IPAM Graduate Summer School: Computer Vision, July 2013

Inverse modelling using optimization to solve imaging tasks

Mila Nikolova

ENS Cachan, CNRS, France

nikolova@cmla.ens-cachan.fr

http://mnikolova.perso.math.cnrs.fr

Computer vision is a young field that arose with the "digital revolution"

Natural images

Finite number of cells in the primary visual cortex

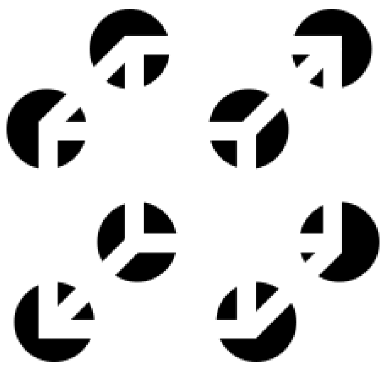
Each cell recognizes a specific geometric shape or color data (D. Hubel, T. Wiesel)

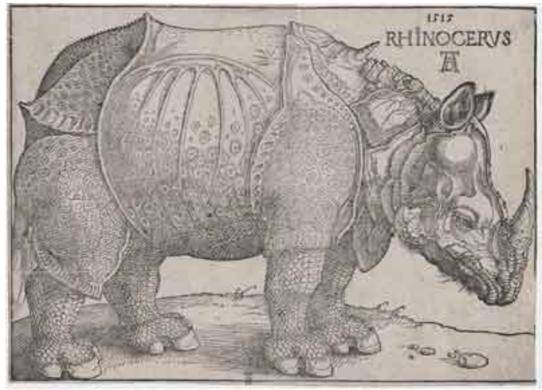
The whole image is produced in another part of our brain

> *"... our perceptions or ideas arise from an active critical principle."* *J.-J. Rousseau*

Gestalt theory of visual perception (since M. Wertheimer, 1923)

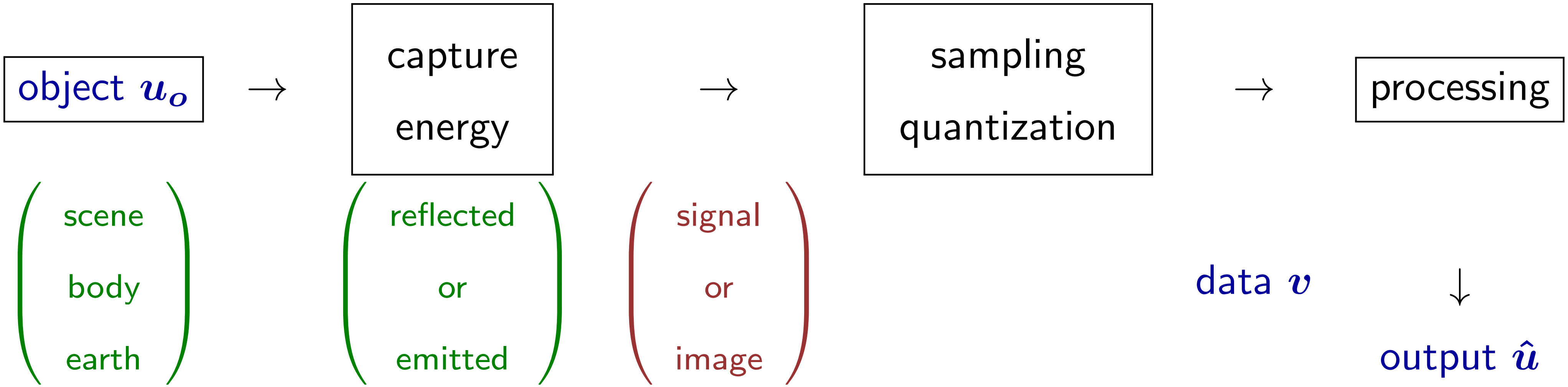
Is there a cube?





The output of imaging devices must satisfy perception (simplifications are enabled)

Objective criteria for image quality is a still open question

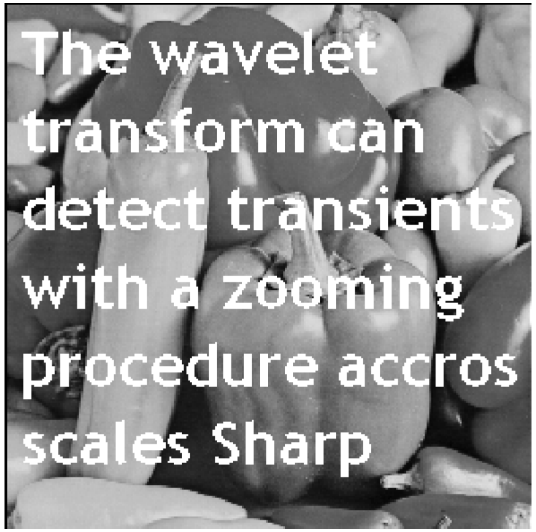


Mathematical model: $v = \text{Transform}(u_o) \bullet (\text{Perturbations})$

Some transforms: loss of pixels, blur, FT, Radon T., frame T. $(\cdots)$

Processing tasks: $\begin{cases} \hat{u} = \text{recover}(u_o) \\ \hat{u} = \text{objects of interest}(u_o) \end{cases} \quad (\cdots)$

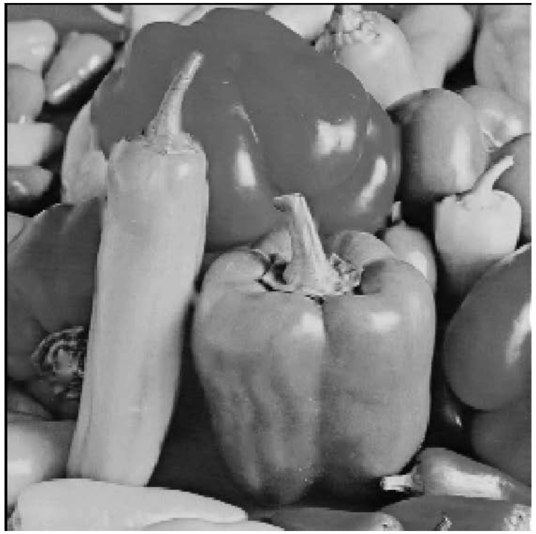
Mathematical tools: PDEs, Statistics, Functional anal., Matrix anal., $(\cdots)$



Editing ↓

Inpainting ↓

Denoising ↓



[Pérez, Gangnet, Blake 04]

[Chan, Steidl, Setzer 08]

[M. Lebrun, A. Buades and J.-M. Morel, 2112]

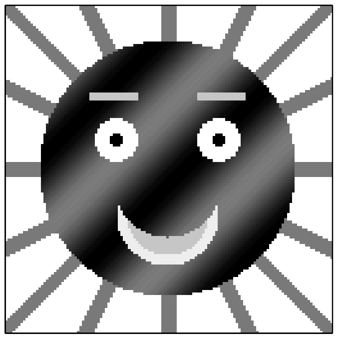
Image/signal processing tasks often require to solve **ill-posed inverse problems**

Out-of-focus picture: $v = a * u_o + \text{noise} = Au_o + \text{noise}$

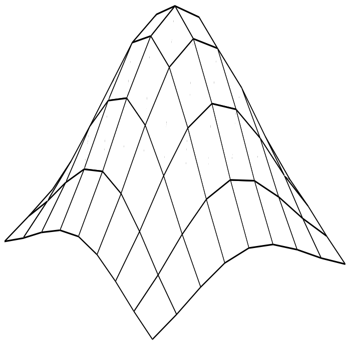
A is ill-conditioned $\equiv$ (nearly) noninvertible

Least-squares solution: $\hat{u} = \arg\min_{u} \left\{ \|Au - v\|^2 \right\}$

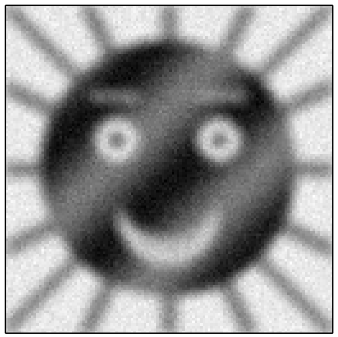
Tikhonov regularization: $\hat{u} \doteq \arg\min_{u} \left\{ \|Au - v\|^2 + \beta \sum_i \|G_i u\|^2 \right\}$ for $\{G_i\} \approx \nabla$, $\beta > 0$



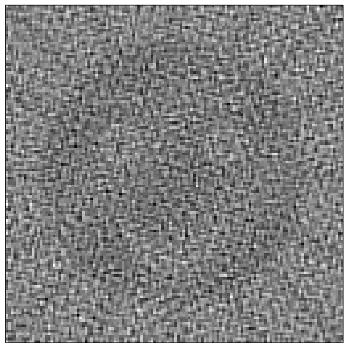
Original u_o



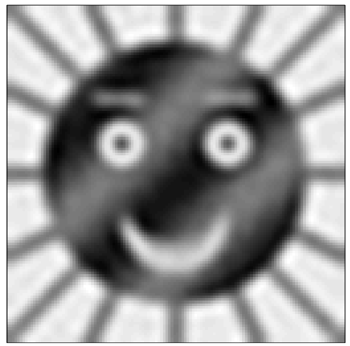
Blur a



Data v



$\hat{u}$: Least-squares



$\hat{u}$: Tikhonov

1. Energy minimization methods

u_o (unknown) $\qquad v$ (data) $=$ Transform(u_o) $\bullet$ (Perturbations)

solution $\hat{u}$ $\nearrow$ coherent with data production model $\Psi(u,v)$ (data-fidelity)

$\searrow$ coherent with priors and desiderata $\Phi(u)$ (prior)

Combining models:

$$\hat{u} = \arg\min_{u\in\Omega} \mathcal{F}_v(u) \qquad (\mathcal{P})$$

$$\mathcal{F}_v(u) = \Psi(u,v) + \beta\Phi(u), \qquad \beta \geqslant 0$$

How to choose $(\mathcal{P})$ to get a good $\hat{u}$?

Applications: Denoising, Segmentation, Deblurring, Tomography, Seismic imaging, Zoom, Superresolution, Learning, Motion estimation, Pattern recognition $(\cdots)$

The $m \times n$ image u is stored in a $p = mn$-length vector, $u \in \mathbb{R}^p$, data $v \in \mathbb{R}^q$

$\boldsymbol{\Psi}$ usually models the production of data $\boldsymbol{v}$ $\Rightarrow$ $\Psi = -\log($ Likelihood $(\boldsymbol{v}|\boldsymbol{u})$ $)$

$v = Au_o + n$ for n white Gaussian noise $\Rightarrow$ $\boxed{\boldsymbol{\Psi(u,v) \propto \|Au - v\|_2^2}}$

$\boldsymbol{\Phi}$ model for the unknown $\boldsymbol{u}$ (statistics, smoothness, edges, textures, expected features)

- Bayesian approach
- Variational approach

Both approaches lead to similar energies

Prior via regularization term $\boldsymbol{\Phi(u) = \sum_i \varphi(\|G_i u\|)}$

$\boldsymbol{\varphi} : \mathbb{R}_+ \to \mathbb{R}_+$ potential function (PF)

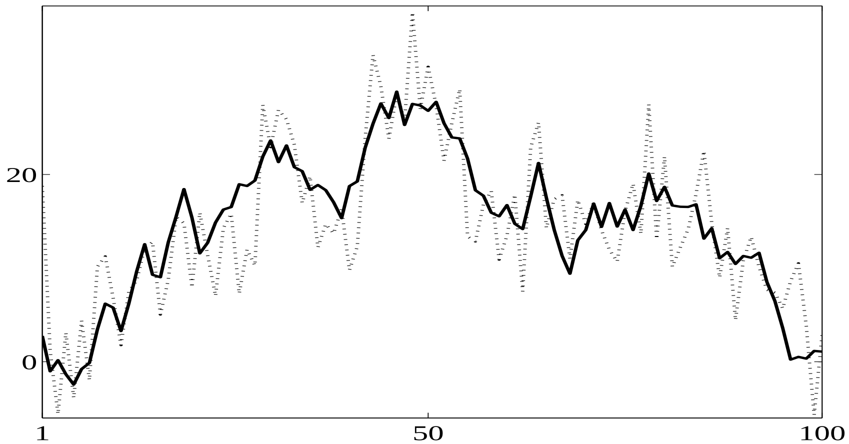
$\boldsymbol{\{G_i\}}$ — linear operators

Bayes: U, V random variables, Likelihood $f_{V|U}(v|u)$, Prior $f_U(u) \propto \exp\{-\lambda\Phi(u)\}$

Maximum a Posteriori (MAP) yields the most likely solution $\hat{\boldsymbol{u}}$ given the data $\boldsymbol{V} = \boldsymbol{v}$:

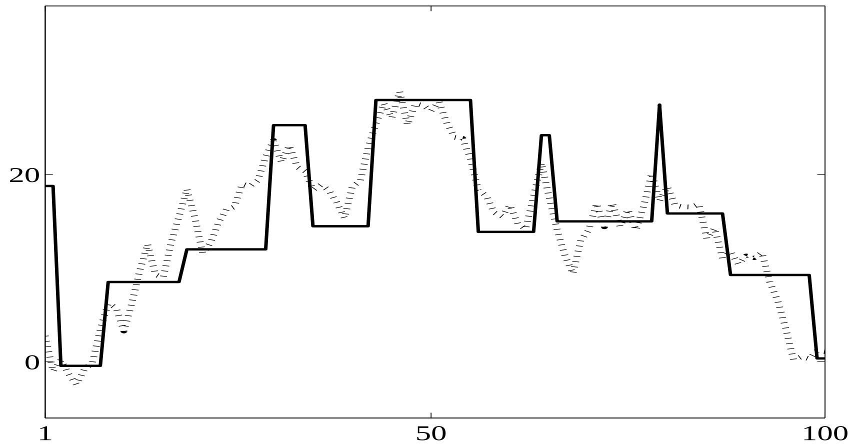
$$\hat{u} = \arg\max_u f_{U|V}(u|v) \quad = \quad \arg\min_u \left(-\ln f_{V|U}(v|u) - \ln f_U(u) \right)$$
$$= \quad \arg\min_u \left(\Psi(u, v) \; + \; \beta\Phi(u) \right) \; = \; \arg\min_u \mathcal{F}_v(u)$$

MAP is a very usual way to combine models on data-acquisition and priors



Original $u_o \sim f_U$ (—)

Data $v = u_o$+noise ($\cdots$), noise $\sim f_{V|U}$



The true MAP $\hat{u}$ (—)

The original u_o ($\cdots$)

- **Minimizer approach** (the core of our tutorials)

  - Analyze the main properties exhibited by the (local) minimizers $\hat{u}$ of $\mathcal{F}_v$ as a function of the shape of $\mathcal{F}_v$

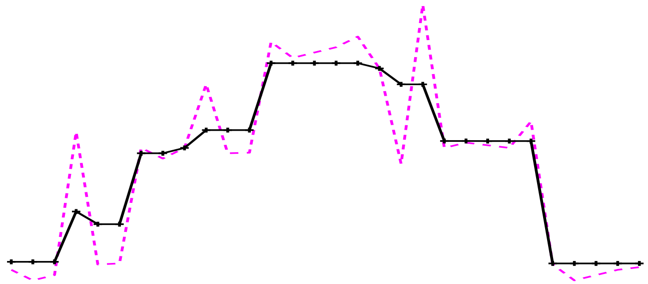
    Strong results...

    Rigorous tools for modelling

  - Conceive $\mathcal{F}_v$ so that $\hat{u}$ satisfy your requirements.

"There is nothing quite as practical as a good theory." *Kurt Lewin*

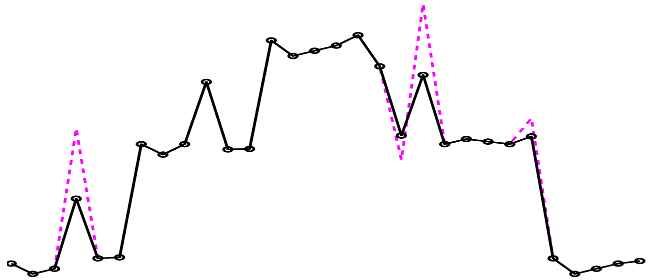
Illustration: the role of the smoothness of $\mathcal{F}_v$



STAIR-CASING

$$\mathcal{F}_v(u) = \sum_{i=1}^{p}(u_i - v_i)^2 + \beta \sum_{i=1}^{p-1} |u_i - u_{i+1}|$$

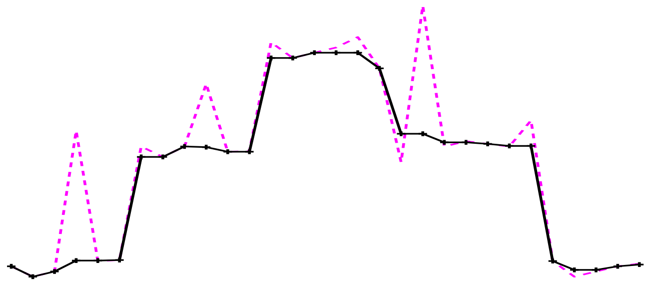
smooth **non-smooth**



EXACT DATA-FIT

$$\mathcal{F}_v(u) = \sum_{i=1}^{p} |u_i - v_i| + \beta \sum_{i=1}^{p-1} (u_i - u_{i+1})^2$$

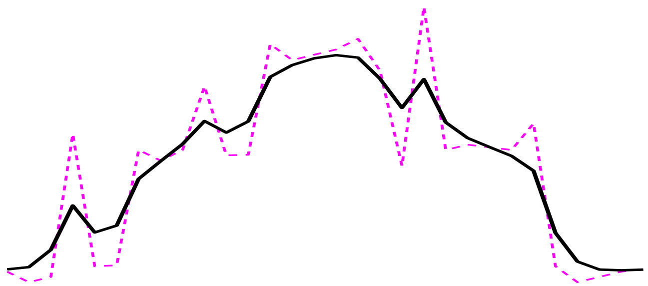
non-smooth **smooth**



BOTH EFFECTS

$$\mathcal{F}_v(u) = \sum_{i=1}^{p} |u_i - v_i| + \beta \sum_{i=1}^{p-1} |u_i - u_{i+1}|$$

non-smooth **non-smooth**



$$\mathcal{F}_v(u) = \sum_{i=1}^{p} (u_i - v_i)^2 + \beta \sum_{i=1}^{p-1} (u_i - u_{i+1})^2$$

smooth **smooth**

Data (— — —), Minimizer (—)

We shall explain why and how to use

Some energy functions

Regularization [Tikhonov, Arsenin 77]: $\boldsymbol{\mathcal{F}_v(u) = \|Au - v\|^2 + \beta\|Gu\|^2}$, $G = I$ or $G \approx \nabla$

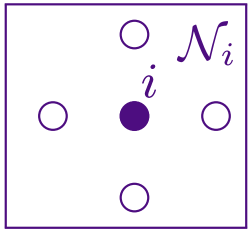
Focus on edges and contours

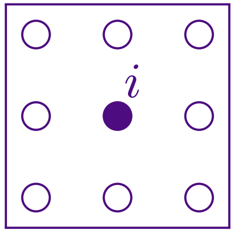
Statistical framework

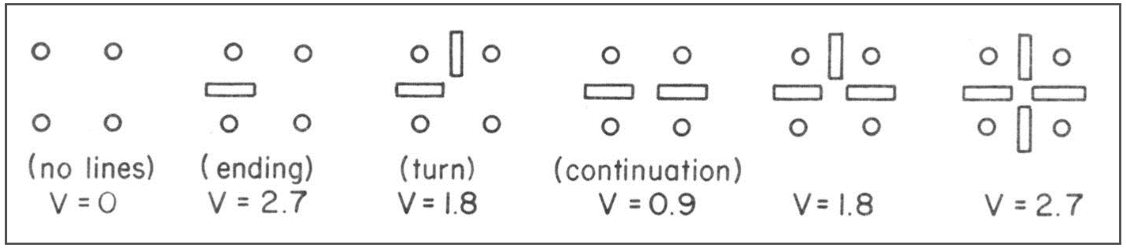
Line process in Markov random field priors [Geman, Geman 84]: $(\hat{u}, \hat{\ell}) = \arg\min_{u,\ell} \mathcal{F}_v(u, \ell)$

$$\boldsymbol{\mathcal{F}_v(u,\ell) = \Psi(u,v) + \beta \sum_i \Big(\sum_{j \in \mathcal{N}_i} \varphi(u[i] - u[j])(1 - \ell_{i,j}) + \sum_{(k,n) \in \mathcal{N}_{i,j}} \mathbf{V}(\ell_{i,j}, \ell_{k,n}) \Big)}$$

$[\ell_{i,j} = 0 \;\Leftrightarrow\;$ no edge$]$, $[\ell_{i,j} = 1 \;\Leftrightarrow\;$ edge between i and $j]$, $\varphi(t) = 1$







Computation: stochastic relaxation and annealing (global convergence with high probability)

PDE's framework

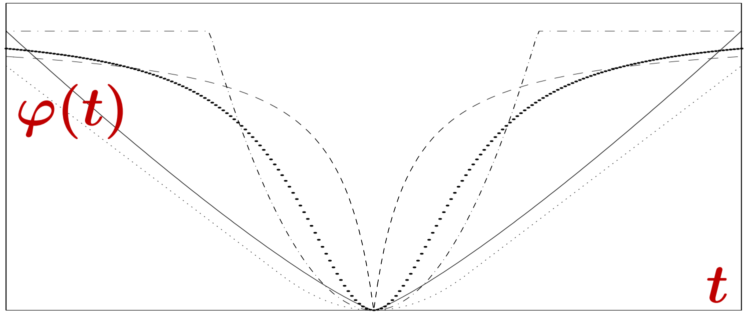
$\boldsymbol{\Phi}(\mathbf{u})$

M.-S. functional [Mumford, Shah 89]: $\mathcal{F}_v(u,L)=\int_\Omega (u-v)^2 dx+\beta\left(\int_{\Omega\setminus L}\|\nabla u\|^2 dx+\alpha\,|L|\right)$

discrete version: $\Phi(u)=\sum_i \varphi(\|G_i u\|)$, $\quad \varphi(t)=\min\{t^2,\alpha\}$, $\quad \{G_i\}\approx\nabla$

Total Variation (TV) [Rudin, Osher, Fatemi 92]: $\mathcal{F}_v(u)=\|u-v\|_2^2+\beta\,\mathrm{TV}(u)$

$$\mathrm{TV}(u)=\int\|\nabla u\|_2\,dx\approx\sum_i\|G_i u\|_2$$



Various edge-preserving functions φ to define Φ

φ is edge-preserving if $\lim_{t\to\infty}\frac{\varphi'(t)}{t}=0$

[Charbonnier, Blanc-Féraud, Aubert, Barlaud 97 ...]

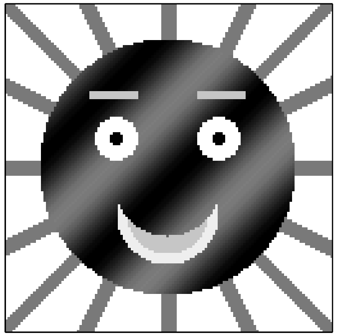
Minimizer approach

ℓ_1- Data fidelity [Nikolova 02]: $\mathcal{F}_v(u)=\|Au-v\|_1+\beta\Phi(u)$

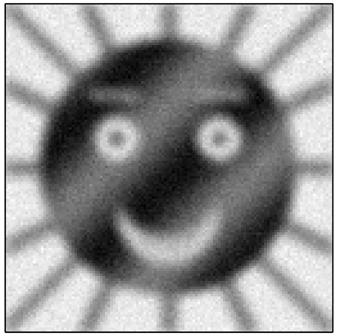
$L_1-\mathrm{TV}$ model [T. Chan, Esedoglu 05]: $\mathcal{F}_v(u)=\|u-v\|_1+\beta\,\mathrm{TV}(u)$

CPU time ! Computers ↑↑

Original u_o

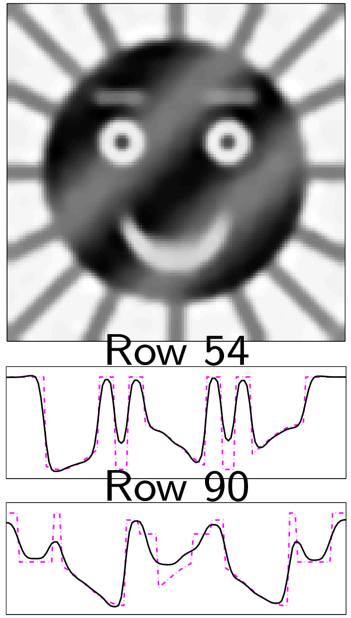


Data $v = a * u_o + n$

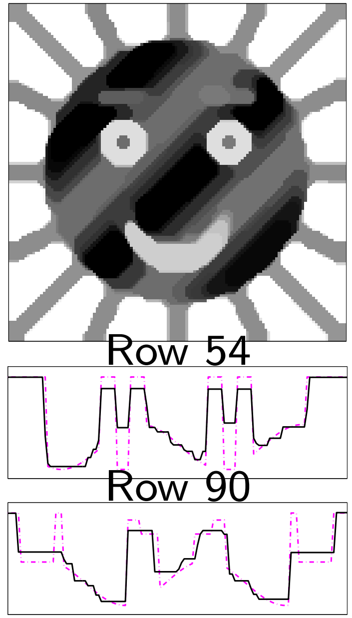


$$\mathcal{F}_v(u) = \|Au - v\|^2 + \beta \sum \varphi(G_i u)$$

$\varphi(t) = |t|^{\alpha \in (1,2)}$



$\varphi(t) = |t|$

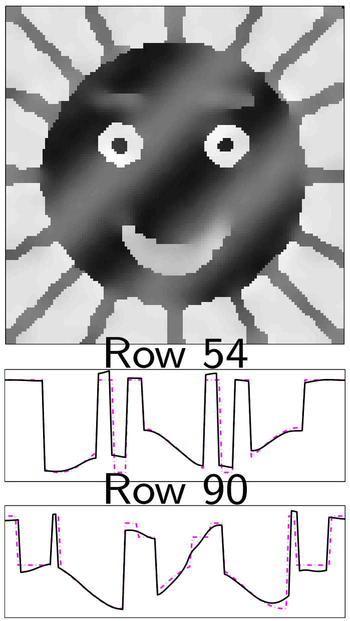


φ convex

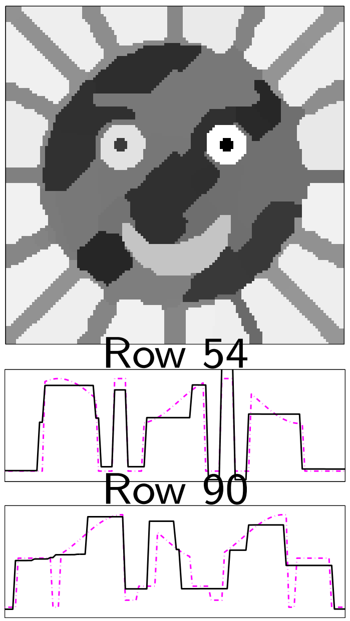
φ smooth at 0

φ nonsmooth at 0

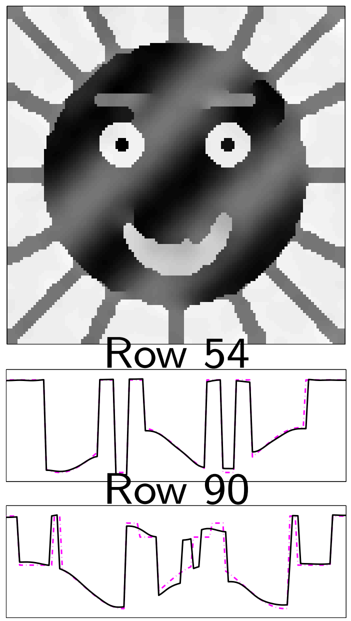
$\varphi(t) = \alpha t^2/(1 + \alpha t^2)$



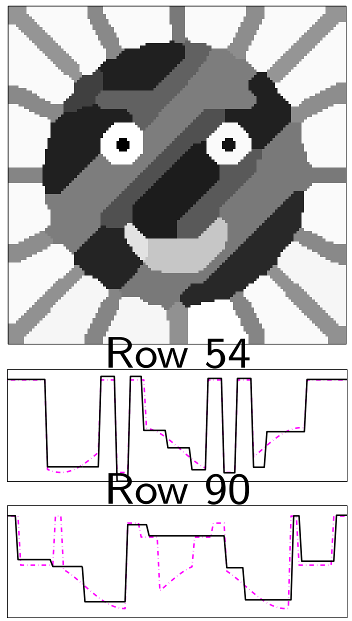
$\varphi(t) = \alpha|t|/(1 + \alpha|t|)$



$\varphi(t) = \min\{\alpha t^2, 1\}$

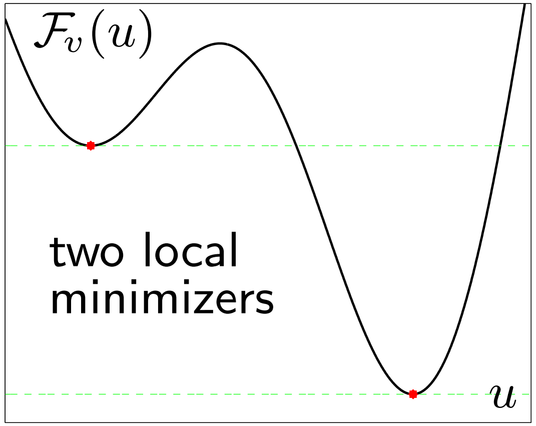


$\varphi(t) = 1 - 1\!\!1_{(t=0)}$

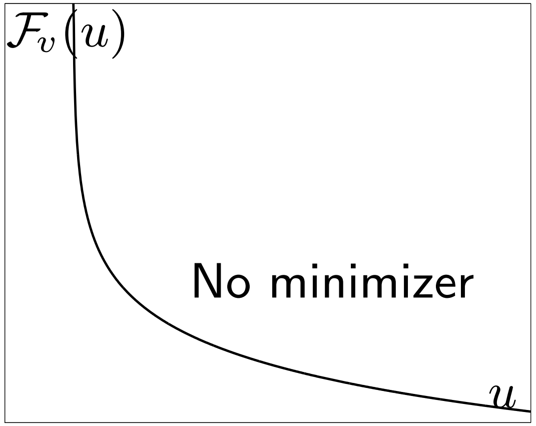


nonconvex

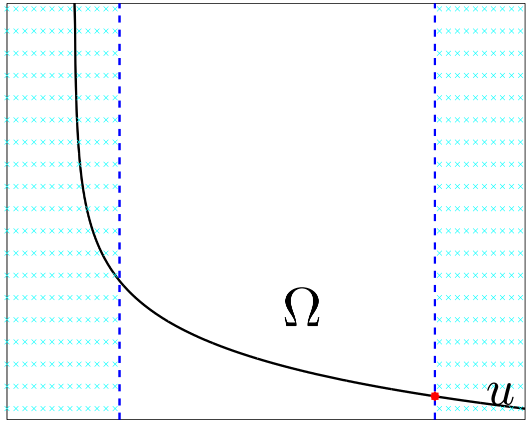
2 Regularity of the optimization problems



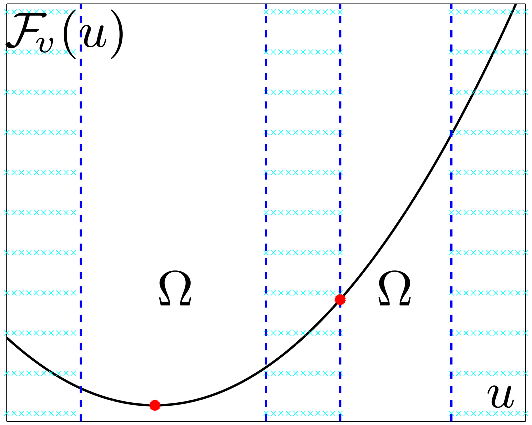
$\mathcal{F}_v$ nonconvex



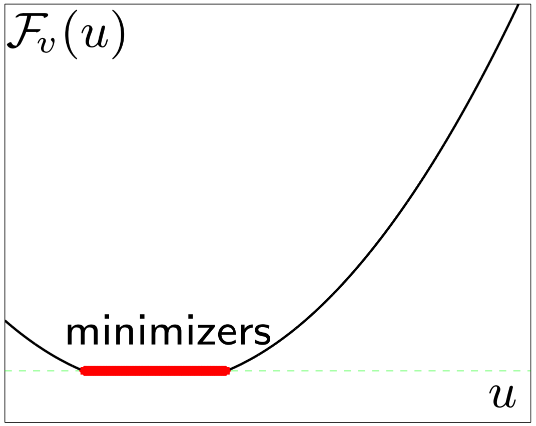
$\mathcal{F}_v$ convex non coercive $\Omega = \mathbb{R}$



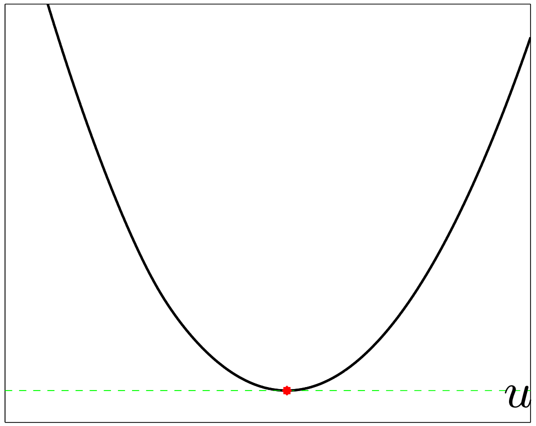
$\mathcal{F}_v$ convex non coercive Ω compact



$\mathcal{F}_v$ strictly convex, Ω nonconvex



$\mathcal{F}_v$ non strictly convex



$\mathcal{F}_v$ strictly convex on $\mathbb{R}$

$$\mathcal{F}_v : \Omega \to \mathbb{R} \qquad \Omega \subset \mathbb{R}^p$$

- Optimal set $\widehat{U} = \{\hat{u} \in \Omega \ : \ \mathcal{F}_v(\hat{u}) \leqslant \mathcal{F}_v(u) \quad \forall\, u \in \Omega\}$

  $\widehat{U} = \{u\}$ if $\mathcal{F}_v$ strictly convex

  $\widehat{U} \neq \varnothing$ if $\mathcal{F}_v$ coercive of if $\mathcal{F}_v$ continuous and Ω compact

  Otherwise – check

  (e.g. see if $\mathcal{F}_v$ is asymptotically level stable [Auslender, Teboulle 03])

- Nonconvex problems:

  Algorithms may get trapped in local minima

  A “good” local minimizer can be satisfying

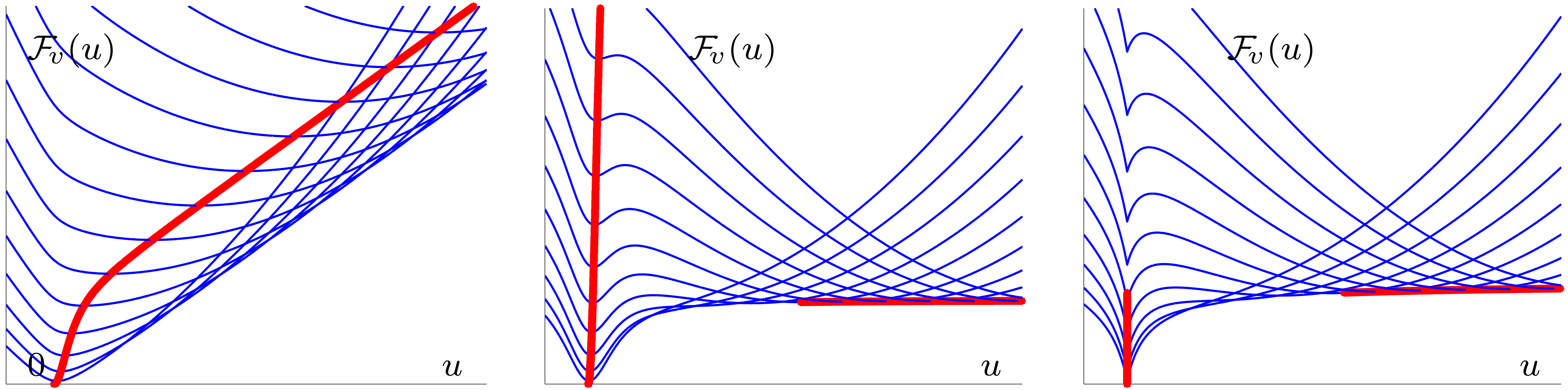
  Global optimization – difficult, but progress, e.g. [Robini Reissman 13]

- Attention to numerical errors

Definition: $\mathcal{U} : O \to \mathbb{R}^p$, $O \subset \mathbb{R}^q$ open, is a (local) minimizer function for

$$\mathcal{F}_O \doteq \{\mathcal{F}_v \;:\; v \in O\} \text{ if } \mathcal{F}_v \text{ has a strict (local) minimum at } \mathcal{U}(v), \; \forall\, v \in O$$

Minimizer functions – a good tool to analyze the properties of minimizers...



$\mathcal{F}_v(u) = (u-v)^2 + \beta\sqrt{\alpha + u^2}$ — minimizer function (red)

$\mathcal{F}_v(u) = (u-v)^2 + \beta\frac{\alpha u^2}{1+\alpha u^2}$ — local minimizer functions (red)

$\mathcal{F}_v(u) = (u-v)^2 + \beta\frac{\alpha |u|}{1+\alpha |u|}$ — global minimizer function (red)

Each blue curve curve: $u \to \mathcal{F}_v(u)$ for $v \in \{0, 2, \cdots\}$

Question 1 What these plots reveal about the local / global minimizer functions?