Recap

A : input space

B : data space

$F : A \longrightarrow B$ (linear)

IP: $b = F a^+ + \varepsilon \overset{?}{\rightsquigarrow} a^+$

F : $+ / \int$ ✓

F^{-1} : $- / \frac{d}{dx}$ ✗

Naive approach:

$$a_{rec} = F^{-1} b = \dots = a^+ + F^{-1}\varepsilon \qquad (F a_{rec} = b)$$

2. Classical approaches: regularization

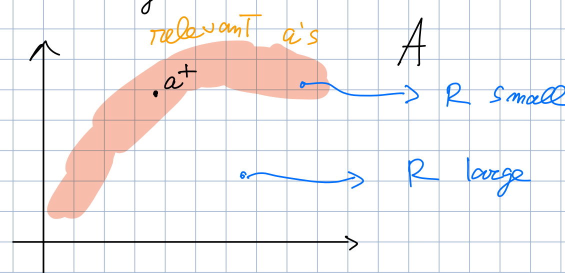
→ is relaxed

$$a_{rec} = \underset{a}{\operatorname{argmin}} \underbrace{\frac{1}{2}\|Fa - b\|^2}_{\text{fidelity term}} + \lambda \underbrace{R(a)}_{\text{penalty term}}$$

fidelity term: $Fa \approx b$

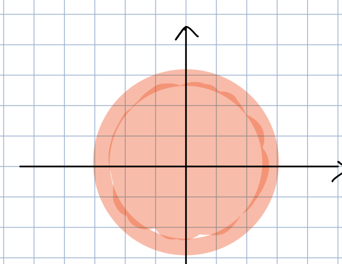
λ → regularization parameter, $\lambda = \lambda(\|\varepsilon\|)$

where $R : A \longrightarrow [0, +\infty]$ is the REGULARISER, which should promote a priori knowledge on a^+:



Examples

- $R(a) = \frac{1}{2}\|a\|^2$ Tikhonov reg.



exercise:

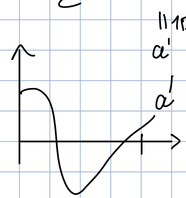
$$a_{rec} = (F^T F + \lambda I)^{-1} F^T b \qquad \left(\overset{\lambda=0}{=} (F^T F)^{-1} F^T b \right)$$

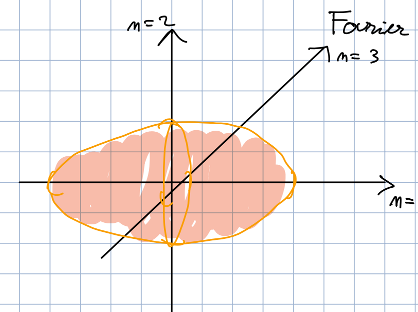
$$= \sum_j \frac{\sigma_j}{\sigma_j^2 + \lambda} \langle b, b_j \rangle a_j$$

• $R(a) = \frac{1}{2} \| \nabla a \|^2$ Sobolev reg.

Gen. Tik. reg.

smoothness promoted





$a' \longleftrightarrow n \cdot \hat{a}(n)$

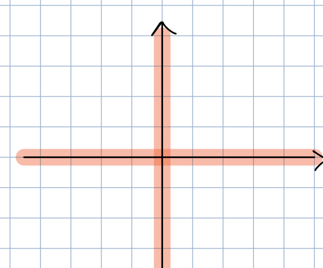
Regularization Theory:

$$\lambda = \lambda(\|\varepsilon\|) \implies \| a^+ - a_{rec}^{\lambda} \| \leq \sigma(\|\varepsilon\|)$$

• $R(a) = \| \Phi a \|_0$ where $\|x\|_0 = \# \{ i = 1..n : x_i \neq 0 \}$

$\Phi : \mathbb{R}^n \to \mathbb{R}^n$ "change of basis"

e.g. $\Phi = I$



$n = 2$

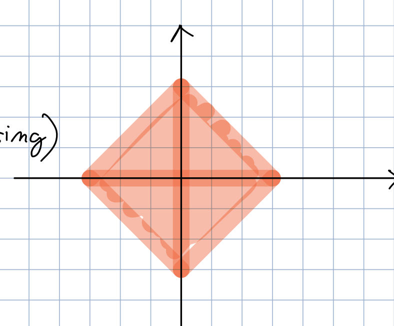
$\| \Phi a \|_0 \leq 1$

$\to$ 1-sparse vectors

$\| \Phi a \|_0$ is relaxed to

$R(a) = \| \Phi a \|_1$

(Basis pursuit - compressed sensing)



$$a_{rec} = \underset{a}{\operatorname{argmin}} \ \frac{1}{2} \| F a - b \|^2 + \lambda \| \Phi a \|_1$$

Proximal gradient descent : $\quad (\Phi^T \Phi = I)$

$$a_0 = 0$$

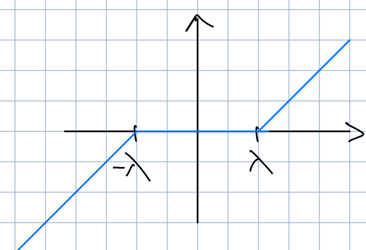
step size

$$a_{j+1} = S_{\Phi,\lambda}\left(a_j - \gamma F^T (F a_j - b)\right)$$

stop $\quad a_K$

acting componentwise

where $\quad S_{\Phi,\lambda} = \Phi^T S_\lambda \Phi \,, \quad S_\lambda$ is the soft-thresholding operator:



$S_{\Phi,\lambda}$ is the PROXIMAL OPERATOR:

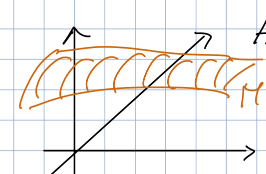
$$S_{\Phi,\lambda} = \mathrm{prox}_{\gamma\lambda\|\Phi\cdot\|_1} \,, \qquad \boxed{\mathrm{prox}_f(a) = \underset{x}{\mathrm{argmin}}\ \frac{1}{2}\|Ix - a\|^2 + f(x)}$$

denoising problem

- $R(\alpha) = \|\nabla \alpha\|_1 \qquad$ TOTAL VARIATION (TV)

  sparse gradient $\rightsquigarrow$ promotes piecewise-constant signals

- $R(\alpha) = \begin{cases} 0 & \alpha \in M \subseteq A \\ +\infty & \alpha \notin M \end{cases}$



"manifold hypothesis"