i\}_{i>k}$
- Solution is $\displaystyle\sum_{i=1}^{k} \frac{1}{\mu_i} \langle y^\delta, \psi_i \rangle \varphi_i$
- How to choose k **optimally?**

First, need convergence

$$q(\alpha,\mu) = \begin{cases} 1 & \alpha \le \mu \\ 0 & \text{else} \end{cases}.$$

(filter)

$$x = \sum_{i=1}^{\infty} \frac{1}{\mu_i} \left\langle y, \psi_i \right\rangle \varphi_i$$

$$x^\alpha = \sum_{i=1}^{\infty} \frac{q\left(\alpha, \mu_i\right)}{\mu_i} \langle y^\delta, \psi_i \rangle \varphi_i =: R_\alpha y^\delta$$

$$\delta x = x^\alpha - x = \sum_{\mu_i \ge \alpha} \frac{1}{\mu_i} \left(\langle y + \delta y, \psi_i \rangle - \langle y, \psi_i \rangle \right) \varphi_i - \sum_{\mu_i < \alpha} \frac{1}{\mu_i} \langle y, \psi_i \rangle \varphi_i$$

$$= \sum_{\mu_i \ge \alpha} \frac{1}{\mu_i} \langle \delta y, \psi_i \rangle \varphi_i - \sum_{\mu_i < \alpha} \frac{1}{\mu_i} \langle y, \psi_i \rangle \varphi_i$$

$$\downarrow \qquad\qquad \downarrow$$

$$R_\alpha(\delta y) \qquad\qquad R_\alpha A x - x$$

(noise amplification error) **(truncation error)**

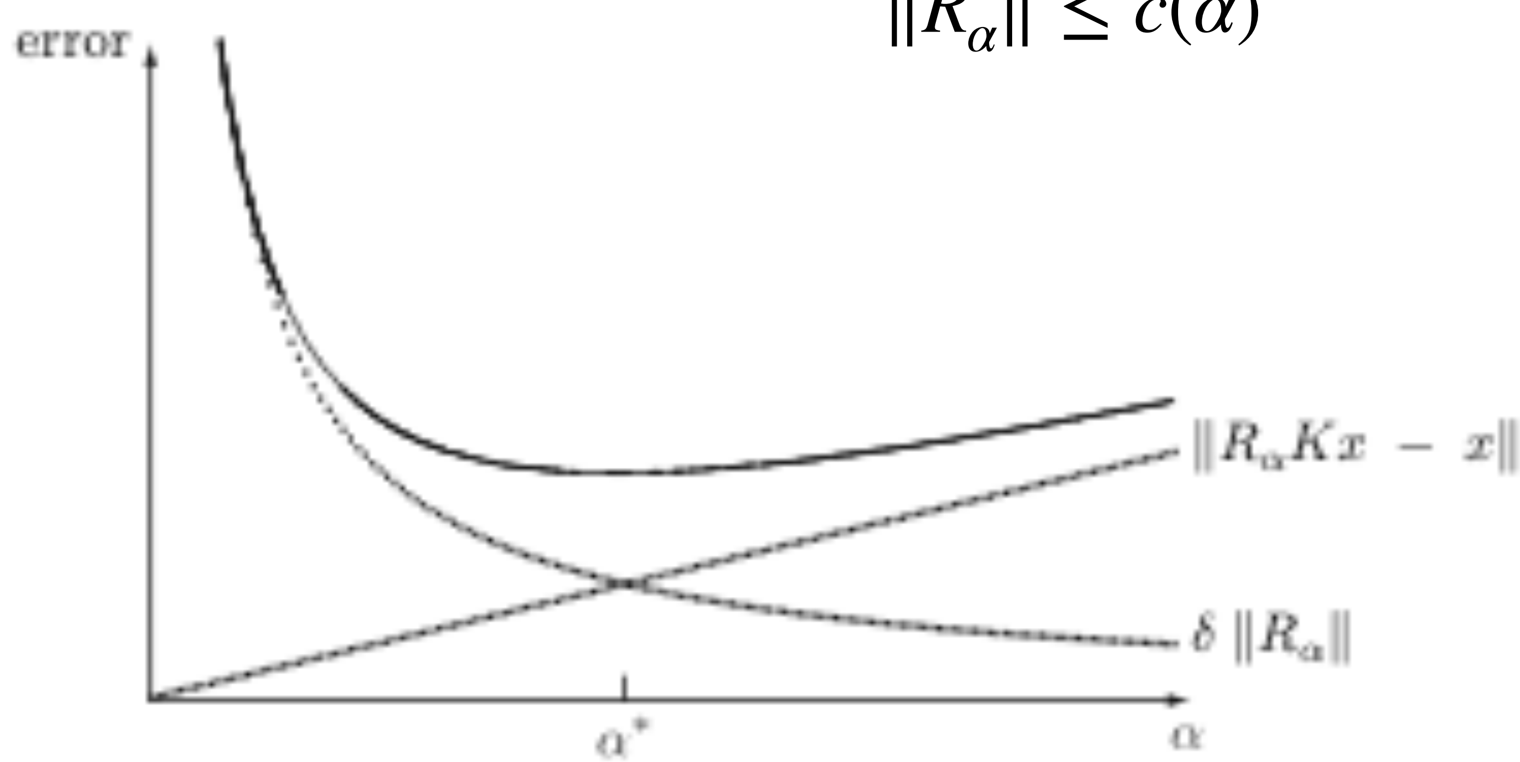
Regularization schemes

$$\|\delta x\| \leq \|R_\alpha \delta y\| + \|R_\alpha A x - x\| \qquad \|\delta y\| \leq \delta$$

Regularization strategy (scheme):

$R_\alpha : Y \to X$ **which satisfies** $\lim_{\alpha \to 0} R_\alpha A x = x$ **for all** $x \in X$

$$\|R_\alpha\| \leq c(\alpha)$$



Convergence: choose $\alpha = \alpha(\delta)$ **such that** $\|\delta x\| \to 0$ **as** $\delta \to 0$.

Cntd.

$$\|\delta x\| \leq \|R_\alpha \delta y\| + \|R_\alpha A x - x\| \qquad \|\delta y\| \leq \delta$$

$$\|R_\alpha \delta y\|^2 = \sum_{\mu_i \geq \alpha} \frac{1}{\mu_i^2} |\langle \delta y, \psi_i \rangle|^2 \leq \frac{1}{\alpha^2} \|\delta y\|^2$$

(Note: $\|R_\alpha\| \leq \dfrac{1}{\alpha}$**)**

Convergence: any $\alpha(\delta) \to 0$ **such that** $\dfrac{\delta}{\alpha(\delta)} \to 0$ **as** $\delta \to 0$.

For example, $\alpha(\delta) = \delta^p, \quad 0 < p < 1$.

Q: Why do we need $\alpha(\delta) \to 0$*?*

Next question: rate of convergence

Range/source conditions

$$\|\delta x\| \leq \|R_\alpha \delta y\| + \|R_\alpha A x - x\|$$

$$\|R_\alpha A x - x\|^2 = \sum_{\mu_i<\alpha} \frac{1}{\mu_i^2} \left|\langle y, \psi_i \rangle\right|^2 = \sum_{\mu_i<\alpha} \frac{1}{\mu_i^2} \left|\langle Ax, \psi_i \rangle\right|^2 = \sum_{\mu_i<\alpha} \frac{1}{\mu_i^2} \left|\langle x, A^* \psi_i \rangle\right|^2 = \sum_{\mu_i<\alpha} \left|\langle x, \varphi_i \rangle\right|^2$$

Can be arbitrarily slow!

Range/source condition: $x = A^* z$ **with** $\|z\| \leq E$. ← **A-priori information (e.g. smoothness)**

(Can have more general constraint: $\|Bx\|_Z \leq E,\ B : X \to Z$)

In this case: $\sum_{\mu_i<\alpha} \left|\langle x, \varphi_i \rangle\right|^2 = \sum_{\mu_i<\alpha} \mu_i^2 \left|\langle z, \psi_i \rangle\right|^2 \leq \alpha^2 \|z\|^2 = \alpha^2 E^2$.

Finally: $\|\delta x\| \leq \dfrac{\delta}{\alpha} + \alpha E$.

Need to choose $\alpha(\delta)$ **to balance the terms.**

$$\alpha(\delta) = c\sqrt{\frac{\delta}{E}}$$ → $$\|\delta x\| \leq (c + 1/c)\sqrt{\delta E}\,.$$

Is it optimal?

Worst-case error

Definition: $\mathscr{F}(E, \delta, \|\cdot\|_*) = \sup\left\{\|x\| : \quad \|Ax\| \leq \delta,\ \|x\|_* \leq E\right\}.$

Claim: **assume $x \in X$ satisfies $\|Ax - y^\delta\| \leq \delta,\ \|x\|_* \leq E$ for some $y^\delta \in Y$.**

Then for any $x_1 \in X$ satisfying these conditions, we have

$$\|x - x_1\| \leq 2\mathscr{F}(E, \delta, \|\cdot\|_*) .$$

Note: $\|\cdot\|_*$ **should be "stronger" than $\|\cdot\|$, otherwise $\mathscr{F} \not\to 0$ as $\delta \to 0$.**

$$\|x\| \leq c\|x\|_*$$

Worst-case error

<u>Definition:</u> $\mathscr{F}(E, \delta, \|\cdot\|_*) = \sup\left\{\|x\| : \quad \|Ax\| \leq \delta,\ \|x\|_* \leq E\right\}.$

<u>Theorem:</u> Let $A : X \to Y$ **compact, injective, with dense range.**
Put $\|x\|_* := \|(A^*)^{-1}x\|$**. Then**

$$\mathscr{F}(E, \delta, \|\cdot\|_*) \leq \sqrt{\delta E}\,.$$

<u>Proof:</u> let $x = A^*z$ **with** $\|z\| \leq E$**. Then**

$$\|x\|^2 = \langle A^*z, x\rangle = \langle z, Ax\rangle \leq \|z\|\,\|Ax\| = \delta E\,.$$

<u>Optimality:</u> consider $x_n = \mu_n E\varphi_n = EA^*\psi_n$**, and** $\delta_n = \mu_n^2 E$**.**
Then $\|Ax_n\| = \delta_n\|x_n\| = \delta_n E$**.**

Summary thus far

$$x^{\delta,\alpha(\delta)} = \sum_{i=1}^{\infty} \frac{q\left(\alpha, \mu_i\right)}{\mu_i} \langle y^\delta, \psi_i \rangle \varphi_i =: R_{\alpha(\delta)} y^\delta$$

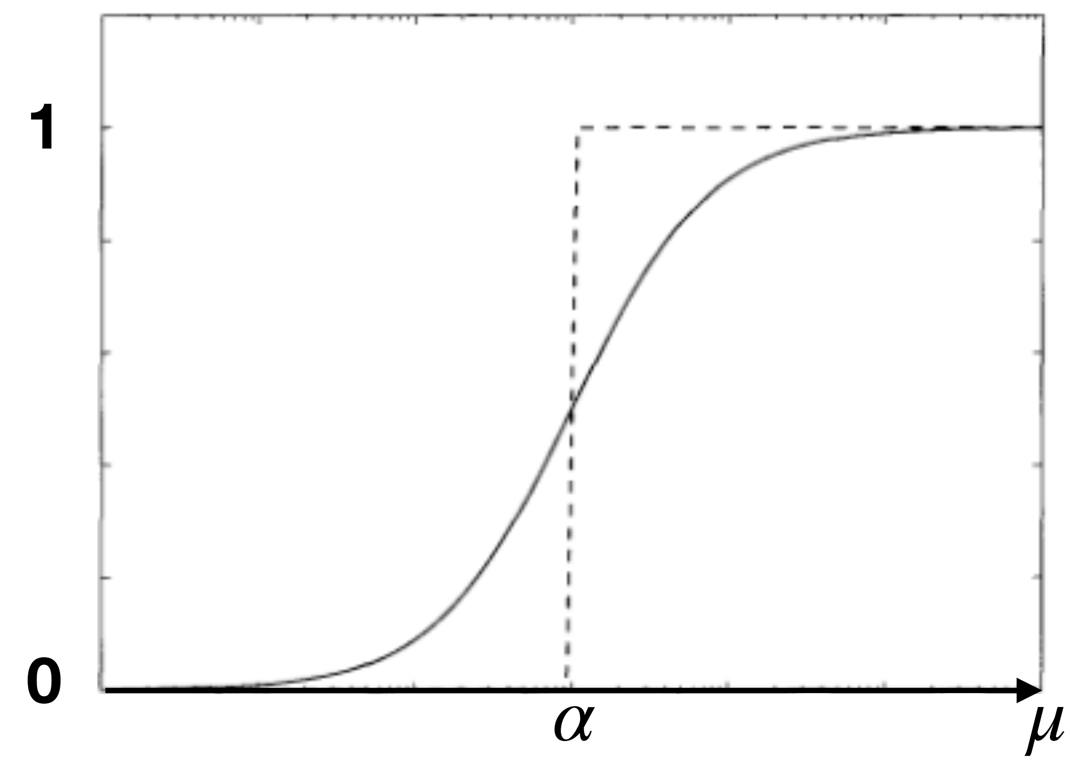
- Started with an ill-posed equation $Ax = y$, $y^\delta = y + \delta y$
- Constructed a sequence of “regularized” solutions $x^{\delta,\alpha(\delta)} = R_{\alpha(\delta)} y^\delta$ and showed that $x^\delta \to x$ as $\delta \to 0$.
- Showed that for certain “prior” information $\|Bx\| \leq E$ we can choose $\alpha(\delta)$ to have the <u>optimal</u> rate $\|x^\delta - x\|$
- This works for general operators B and filters $q(\alpha, \mu)$ satisfying certain technical conditions (not too complicated…)

Tikhonov regularisation

Three points of view

$$x^{\delta,\alpha} = \sum_{i=1}^{\infty} \frac{q\left(\alpha, \mu_i\right)}{\mu_i} \langle y^\delta, \psi_i \rangle \varphi_i =: R_\alpha y^\delta$$

$$q(\alpha, \mu) = \frac{\mu^2}{\mu^2 + \alpha^2}$$



Replace ill-posed problem $Ax^\delta = y^\delta$ **whose solution is** $x^\delta = (A^*A)^{-1}A^*y^\delta$

with $x^{\delta,\alpha} = (A^*A + \alpha^2 I)^{-1}A^*y^\delta$

$$x^{\delta,\alpha} = \arg\min_x \|Ax - y^\delta\|^2 + \alpha^2\|x\|^2$$

(penalize large x**)**

Tikhonov regularisation

$$x^{\delta,\alpha} = (A^*A + \alpha^2 I)^{-1} A^* y^\delta$$

Can be shown that:

Convergence: any $\alpha(\delta) \to 0$ **such that** $\dfrac{\delta}{\alpha(\delta)} \to 0$ **as** $\delta \to 0$.

For example, $\alpha(\delta) = \delta^p, \quad 0 < p < 1.$

Optimality under range conditions:

$$\|(A^*)^{-1}x\| \leq E \qquad \alpha(\delta) = c\sqrt{\frac{\delta}{E}} \qquad \|\delta x\| \leq (c + 1/c)\sqrt{\delta E}\,.$$

Discrepancy principle

- Can we have a-posteriori parameter choice strategy?
- Idea: make sure that $||Ax^{\delta,\alpha} - y^{\delta}|| = \delta$, where δ is an upper bound on the *discrepancy of the true solution.*
- Turns out that we can solve this nonlinear equation easily
- Can show convergence and optimality