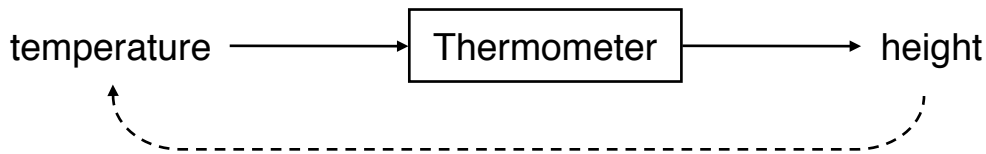
Eric Debreuve - debreuve@i3s.unice.fr

Definition, Ill-posedness, Regularization

- **Purpose**
  - To determine the magnitude of a (set of) characteristic(s) from some measurements

- **Ex. temperature**
  - Thermometer
    - Measurement: height of alcohol or mercury in the tube
    - Conversion of this height into a temperature



    - Forward problem: Physical phenomenon: height = F(temperature)
    - Inverse problem: Find G such that the temperature = G(height)
      - Ideally, $G = F^{-1}$

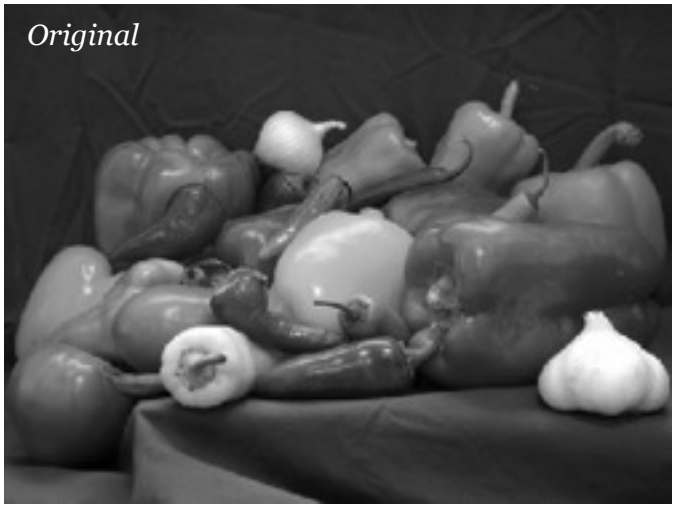
- **Difficulties**
  - Quality of the instrument
  - Complexity of the phenomenon
    - Is it linear? F(t) = a t

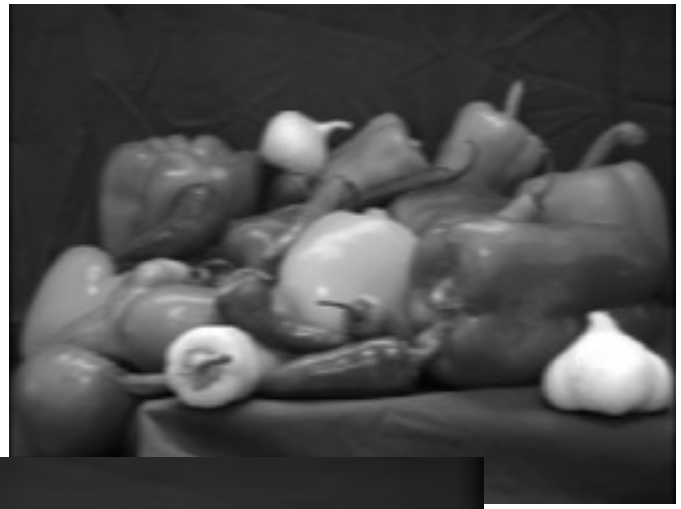
  - Ability to find G or a procedure allowing to approximate it?

- **Context**
  - Shooting: fixed camera
  - Scene: fixed and moving objects
  - Physical phenomena: light measurement over a time interval
  - Consequence: motion blur

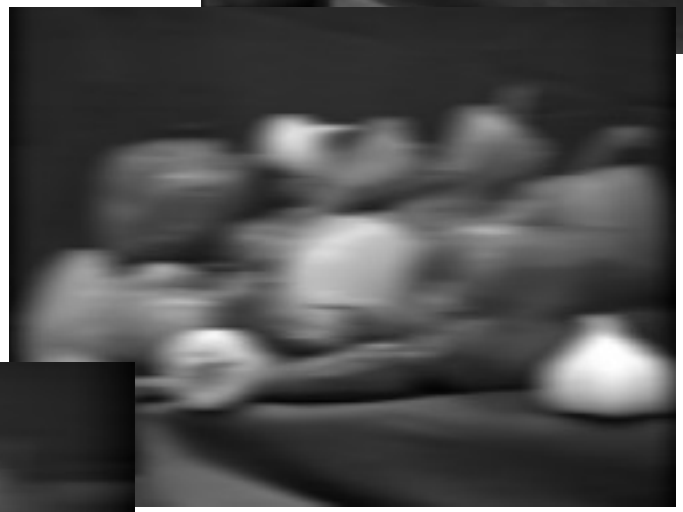
- **Goal**
  - Given the blurry observation y, find a deblurred image x

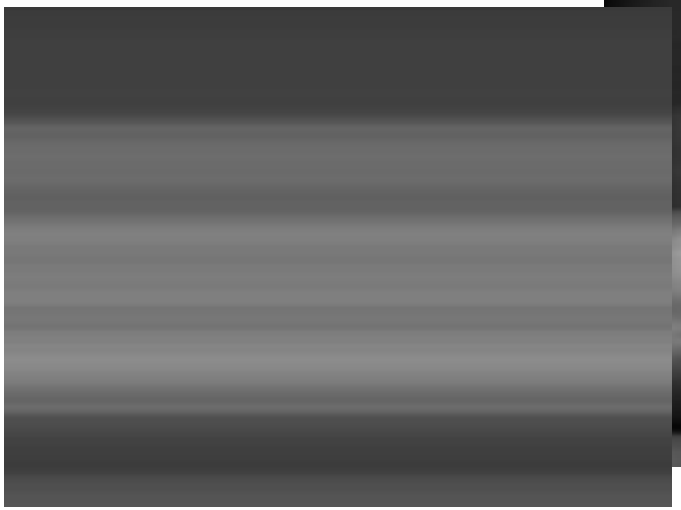
- **Method**
  - Model the physical phenomenon ➜ simplified representation: m
  - Express the solution as x = f(m,y) or build an algorithm (m,y) ➜ x





Several levels of (horizontal) blur



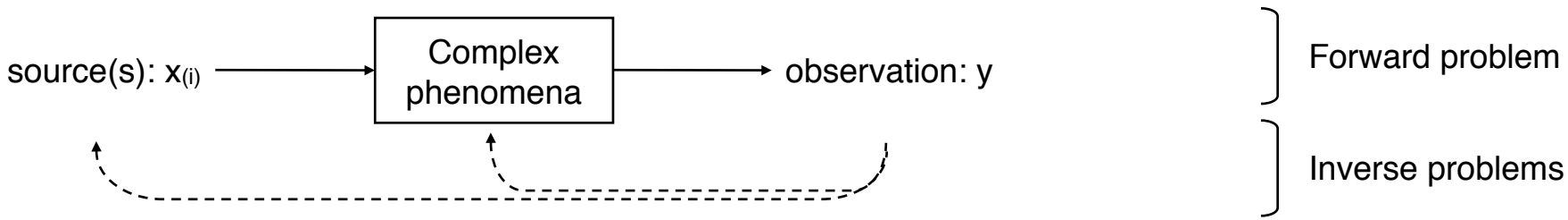


Lecture: Definition, ill-posedness, regularization

- **Overview**
  - Terminology
  - Types of inverse problems
- **Some reminders**
  - Convolution, Fourier
- **Forward problem**
  - Model
  - Deterministic and stochastic points of view
- **Ill-posed problem**
  - Existence
  - Uniqueness
  - Stability
- **Regularization**
  - Constrained approach
  - Lagrangian approach
    - Equivalence
  - Derivation of a smoothness constraint
  - L1-regularization

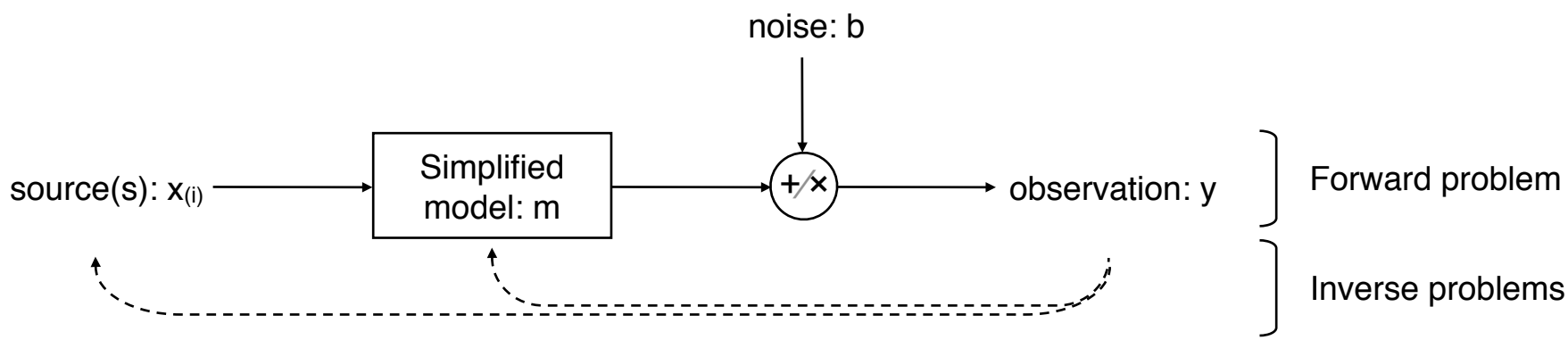
- **Terminology**
  - What we are looking for: Unknown(s) / Source(s) / Original image
  - What we have at our disposal: Observation(s) / Data / Acquisition / Observed image / Degraded image

- **Real problem: often insoluble**



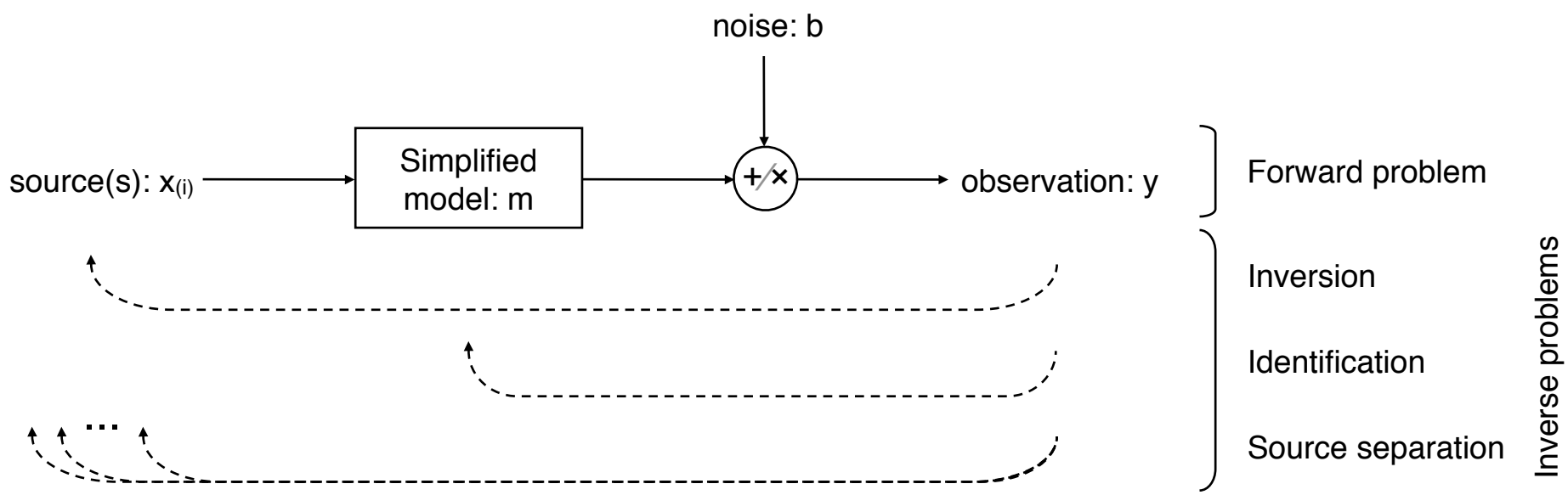
  - What is the precision on the observation?
  - How well are the phenomena known?

- **Simplified problem**

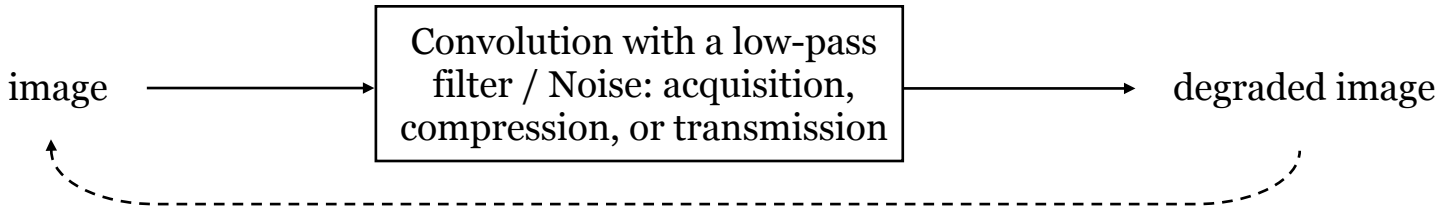


- **"Observable" unknown**
  - Although after degradation
    - Imperfect/disturbed transmission channel
    - Lossy compression
    - Measurement noise
- **Unobservable unknown**
  - Information « hidden » within an unreachable medium
    1. Use of an external source
    2. Observation of the response
    3. Inference of the unknown information

- **Three types of problems:** $y = m(x) \# b$ or $y = m(x_1, x_2 \ldots x_n) \# b$ (operator #: "+" or "x")
  - Inversion: m is known (modeled), y is observed ➜ find x
    - y is observed ➜ find x and m (blind deconvolution)
  - Identification: x is known, m is partly unknown, y is observed ➜ find m
  - Source separation: m is partly unknown, y is observed ➜ find m and $x_1, x_2 \ldots x_n$



- **Image restoration**
  - Denoising
  - Deconvolution

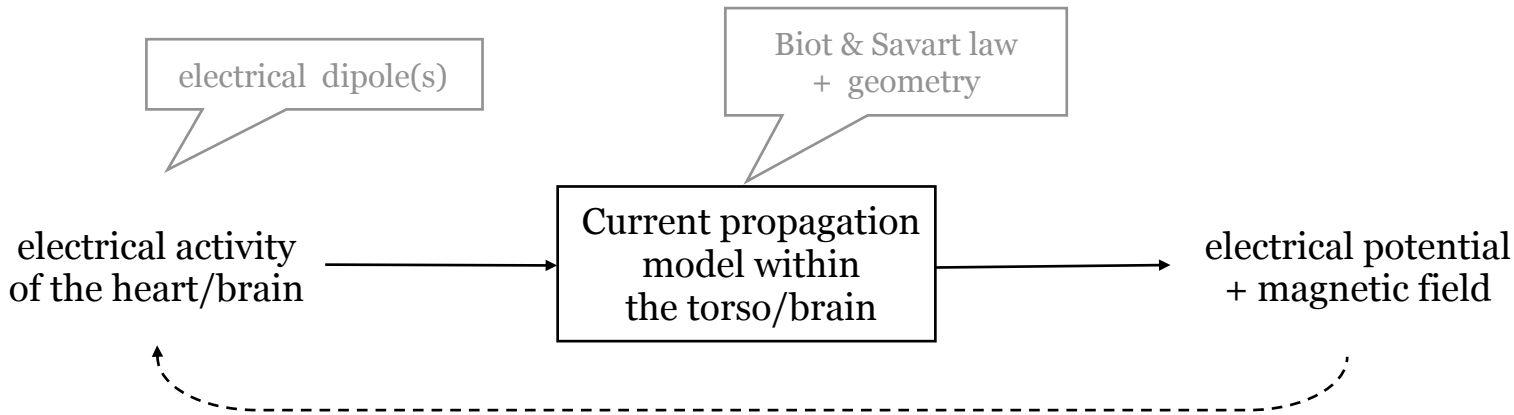


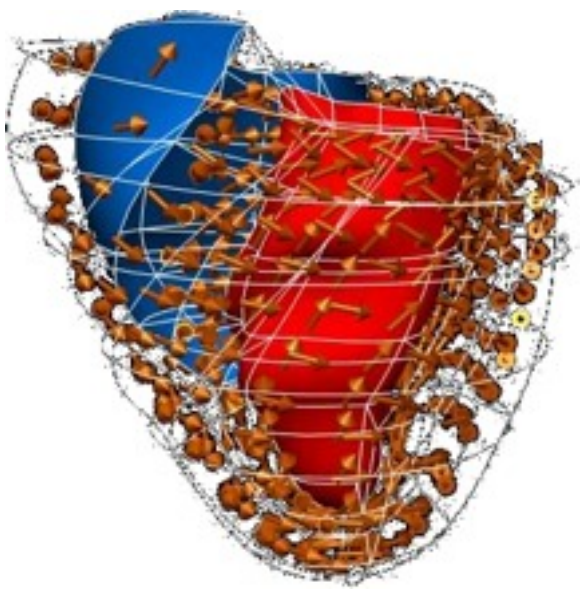
Original image

Degraded image: low-pass filter + noise

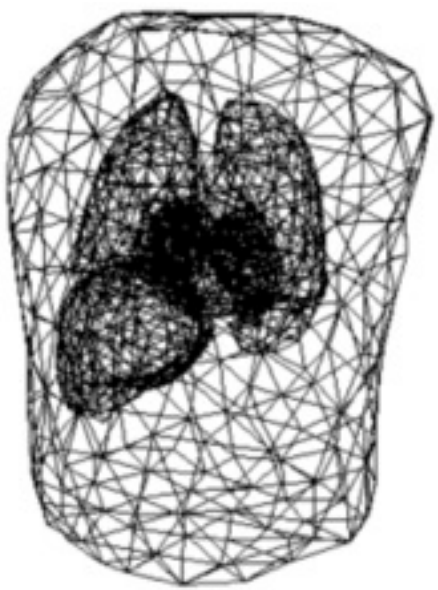
Restored image

- **Myocardium: ElectroCardioGraphy (ECG or EKG) + MagnetoCardioGraphy (MCG)**
- **Brain: ElectroEncephaloGraphy (EEG) + MagnetoEncephaloGraphy (MEG)**





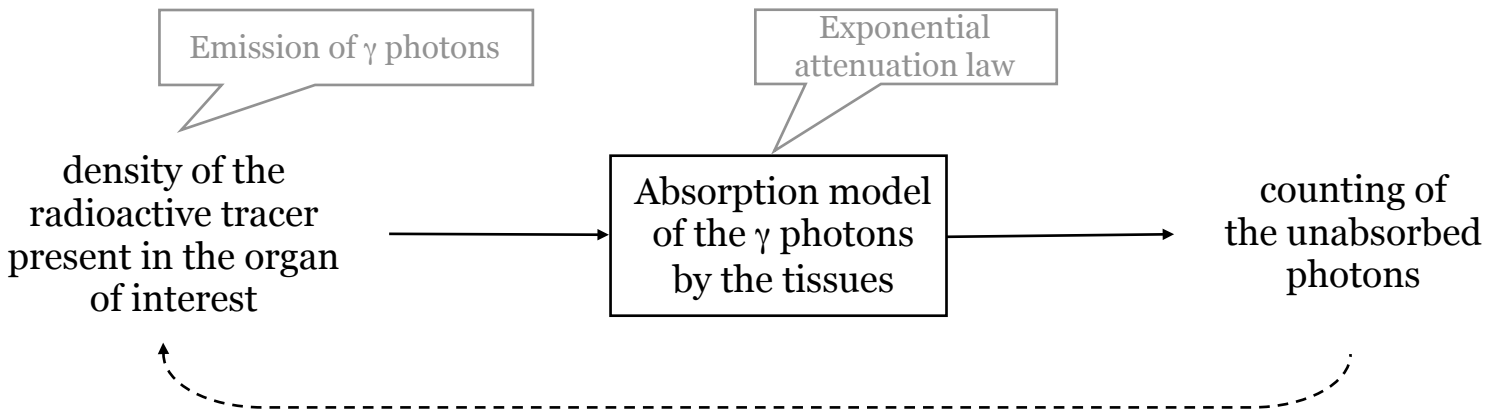
Bioengineering Research Group,
The University of Auckland, New Zealand

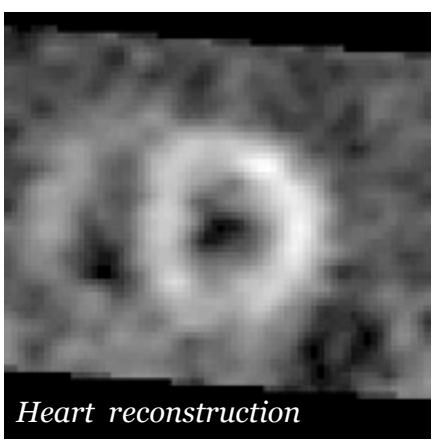


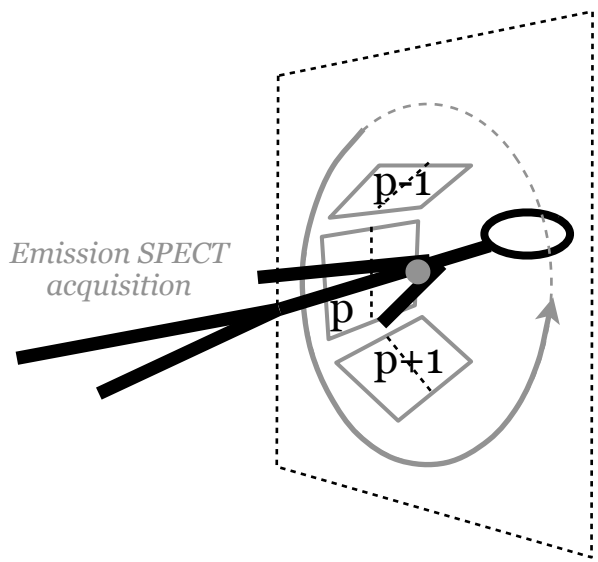
Conductivity model: set of regions
with constant conductivity

Observed electrical potential

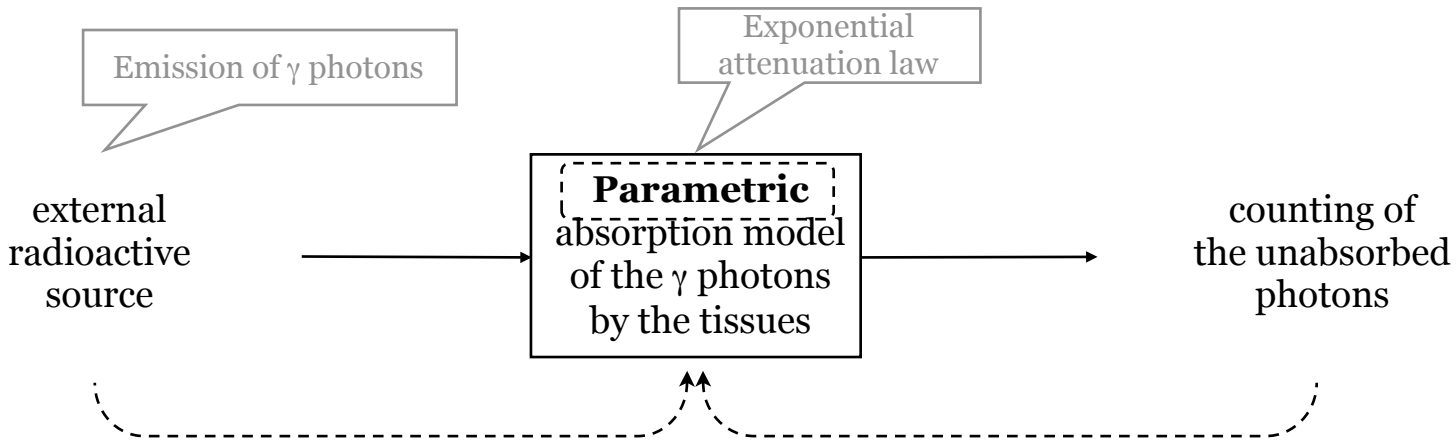
- **Nuclear medicine imaging**
  - Emission Single-photon emission computed tomography (SPECT)
  - Positron emission tomography (PET)

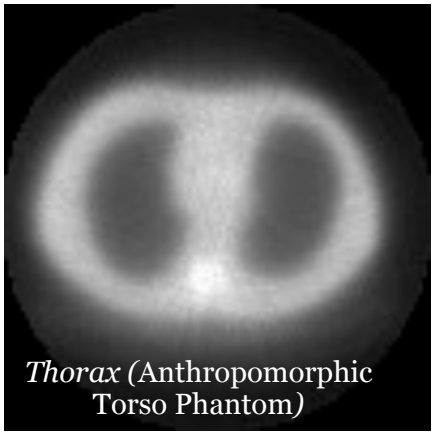


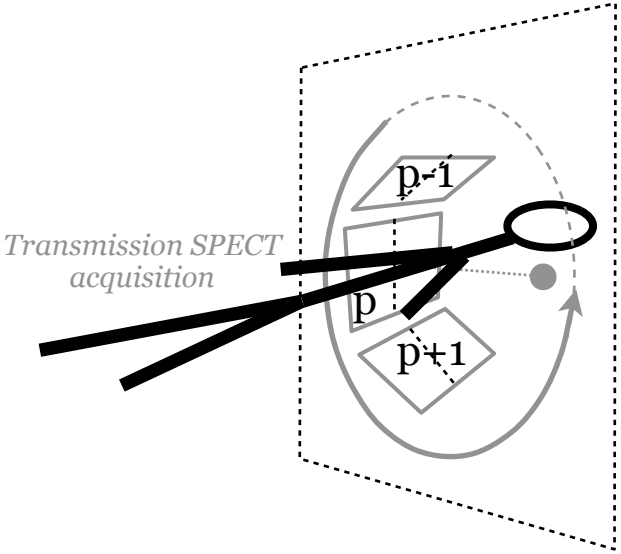




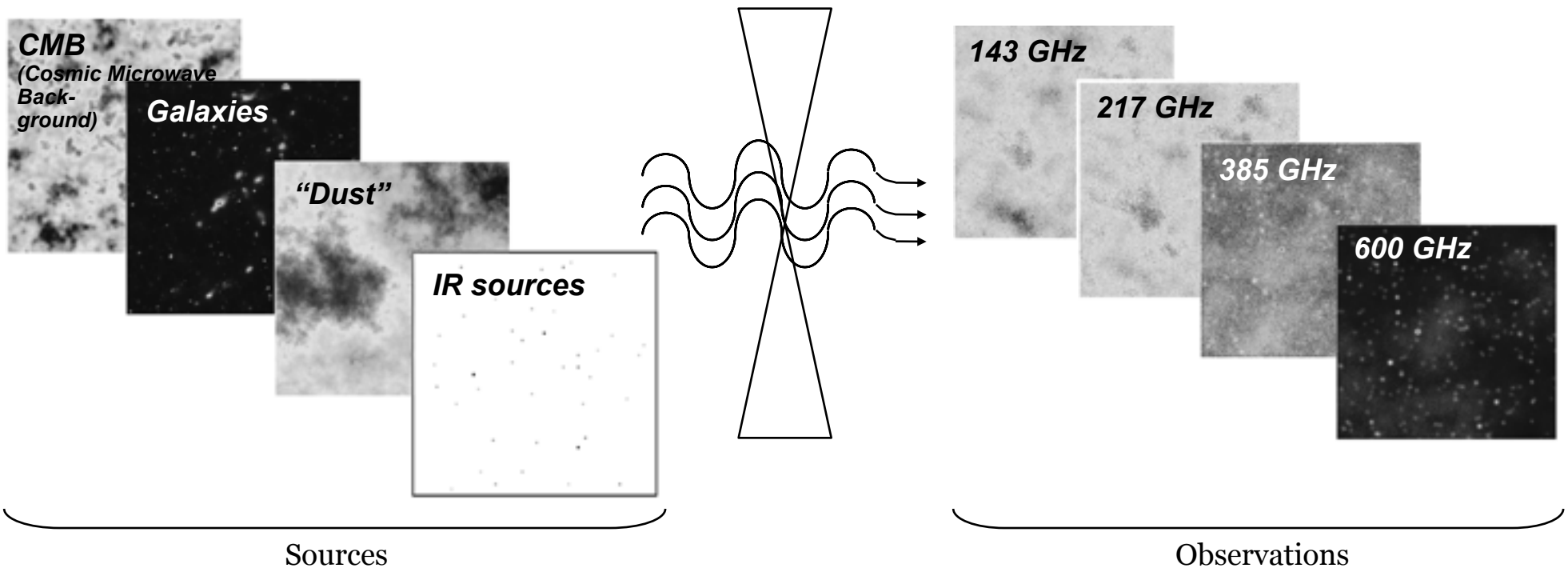
- **Nuclear medicine imaging: Transmission SPECT**
- **Scanner X**







- **Astronomy**
  - Sources: different physical phenomena
  - Observations
    - Several observations measured at distinct wavelengths
    - Convolutive mixture of the sources
  - Linear model
    - Source-dependent mixture
    - Wavelength-dependent convolution filters

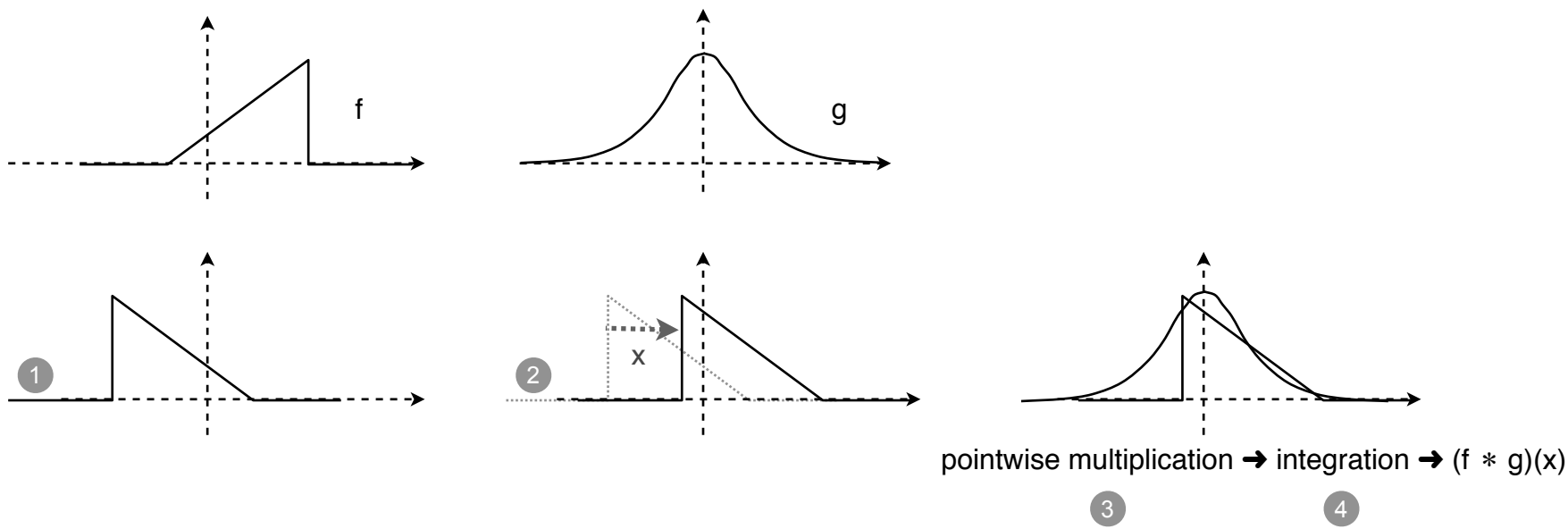


Lecture: Definition, ill-posedness, regularization

- **Overview**
  - Terminology
  - Types of inverse problems
- **Some reminders**
  - Convolution, Fourier
- **Forward problem**
  - Model
  - Deterministic and stochastic points of view
- **Ill-posed problem**
  - Existence
  - Uniqueness
  - Stability
- **Regularization**
  - Constrained approach
  - Lagrangian approach
    - Equivalence
  - Derivation of a smoothness constraint
  - L1-regularization

- **Definition (1-D)**
  - f and g are two real or complex functions
  - Continuous $(f * g)(x) = \int_{-\infty}^{+\infty} f(x-t)\ g(t)\ dt$
  - Discrete $(f * g)(n) = \sum_{m=-\infty}^{+\infty} f[n-m]\ g[m]$

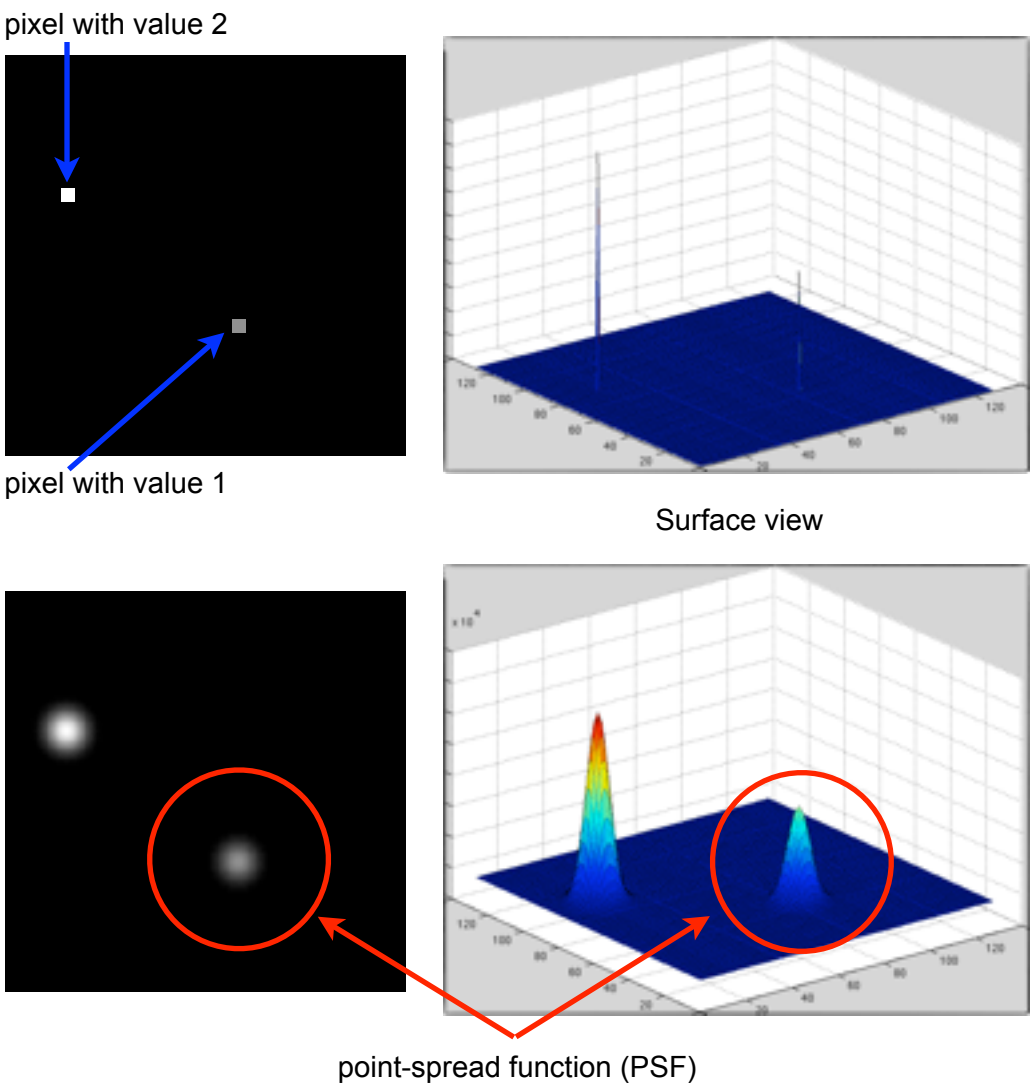
- **Interpretation**



Reminder > Convolution (or convolution product)

- **Discrete (1-D):** $(f * g)(n) = \sum_{m=-\infty}^{+\infty} f[n-m]\, g[m]$

- **Blurring ➜ convolution (discrete, 1-D)**
  - Signal with a single bright point
    - Sequence x[n] = δ[n]: all zeros expect δ[0] = 1
  - After blurring: y[n]
    - Bright point has spread out into a region
    - ▸ The point-spread function (PSF): h[n]
  - Hypothesis
    - Blurring: linear and spatially invariant
  - Signal = set of shifted bright points with ≠ amplitudes
    - $x[n] = \sum_m \alpha_m \delta[n-m]$
  - After blurring
    - y[n] = same linear combination of shifted PSFs
      $= \sum_m \alpha_m h[n-m]$
    - α_k ? $\alpha_k = x[k]$ ➜ $y[n] = \sum_m x[m]\, h[n-m]$



- **Properties**
  - Commutativity / symmetry: $f * g = g * f$
  - Associativity: $f * (g * h) = (f * g) * h$
  - Linearity: $f * (\alpha g + h) = \alpha (f * g) + (f * h)$
  - Identity or neutral element: $f * \delta = f$
    - δ is the Dirac distribution (incorrectly called the Dirac *function*) ≈ intuitively, *a function* with infinite value at 0 and equal to zero elsewhere whose integral is equal to 1
  - Differentiation: $\frac{d}{dx}(f * g) = \frac{df}{dx} * g = f * \frac{dg}{dx}$
  - Convolution theorem:
    - $\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \times \mathcal{F}\{g\}$
    - $\mathcal{F}\{f\ g\} = \mathcal{F}\{f\} * \mathcal{F}\{g\}$
    - Therefore: $f * g = \mathcal{F}^{-1}\left\{\mathcal{F}\{f\} \times \mathcal{F}\{g\}\right\}$
    - where $\mathcal{F}$ is the Fourier transform

- **Matrix expression (1-D)**
  - Discrete convolution: $y(t) = (M * x)(t)$

$$= \sum_{u=-\infty}^{+\infty} M(u)\, x(t-u)$$

  - Filter with finite support ➜ $$= \sum_{u=-k}^{+k} M(u)\, x(t-u)$$

  - If x and y are composed of n samples
  - **Boundary condition: zero outside the image domain** (see slide on Boundary conditions)

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n-1) \end{bmatrix} = \begin{bmatrix} M(0) & M(-1) & \cdots & M(-k) & 0 & \cdots & 0 \\ M(1) & & & & & & \vdots \\ \vdots & & & & \ddots & & 0 \\ M(k) & & & & & & M(-k) \\ 0 & & \ddots & & & & \vdots \\ \vdots & & & & & & M(-1) \\ 0 & \cdots & 0 & M(k) & \cdots & M(1) & M(0) \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(n-1) \end{bmatrix}$$

- M = Toeplitz (or diagonal-constant) matrix: $M_{i,j} = m(i - j)$

- **Boundary condition: periodicity of the image** (see slide on Boundary conditions)

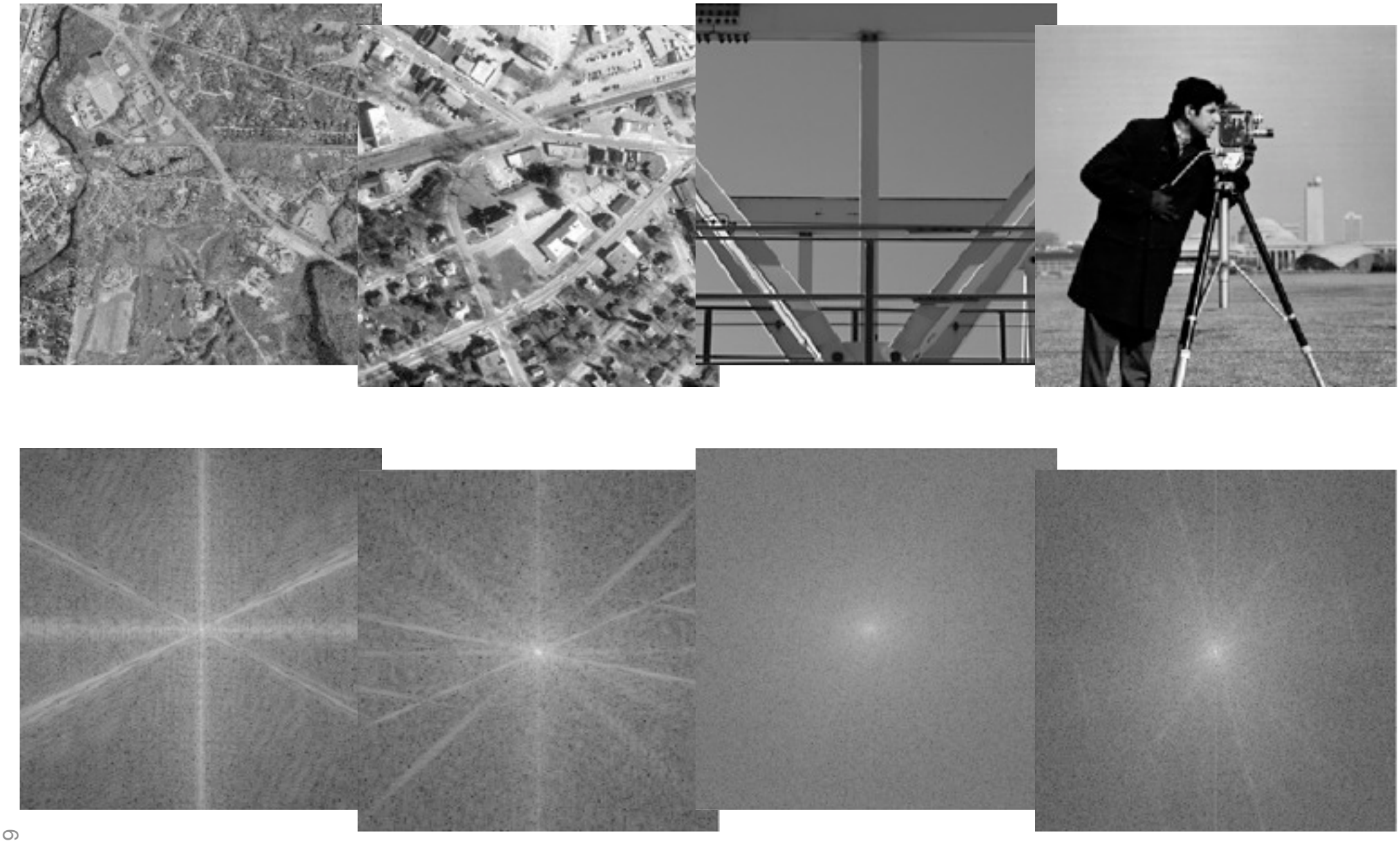
$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(n-1) \end{bmatrix} = \begin{bmatrix} M(0) & M(-1) & \cdots & M(-k) & 0 & \cdots & 0 & M(k) & & \cdots & M(2) & M(1) \\ M(1) & & & & & & & & & & & M(2) \\ \vdots & & & & & & & & & & & \vdots \\ M(k-1) & & & & & & & & & & & M(k) \\ M(k) & & & & & & & & & & & 0 \\ 0 & & & & & \ddots & & & & & & \vdots \\ \vdots & & & \ddots & & & & & & & & \vdots \\ 0 & & & & & & & & & & & M(-k) \\ M(-k) & & & & & & & & & & & M(-k+1) \\ \vdots & & & & & & & & & & & \vdots \\ M(-1) & M(-2) & \cdots & M(-k) & 0 & \cdots & 0 & M(k) & M(k-1) & \cdots & & M(0) \end{bmatrix} \begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(n-1) \end{bmatrix}$$

- M = circulant matrix: Toeplitz matrix of the following form (m(i) = m(i + n))

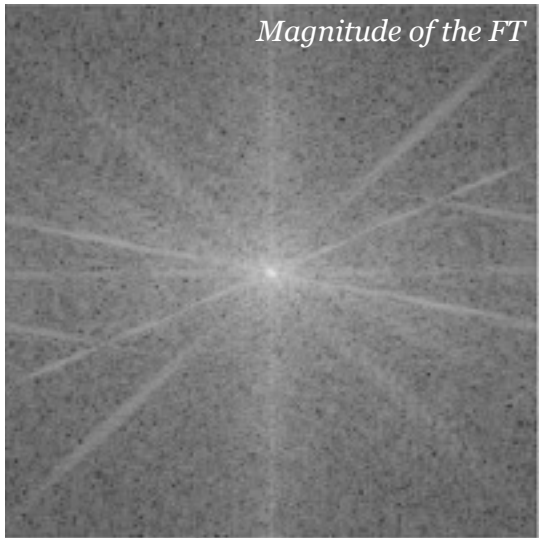
$$\begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & & c_{n-3} \\ \vdots & & & \ddots & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_0 \end{bmatrix}$$

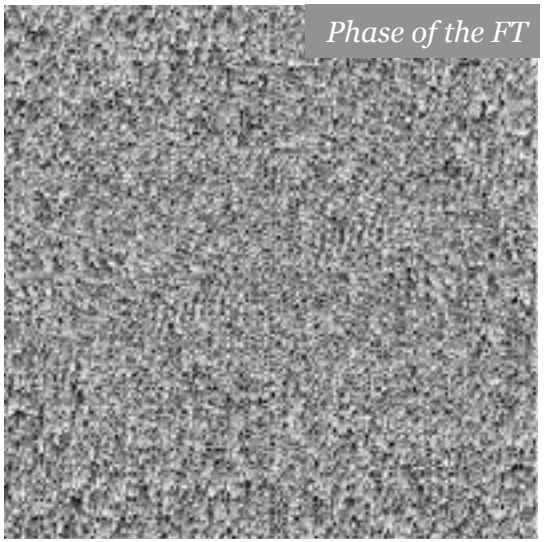
  - Property: the eigenvalues correspond to the discrete Fourier transform of the first row (see discrete Fourier transform)

Reminder > Fourier transform



- **Definition**
  - (Forward transform): $$\mathcal{F}\{f\}(\lambda) = \int_{-\infty}^{+\infty} f(t)\ e^{-2i\pi\lambda t}\ dt$$
  - Inverse transform: $$f(t) = \int_{-\infty}^{+\infty} \mathcal{F}\{f\}(\lambda)\ e^{+2i\pi\lambda t}\ d\lambda$$
- **An alternative definition:** $$\mathcal{F}\{f\}(\lambda) = \int_{-\infty}^{+\infty} f(x)\ e^{-i\lambda x}\ dx \quad \textbf{and} \quad f(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \mathcal{F}\{f\}(\lambda)\ e^{+i\lambda x}\ d\lambda$$
- **Dimension 2** $$\mathcal{F}\{f\}(\lambda,\mu) = \int_{R^2} f(t,u)\ e^{-2i\pi(\lambda t+\mu u)}\ dt\,du$$



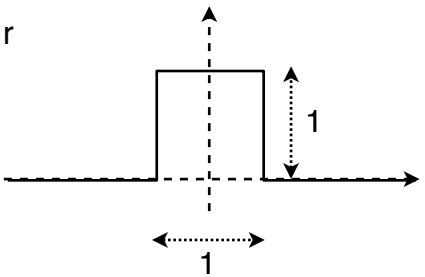


- **Properties**
  - Linearity: $\mathcal{F}\{\alpha\, f + g\} = \alpha\, \mathcal{F}\{f\} + \mathcal{F}\{g\}$
  - Scaling: $\mathcal{F}\{f(a\ x)\}(\lambda) = \dfrac{1}{|a|}\ \mathcal{F}\{f(x)\}\left(\dfrac{\lambda}{a}\right)$
    - If the function *shrinks spatially* (faster variations), its transform spreads (more energy in high frequencies), and conversely
  - Translation: $\mathcal{F}\{f(x-b)\}(\lambda) = e^{-2i\pi\lambda b}\ \mathcal{F}\{f(x)\}(\lambda)$
  - Derivative: $\mathcal{F}\{f^{(n)}\}(\lambda) = (2i\pi\lambda)^n\ \mathcal{F}\{f\}(\lambda)$
  - Convolution theorem...(see convolution)
  - Plancherel theorem: $\displaystyle\int_{-\infty}^{+\infty} |f(x)|^2\ dx = \int_{-\infty}^{+\infty} |\mathcal{F}\{f\}(\lambda)|^2\ d\lambda$ (the Fourier transform preserves the energy of the signal)
  - Symmetry and domain

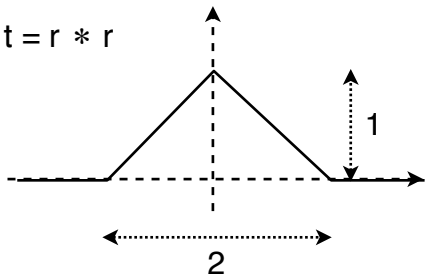
	f	$\mathcal{F}$
f is even	real	real
	imaginary	imaginary
	complex	complex
f is odd	real	imaginary
	imaginary	real
	complex	complex

- **Some functions/distributions and their transform**

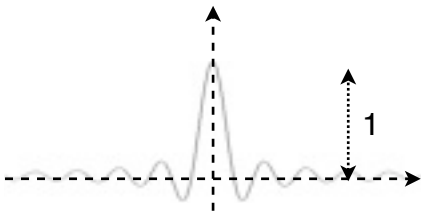
  - Rectangular function r



  - Triangular function t = r ∗ r



  - Normalized sinc function sinc(x) = sin(π x) / (π x)



f	$\mathcal{F}$
r	sinc
sinc	r
t	sinc^2
sinc^2	t
Gaussian $\exp(-x^2)$	Gaussian $\sqrt{(\pi)}\ \exp(-(\pi\ \lambda)^2)$
1	δ
δ	1

- **Definition**
  - One-dimensional signal f composed of n samples
    - (Forward transform): $\mathcal{F}\{f\}(k) = \sum_{t=0}^{n-1} f(t)\ e^{-2i\pi k \frac{t}{n}}$
    - Inverse transform: $f(t) = \frac{1}{n} \sum_{k=0}^{n-1} \mathcal{F}\{f\}(k)\ e^{2i\pi k \frac{t}{n}}$
    - Remark: $\mathcal{F}\{f\}$ is periodic with period n
      - Space/time discrete ➜ periodic in frequency (and conversely)

- **Parseval theorem:** $\sum_{t=0}^{n-1} |f(t)|^2 = \frac{1}{n} \sum_{k=0}^{n-1} |\mathcal{F}\{f\}(k)|^2$

Lecture: Definition, ill-posedness, regularization

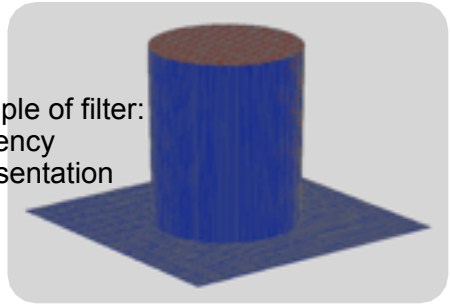
- **Overview**
  - Terminology
  - Types of inverse problems
- **Some reminders**
  - Convolution, Fourier
- **Forward problem**
  - Model
  - Deterministic and stochastic points of view
- **Ill-posed problem**
  - Existence
  - Uniqueness
  - Stability
- **Regularization**
  - Constrained approach
  - Lagrangian approach
    - Equivalence
  - Derivation of a smoothness constraint
  - L1-regularization

- **Model**
  - Implicit: m(x,y,b) = 0 (the noise b accounts for acquisition errors and model mismatch)
  - Explicit: y = m(x) # b

- **Terminology**
  - Transformation m (ex. convolution)
    - Characterized by the impulse response (also point spread function (PSF))
  - Perturbation
    - Noise
  - Discretization (additive noise): y = M x + b
    - Ex. reconstruction: backprojection: $s = M^t y$
      - Not an inversion since $M^t M \neq Id$ (in general)

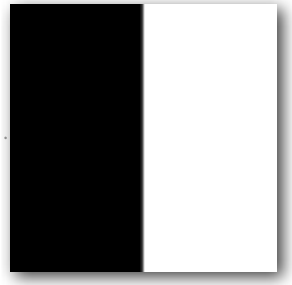
- **Convolution:** $y = m * x + b$
  - Integral over the domain of x
  - In practice, x has a finite support
  - Impulse response
    - x = Dirac
    - Response = m
- **Finite impulse response filter**
  - m has a finite support
    - Discretization of the convolution: finite sum
  - Stable filter (in the sense: output stable with respect to input)
- **Infinite impulse response filter**
  - m has an infinite support
    - Particular case: recursive filter
    - y = sum of a function of x and a function of y
      - Necessity of initial conditions
  - Stability not always guaranteed (stability study (discrete implementation): Z-transform)

- **Fourier series**
  - Decomposition of non-smooth functions: oscillations at discontinuities
- **In signal processing**
  - Fourier transform
    - Removing high-frequencies from the representation = low-pass filtering
    - Filtering: Fourier transform of the image .* Filter frequency representation



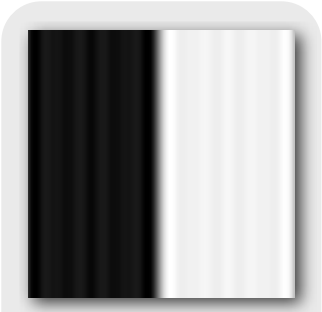
Example of filter: frequency representation

  - Image with a discontinuity

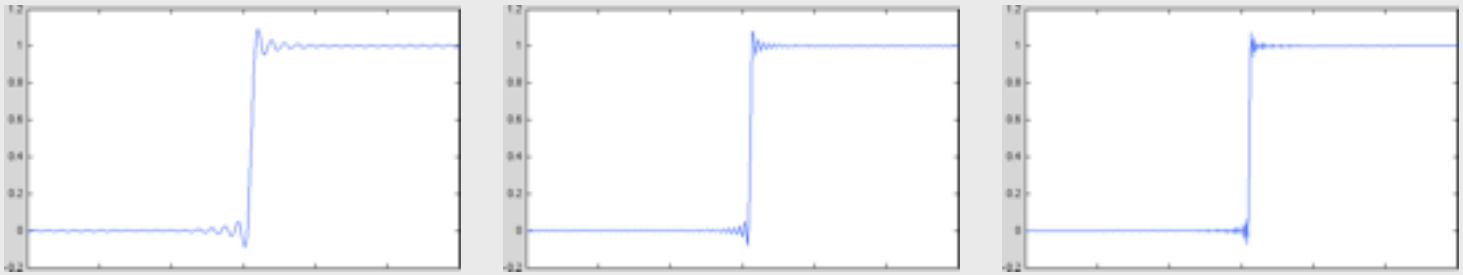


Example on a real image: ringing artifacts

    - Expected result: discontinuity ➜ smooth transition
    - Actual result: smoother transition + oscillations

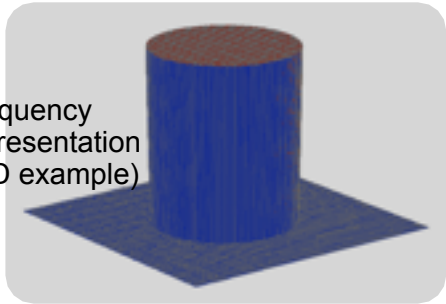


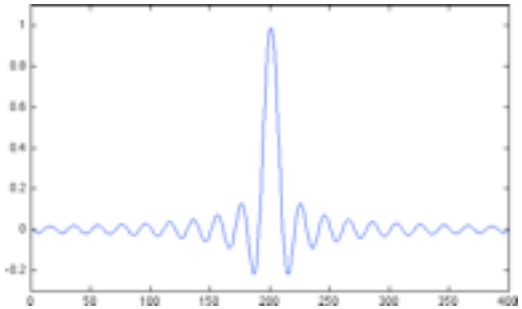
Wider mask in Fourier



- **Explanation**
  - Pointwise multiplication in Fourier ⇔ convolution by a sinc function in space

Frequency representation (2-D example)



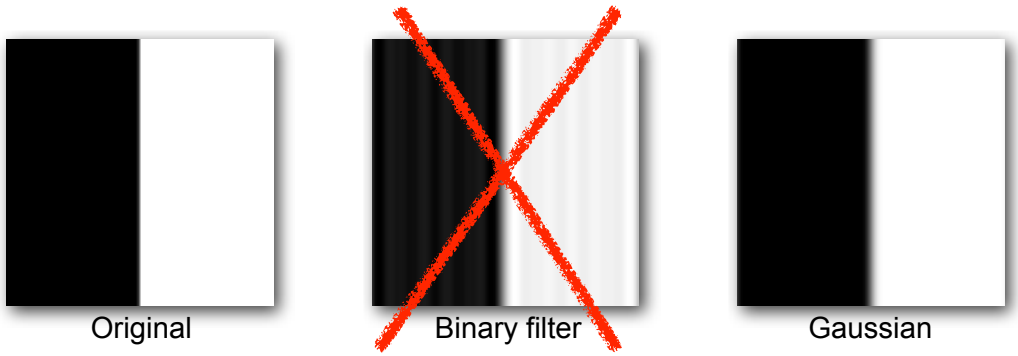


Spatial representation (1-D)

  - Oscillations

- **Solution**
  - Multiplication of the filter by an apodization (or tapering) window
    - Cosinus, Hanning, Hamming, Blackman…

  - Use a smooth filter: example with a Gaussian



Original | Binary filter | Gaussian

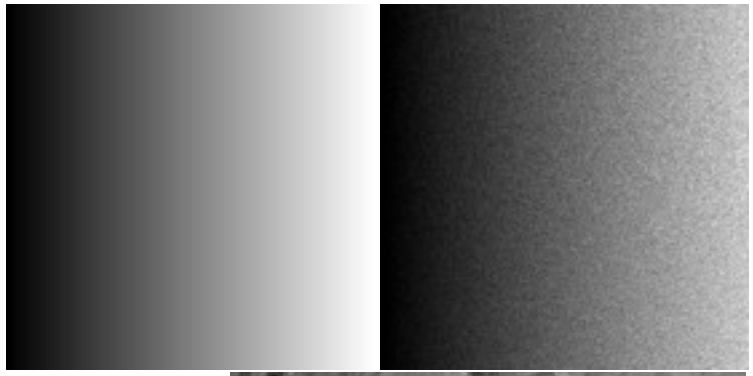
- **Additive Gaussian noise: y = m(x) + b**
  - b: centered Gaussian noise with covariance matrix Cov

$$p(b) = \left((2\pi)^{\frac{N}{2}}\sqrt{\det(\mathrm{Cov})}\right)^{-1} \exp\left\{-\frac{1}{2} b^{T}\ \mathrm{Cov}^{-1}\ b\right\}$$

  - If b is a white noise: Cov = σ^2 I

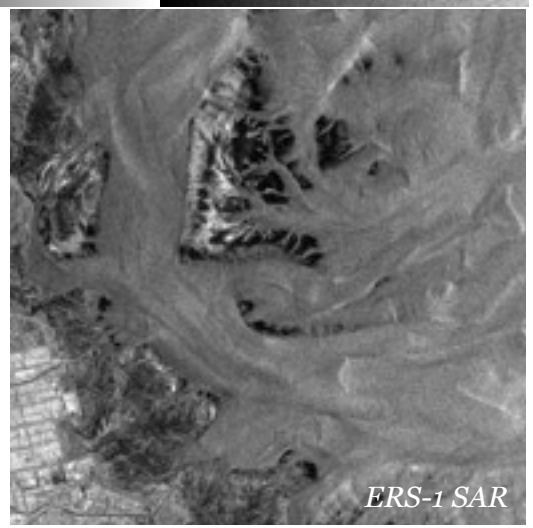
- **Poisson process**
  - Photon counting at low flow
    - Astronomy
    - Nuclear medicine imaging (SPECT/PET)

$$p(y|x) = \Pi_{i=1}^{N} e^{-x_i}\ \frac{x_i^{y_i}}{y_i!}$$



- **Multiplicative noise: Speckle**
  - Synthetic-aperture radar (SAR) imaging
    - Interference between close wavelengths
  - b: white noise

$$p(b) = \begin{cases} \sigma^{-2}\exp\left\{-\frac{b-1}{\sigma^2}\right\}, & b \geq 1 \\ 0 & \text{otherwise} \end{cases}$$



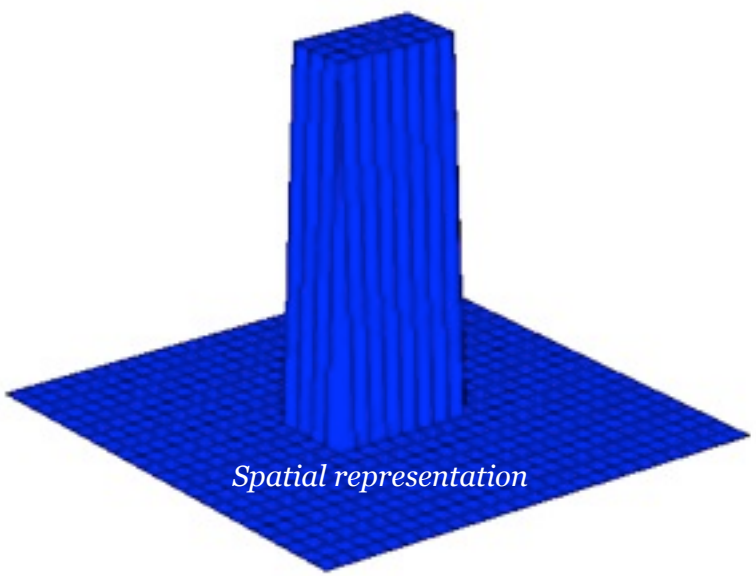
- **Deterministic**
  - The only random component is the noise (measure, compression…)
  - Algebraic (discrete)
    - "System solving" $\text{observation} = \mathsf{M} \times \text{unknown}$
    - Finite elements: resolution on a basis $\text{observation} = \int_{\text{domain}} \mathsf{M}(\text{unknown})$
  - Analytical: resolution of the continuous problem $\text{observation} = \int_{\text{domain}} \mathsf{M}(\text{unknown})$

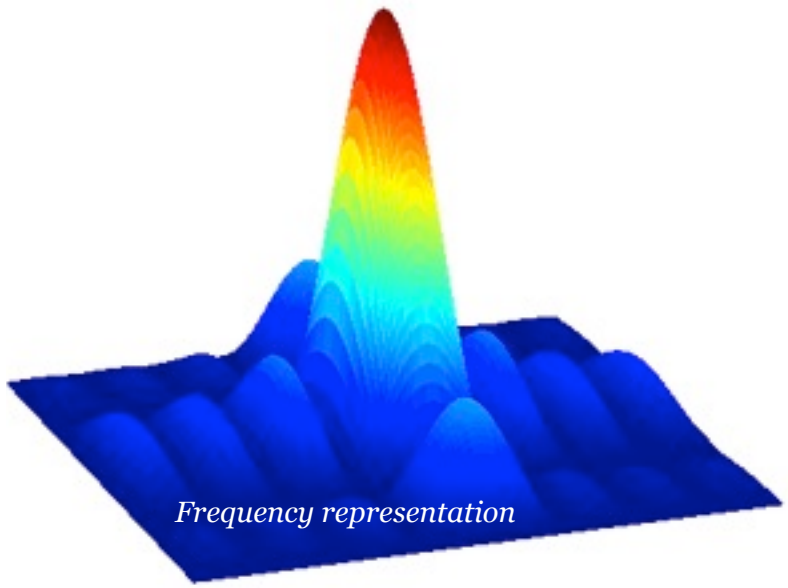
- **Stochastic**
  - The unknown is a realization of a random variable
  - Wiener filter
  - Random walk
  - Maximum likelihood
  - Maximum a posteriori
  - Maximum entropy
  - Minimum relative entropy

Lecture: Definition, ill-posedness, regularization

- **Overview**
  - Terminology
  - Types of inverse problems
- **Some reminders**
  - Convolution, Fourier
- **Forward problem**
  - Model
  - Deterministic and stochastic points of view
- **Ill-posed problem**
  - Existence (over-determined problems with incoherences)
  - Uniqueness
  - Stability
- **Regularization**
  - Constrained approach
  - Lagrangian approach
    - Equivalence
  - Derivation of a smoothness constraint
  - L1-regularization

- **Model:** $y(u) = \int_R m(v)\, x(u-v)\, dv$
  - Convolution by m: low-pass filter





  - Some frequencies are set to zero during the filtering
  - They cannot be perfectly recovered during the inverting process

- **Model:** $y(u) = \int_a^b m(u,v)\ x(v)\ dv$ (convolution: m(u,v) = m(u-v))

- **Riemann-Lebesgue lemma**
  - Condition: m(u,v) integrable on [a,b]

$$\lim_{n\to\infty} \int_a^b m(u,v)\ \sin(nv)\ dv = 0$$

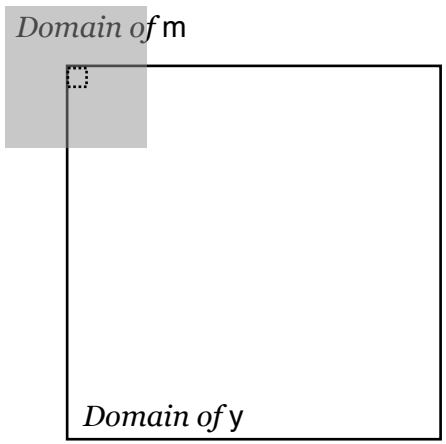
$$\lim_{n\to\infty} \int_a^b m(u,v)\ \cos(nv)\ dv = 0$$

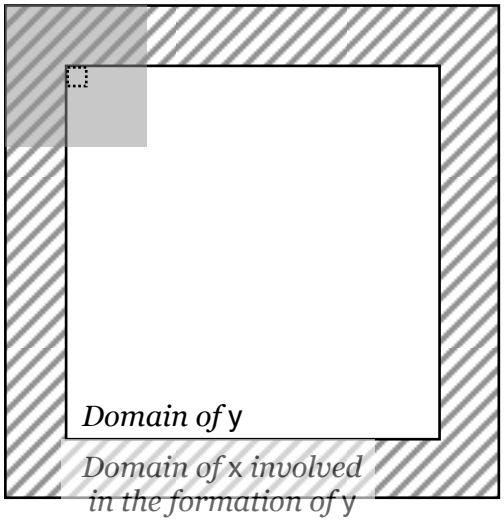
- **Consequence**

$$\lim_{n\to\infty} \int_a^b m(u,v)\ (x(v) + \sin(nv))\ dv = y(u)$$

  - The solution is not unique since
    - if x is a solution
    - then (x + high-frequency signal) is also a solution

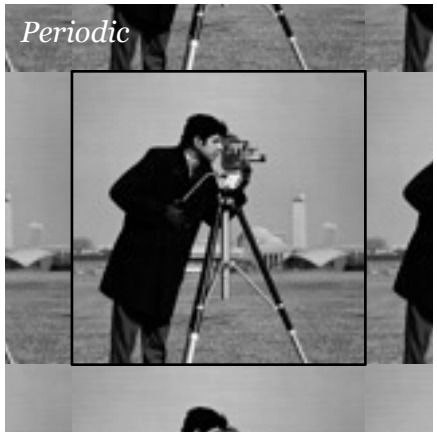
- **Model:** $y(u) = \int_R m(v)\, x(u-v)\, dv$
  - Convolution by m

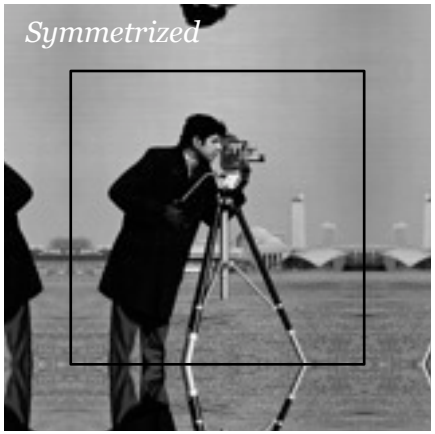


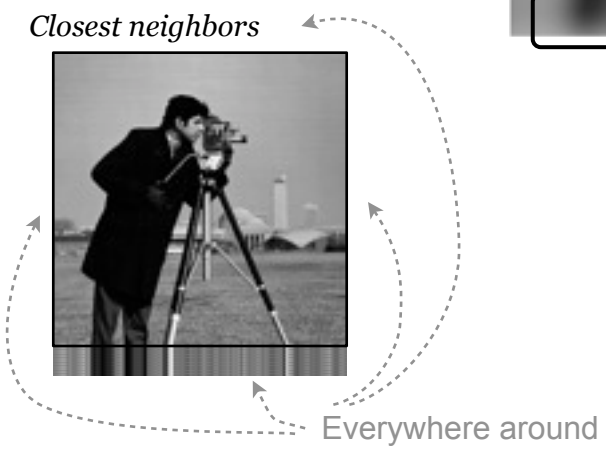


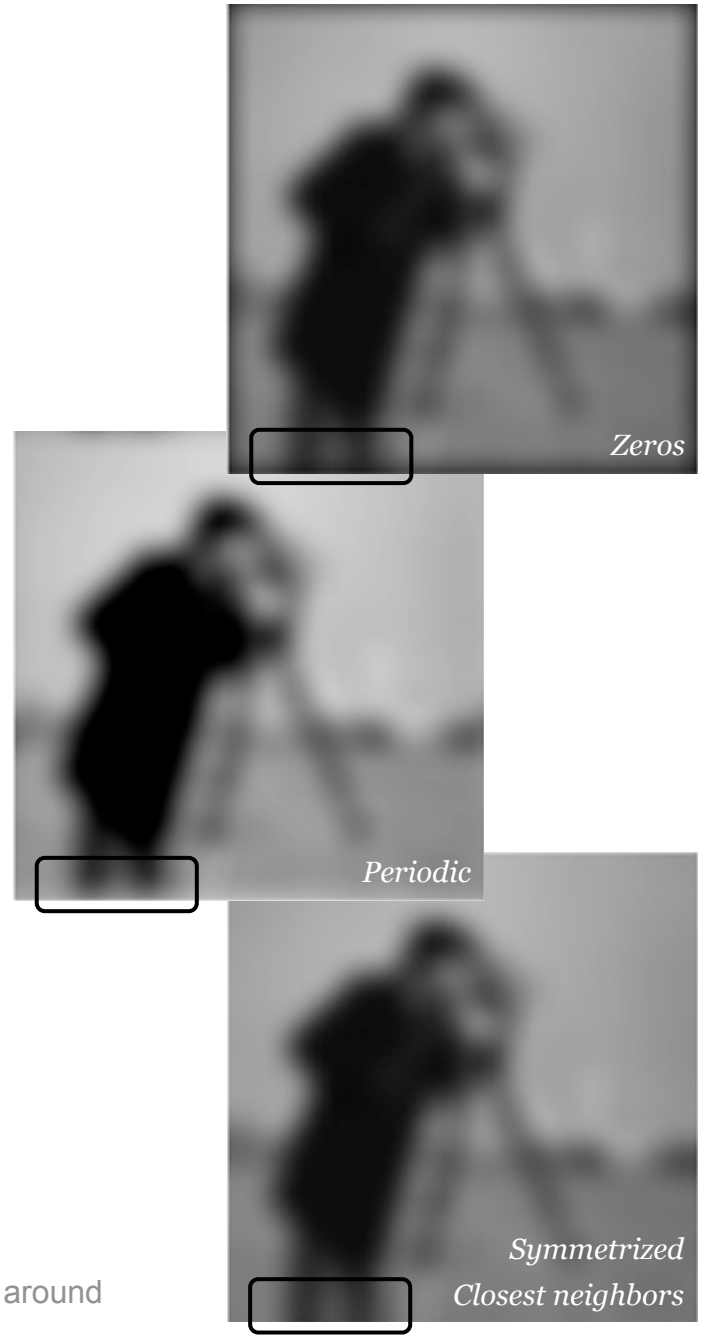
  - Under-determined problem: more unknowns than observations
    - Necessity to impose boundary conditions (set the missing information to some arbitrary values)

- **Arbitrary values in: domain(x) \ domain(y)**
  - Filling with zeros
  - Assume periodic image (discrete Fourier transform)
    - Problem of continuity
  - Image symmetrization
  - Closest neighbors

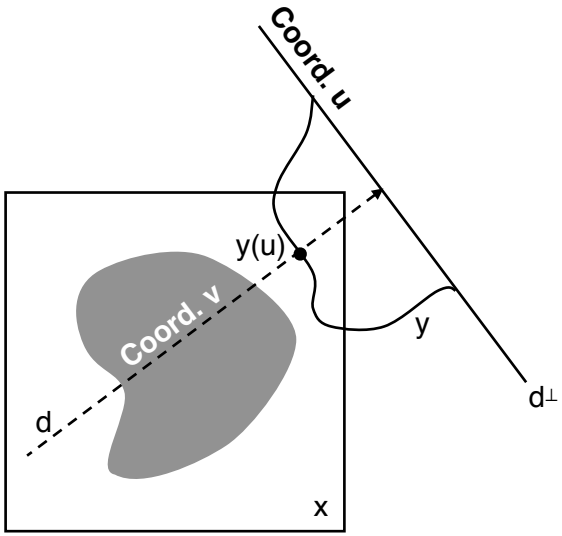


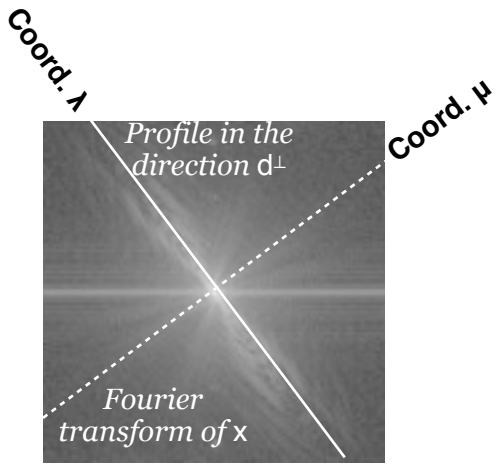






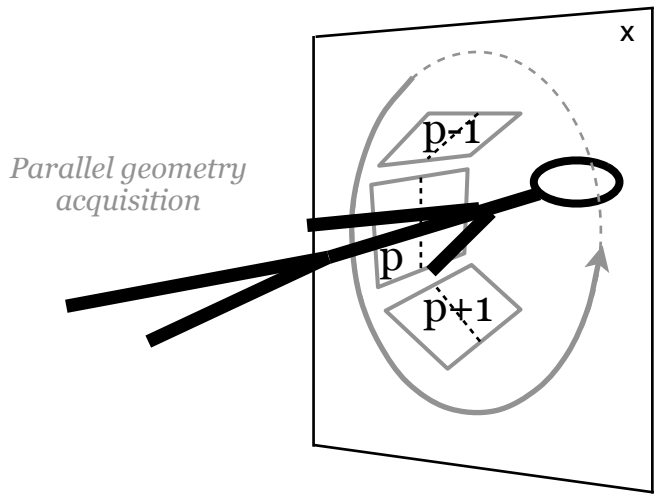
- **Dimension 2**
  - The 1-D Fourier transform of the projection y in the direction d of a 2-D function x is equal to the profile in the direction $d^{\perp}$ passing through the origin of the (2-D) Fourier transform of x

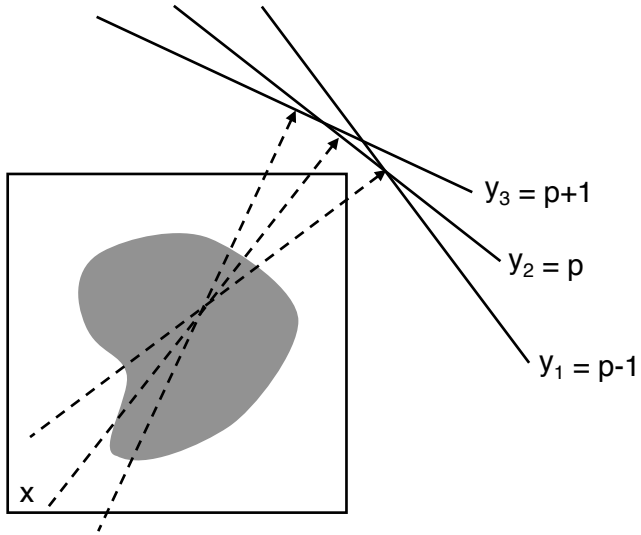




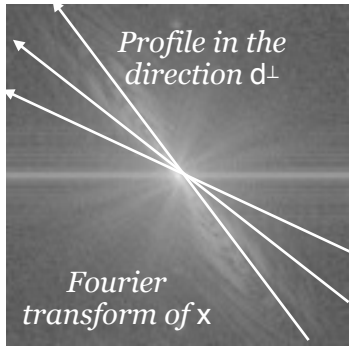
- **Extension to dimension N**

- **Nuclear medicine imaging / X-ray scan**
  - Parallel geometry
    - N projections over 360°





- **Fourier space filling**



  - Partial ➜ the solution is not unique
  - Higher density in the low frequencies ➜ homogeneous areas better *represented* than contours

Lecture: Definition, ill-posedness, regularization

- **Overview**
  - Terminology
  - Types of inverse problems
- **Some reminders**
  - Convolution, Fourier
- **Forward problem**
  - Model
  - Deterministic and stochastic points of view
- **Ill-posed problem**
  - Existence
  - Uniqueness
  - Stability
- **Regularization**
  - Constrained approach
  - Lagrangian approach
    - Equivalence
  - Derivation of a smoothness constraint
  - L1-regularization

- **System:** M x = y

model source observation

$$\begin{pmatrix} 1000 & 999 \\ 999 & 998 \end{pmatrix} x = \begin{pmatrix} 1999 \\ 1997 \end{pmatrix}$$

- Solution $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$
- Perturbation on the observations $\delta y = \begin{pmatrix} -0.01 \\ 0.01 \end{pmatrix}$
- Solution of the new system $x + \delta x = \begin{pmatrix} 20.97 \\ -18.99 \end{pmatrix}$
- Variation
  - of the observations 5e-4 %
  - of the solution 2000 %

- **System:** M x = y

- **Matrix norm**

  - $|\mathsf{M}| = \sup_{\mathsf{x}\neq 0} \frac{|\mathsf{M}\ \mathsf{x}|}{|\mathsf{x}|}$ (measures by how much at most the transformation M *stretches* a vector x)

    $= \sup_{|\mathsf{x}|=1} |\mathsf{M}\ \mathsf{x}|$
  - Example: $|\mathsf{M}|_2$ = square root of the largest eigenvalue of $\mathsf{M}\ \mathsf{M}^T$

- **Condition number**

$$\kappa(\mathsf{M}) = |\mathsf{M}^{-1}|\ |\mathsf{M}|$$ (invertible matrix)

$$= \frac{\sigma_{\max}(\mathsf{M})}{\sigma_{\min}(\mathsf{M})}$$ (L^2 norm / singular values σ = square root of the eigenvalues of $\mathsf{M}\ \mathsf{M}^T$)

$$= \frac{\lambda^{\text{abs}}_{\max}(\mathsf{M})}{\lambda^{\text{abs}}_{\min}(\mathsf{M})}$$ (L^2 norm / min and max eigenvalues in absolute value)

$$\geq 1$$

- Close to 1: well-conditioned matrix (equal to 1: rotations, which includes identity)
- Large: badly-conditioned matrix
- Infinite: singular matrix
  - System of previous slide: κ(M) ~ 400000

- **System:** M x = y, M invertible
  - Perturbation: y + δy = M (x + δx)
  - Upper bound

$$|\delta \mathrm{x}| \leq |\mathrm{M}^{-1}| \; |\delta \mathrm{y}|$$

$$\frac{1}{|\mathrm{x}|} \leq |\mathrm{M}| \; \frac{1}{|\mathrm{y}|}$$

$$\frac{|\delta \mathrm{x}|}{|\mathrm{x}|} \leq \kappa(\mathrm{M}) \; \frac{|\delta \mathrm{y}|}{|\mathrm{y}|}$$

- **Convolution:** $y = m * x + b$

- **Discretization**
  - Matrix formulation: $y = M x + b$
  - If x is periodic (or boundary condition = periodic assumption – see Boundary conditions)
  - Then M is circulant
    - Its first row is composed of the filter coefficients
  - Property: eigenvalues of M = discrete Fourier transform of its first row

- **Low-pass filter**
  - The Fourier transform magnitude decreases with frequency
  - M has some small eigenvalues ➜ badly conditioned

- **Restoration of a low-pass filtered image: always ill-posed**

- **Convolution:** $y = m * x + b$
  - Fourier: $\hat{Y} = \hat{M}\,\hat{X} + \hat{B}$
  - Solution: $\hat{S} = \dfrac{\hat{Y}}{\hat{M}}$ (approximation: the noise is neglected)

    $= \hat{X} + \dfrac{\hat{B}}{\hat{M}}$ (actually)

- **Low-pass filter: $|\hat{M}|$ is small in high frequencies**
  - Noise = high-frequency signal
    - $|\hat{B}|$ is large in high frequencies
  - The noise is amplified

- **Hadamard's definition**
  - An inverse problem is well-posed if and only if it fulfills the following 3 conditions
    - There exists a solution: S (existence)
    - S is unique (uniqueness)
    - S depends continuously on the observations (stability)

- **If a problem is ill-posed**
  - It cannot be solved without a priori information (i.e., constraints on the solution)
  - Adding such information is known as regularization

  - Terminology
    - Primary (unregularized) problem: $y = m(x)$
      - Expressed as $\min_x \psi(y, x)$, for example: $\min_x |y - m(x)|^2$
      - Referred to as data fidelity [the solution x must be such that m(x) is close to the actual observation (or data) y]

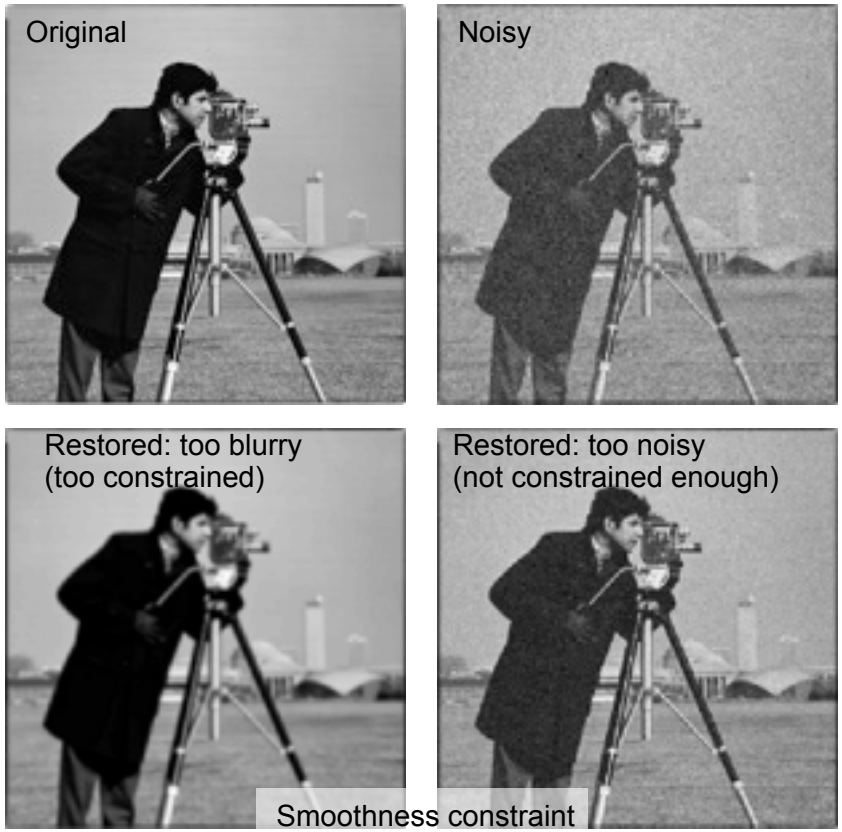
    - Regularization: $\varphi(x)$
      - Also referred to as a priori information, a priori knowledge, constraint
        - where "a priori" is a shortcut for "before taking into account the observation"

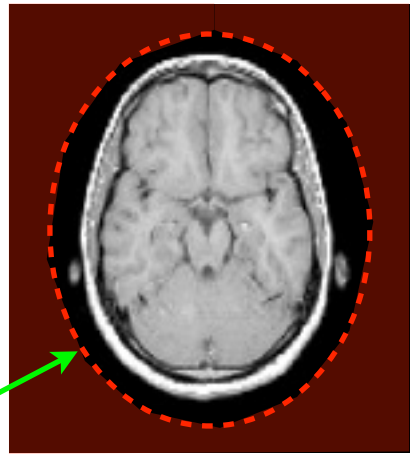
    - A posteriori context: when taking into account both the a priori information and the observation

Lecture: Definition, ill-posedness, regularization

- **Overview**
  - Terminology
  - Types of inverse problems
- **Some reminders**
  - Convolution, Fourier
- **Forward problem**
  - Model
  - Deterministic and stochastic points of view
- **Ill-posed problem**
  - Existence
  - Uniqueness
  - Stability
- **Regularization**
  - Constrained approach
  - Lagrangian approach
    - Equivalence
  - Derivation of a smoothness constraint
  - L1-regularization

- **Parameterization**
  - Solution in a given parametric space
- **Smoothness**
  - Global (Ex. Tikhonov)
    - Smoothes out the noise in homogeneous areas
      - Also smoothes the edges
    - Regularization weight tends toward infinity
      - Enforces a constant image
  - Local (half-quadratic)
    - Edge-preserving smoothing
- **Positivity**
  - Usually an image is positive
- **Minimal support**
  - The solution should be equal to zero outside a given domain
    - Medical imaging
  - Enforced by projection
    - At each iteration of the minimization
  - By means of a constraint





Outside the ellipse, it is *known* that there is no information

- **Observations**
  - Noisy
  - Incomplete
    - Nuclear medicine imaging, X-ray scan… (see the projection theorem)

- **Minimization**
  - A priori information on the solution
  - Expressed as: $\min_x J_2(x)$
    - For example: $J_2(x) = \int \varphi(|\nabla x|)$

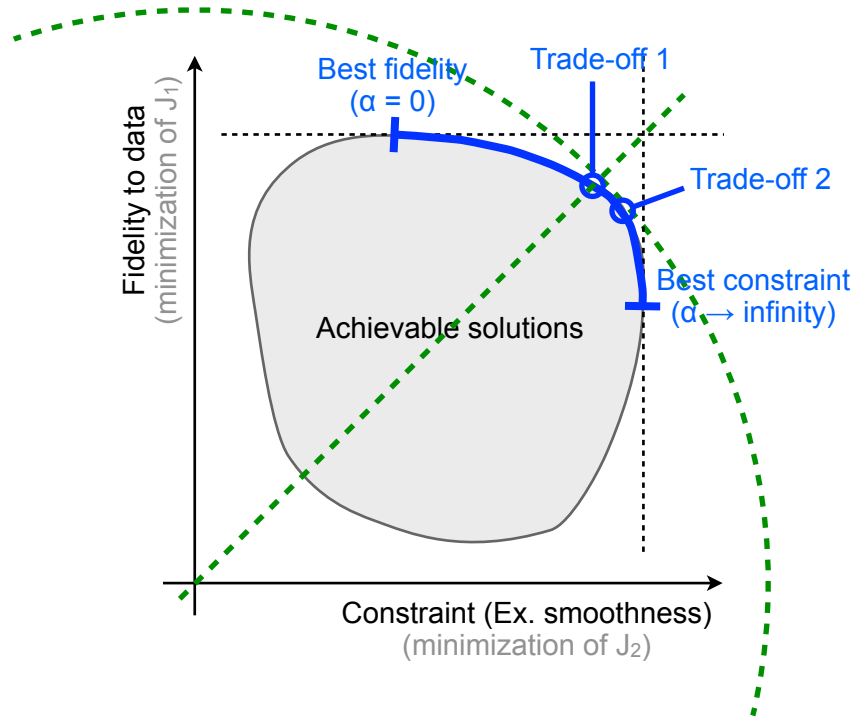
- **Under the constraint**
  - Model: $y = m(x) + b$
    - Set of admissible solutions
  - Expressed as: $J_1(x,y) = y - m(x) \approx 0$

$$\begin{cases} \min_x J_2(x) \\ \text{such that } J_1(x,y) = 0 \end{cases}$$

- **Constrained minimization**
  - $J(x, \alpha) = J_2(x) + \alpha\, J_1(x,y)$

  - If x_0 is a solution of the problem "$\min_x J_2(x)$ with the constraint $J_1(x,y) = 0$"
  - Then there exists α_0 such that $\nabla J(x_0, \alpha_0) = 0$
    - α_0: Lagrange multiplier

- **Observations**
  - Noisy
  - Incomplete
    - Nuclear medicine imaging, X-ray scan…
- **A priori information**
  - Properties of the solution
  - Expressed as: $\min_x J_2(x)$
- **Model:** $y = m(x) + b$
  - Set of admissible solutions
  - Expressed as: $\min_x J_1(x,y)$
    - Ex. $\min_x |y - m(x)|^2$
- **Regularized problem**
  - $\min_x J_1(x,y) + \alpha J_2(x)$
    - α: regularization parameter to tune elative influence of the 2 terms
    - Sometimes called hyperparameter



- **Linear**
  - E.g.: Tikhonov $|x|^2$
    - $\underbrace{M\,x = y}_{\text{Unregularized}}$ ➜ $\underbrace{\min_x |y - M\,x|^2 + \alpha\,|x|^2}_{\text{Possible regularization}}$

model source observation

$$\begin{pmatrix} 1000 & 999 \\ 999 & 998 \end{pmatrix} x = \begin{pmatrix} 1999 \\ 1997 \end{pmatrix}$$

  - Solution (see Pseudo-inverse)
    - $x = (M^T M + \alpha\, I)^{-1} M^T y$

Unregularized solution α = 0	Regularized solution α = 5e-3
$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$x = \begin{pmatrix} 1.0005 \\ 0.9995 \end{pmatrix}$
$\delta y = \begin{pmatrix} -0.01 \\ 0.01 \end{pmatrix}$ 5e-4%	
$x + \delta x = \begin{pmatrix} 20.97 \\ -18.99 \end{pmatrix}$ 2000%	$x + \delta x = \begin{pmatrix} 1.0015 \\ 0.9985 \end{pmatrix}$ 1%

    - Increasing α:
      - increases stability
      - decreases the norm of the solution (by making the a priori stronger)

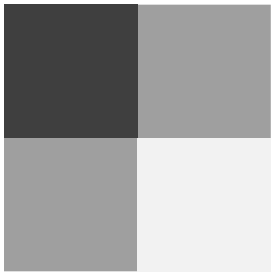
- **Non linear**
  - Ex. L^p norm of x, $p \neq 2$

- **Tikhonov regularization on gradient**
  - Smoothness: strong correlations between neighboring pixels
  - Minimize: $J_2^T(x) = \int |\nabla x|^2 \, dx$

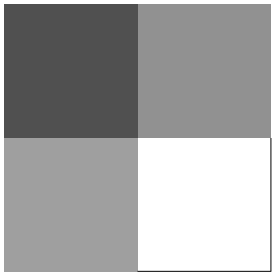
- **Maximum entropy-like regularization**
  - Minimum correlation between the pixels
  - Minimize: $J_2^H(x) = \int x \log(x) \, dx$

- **Example: image with 4 pixels**
  - Observations
    - $p_1 + p_2 + p_3 + p_4 = 36$
    - $p_1 + p_2 = 12$
    - $p_1 + p_4 = 12$

p_1	p_4
p_2	p_3



Solution regularized with Tikhonov



Solution regularized with Entropy

  - $J_2^T(x^T) = 144$ $\quad$ $J_2^T(x^H) = 160$

    $J_2^H(x^H) = 83.2$ $\quad$ $J_2^H(x^T) = 83.5$

- **Model:** y = m(x) + b ➜ $\min_x$ |y - m(x)|²
  - Continuous framework: $|y - m(x)|^2 = \int (y - m(x))^2$
    - Data fidelity: $J_1(x, y) = \int (y - m(x))^2$
    - Regularization: $J_2(x) = \int \varphi(|\nabla x|)$
- **Energy to be minimized:** $J(x) = J_1(x, y) + \lambda \int \varphi(|\nabla x|)$
- **Derivative** (see next slides for proof)
  - Euler-Lagrange equation: x such that $J'(x) = \frac{\partial J_1}{\partial x}(x, y) - \mathrm{div}\left(\frac{\varphi'(|\nabla x|)}{|\nabla x|}\ \nabla x\right) = 0$

- **One dimensional**
  - x: a function of u, a real number in [a,b]

$$J(x) = \int_a^b L(x(u), x'(u), u)\ du$$

  - Notation

$$J(x) = \int_a^b L(x, x', u)\ du$$

  - Assumption: J reaches its minimum at x_0

- ***Displacement*** **h such that h(a) = h(b) = 0**

$$J(\epsilon) = \int_a^b L(x_0 + \epsilon h, x_0' + \epsilon h', u)\ du$$

  - Euler-Lagrange condition

$$J'(0) = \int_a^b \left[h\ \frac{\partial L}{\partial x} + h'\ \frac{\partial L}{\partial x'} \right]\ du = 0$$

Integration by parts

$$\int_a^b h'\ \frac{\partial L}{\partial x'}\ du = \left[h\ \frac{\partial L}{\partial x'} \right]_a^b - \int_a^b h\ \frac{d}{du}\frac{\partial L}{\partial x'}\ du$$

Therefore

$$J'(0) = \int_a^b \left[\frac{\partial L}{\partial x} - \frac{d}{du}\frac{\partial L}{\partial x'} \right] h\ du = 0$$

Then, based on the fundamental lemma of calculus of variations

$$\underbrace{\frac{\partial L}{\partial x} - \frac{d}{du}\frac{\partial L}{\partial x'}}_{J'(x)} = 0 \qquad \text{for all } u \in [a,b]$$

- **The unknown is a function of n variables**

$$J(x) = \int_D L\left(x, \frac{\partial x}{\partial u_1}, \frac{\partial x}{\partial u_2} \cdots \frac{\partial x}{\partial u_n}, u_1, u_2 \ldots u_n\right) du$$

  - Euler-Lagrange equation: $\frac{\partial L}{\partial x} - \sum_{i=1}^{n} \frac{d}{du_i} \frac{\partial L}{\partial \frac{\partial x}{\partial u_i}} = 0$

- **Applied to images**

  - Two variables u and v (horizontal and vertical axes): x = x(u,v)
  - $$L\left(x, \frac{\partial x}{\partial u_1}, \frac{\partial x}{\partial u_2} \cdots \frac{\partial x}{\partial u_n}, u_1, u_2 \ldots u_n\right) = \varphi(|\nabla x|) = \varphi\left(\frac{\partial x}{\partial u}, \frac{\partial x}{\partial v}\right)$$
  - Euler-Lagrange equation: $$\frac{\partial L}{\partial x} - \frac{d}{du} \frac{\partial L}{\partial \frac{\partial x}{\partial u}} - \frac{d}{dv} \frac{\partial L}{\partial \frac{\partial x}{\partial v}} = 0 \iff -\operatorname{div}\left(\frac{\partial L}{\partial \nabla x}\right) = 0$$

- **Divergence expression**

$$-\operatorname{div}\left(\frac{\partial \mathsf{L}}{\partial \nabla \mathsf{x}}\right) = 0$$

$$\iff \quad -\operatorname{div}\left(\frac{\mathsf{d}\varphi(|\nabla \mathsf{x}|)}{\mathsf{d}\nabla \mathsf{x}}\right) = 0$$

  - where: $$\frac{\mathsf{d}\varphi(|\nabla \mathsf{x}|)}{\mathsf{d}\nabla \mathsf{x}} = \frac{\mathsf{d}|\nabla \mathsf{x}|}{\mathsf{d}\nabla \mathsf{x}}\ \varphi'(|\nabla \mathsf{x}|) = \frac{\nabla \mathsf{x}}{|\nabla \mathsf{x}|}\ \varphi'(|\nabla \mathsf{x}|)$$

- **Finally:** $$\mathsf{J}'(\mathsf{x}) = \frac{\partial \mathsf{J}_1}{\partial \mathsf{x}}(\mathsf{x},\mathsf{y}) - \operatorname{div}\left(\frac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}\ \nabla \mathsf{x}\right)$$

- **Iterative minimization**

  - Gradient descent: $$\frac{\partial \mathsf{x}}{\partial \tau} = -\frac{\partial \mathsf{J}_1}{\partial \mathsf{x}}(\mathsf{x},\mathsf{y}) + \operatorname{div}\left(\frac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}\ \nabla \mathsf{x}\right)$$

- **Divergence expression**

$$\mathrm{div}\left(\frac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}\ \nabla \mathsf{x}\right)$$

- Notation: $\mathsf{b}(|\nabla \mathsf{x}|) = \dfrac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}$

$$\begin{aligned}\mathrm{div}(\mathsf{b}(|\nabla \mathsf{x}|)\ \nabla \mathsf{x}) &= \nabla \mathsf{b}(|\nabla \mathsf{x}|) \cdot \nabla \mathsf{x} + \mathsf{b}(|\nabla \mathsf{x}|)\ \underbrace{\mathrm{div}(\nabla \mathsf{x})}_{\Delta \mathsf{x}} \\ &= \nabla|\nabla \mathsf{x}|\ \mathsf{b}'(|\nabla \mathsf{x}|) \cdot \nabla \mathsf{x} + \mathsf{b}(|\nabla \mathsf{x}|)\ \Delta \mathsf{x}\end{aligned}$$

- Hessian matrix: H_x

$$\nabla|\nabla \mathsf{x}| = \frac{1}{|\nabla \mathsf{x}|}\ \mathsf{H}_\mathsf{x}\ \nabla \mathsf{x}$$

- Therefore:

$$\begin{aligned}\mathrm{div}(\mathsf{b}(|\nabla \mathsf{x}|)\ \nabla \mathsf{x}) &= \mathsf{b}'(|\nabla \mathsf{x}|)\left(\frac{1}{|\nabla \mathsf{x}|}\ \mathsf{H}_\mathsf{x}\ \nabla \mathsf{x}\right) \cdot \nabla \mathsf{x} + \mathsf{b}(|\nabla \mathsf{x}|)\ \Delta \mathsf{x} \\ &= \mathsf{b}'(|\nabla \mathsf{x}|)\ \frac{\nabla \mathsf{x}^{\mathrm{T}}}{|\nabla \mathsf{x}|}\ \mathsf{H}_\mathsf{x}\ \nabla \mathsf{x} + \mathsf{b}(|\nabla \mathsf{x}|)\ \Delta \mathsf{x}\end{aligned}$$

- **Divergence expression**

$$\mathrm{div}(\mathsf{b}(|\nabla \mathsf{x}|)\ \nabla \mathsf{x}) = \mathsf{b}'(|\nabla \mathsf{x}|)\ \frac{\nabla \mathsf{x}^{\mathrm{T}}}{|\nabla \mathsf{x}|}\ \mathsf{H}_{\mathsf{x}}\ \nabla \mathsf{x} + \mathsf{b}(|\nabla \mathsf{x}|)\ \Delta \mathsf{x}$$

- **Property of the Hessian matrix:** $\mathsf{a}^{\mathsf{T}}\ \mathsf{H}_{\mathsf{x}}\ \mathsf{b} = \mathsf{D}_{\mathsf{a}}\mathsf{D}_{\mathsf{b}}\mathsf{x}$ if $|\mathsf{a}|=|\mathsf{b}|=1$

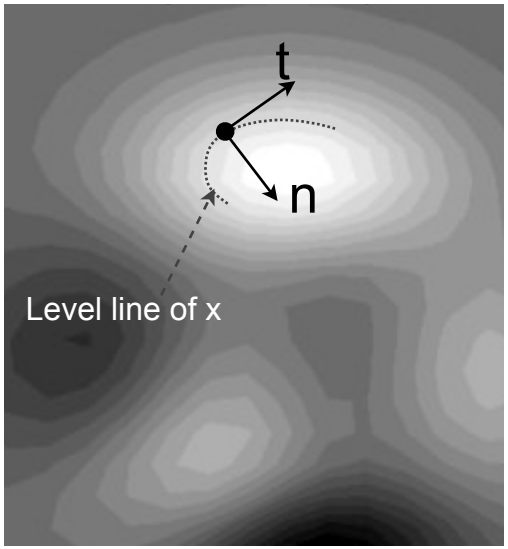
$$\mathsf{n} = \frac{\nabla \mathsf{x}}{|\nabla \mathsf{x}|} \quad \rightarrow \quad \frac{\nabla \mathsf{x}^{\mathrm{T}}}{|\nabla \mathsf{x}|}\ \mathsf{H}_{\mathsf{x}}\ \nabla \mathsf{x} = |\nabla \mathsf{x}|\ \mathsf{x}_{\mathsf{nn}}$$

  - Property of the Laplacian: $\mathsf{t} = \mathsf{n}^{\perp} \quad \rightarrow \quad \Delta \mathsf{x} = \mathsf{x}_{\mathsf{nn}} + \mathsf{x}_{\mathsf{tt}}$

  - Therefore: $\mathrm{div}(\mathsf{b}(|\nabla \mathsf{x}|)\ \nabla \mathsf{x}) = |\nabla \mathsf{x}|\ \mathsf{b}'(|\nabla \mathsf{x}|)\ \mathsf{x}_{\mathsf{nn}} + \mathsf{b}(|\nabla \mathsf{x}|)\ (\mathsf{x}_{\mathsf{nn}} + \mathsf{x}_{\mathsf{tt}})$

    - where: $\mathsf{b}'(|\nabla \mathsf{x}|) = \dfrac{\varphi''(|\nabla \mathsf{x}|)\ |\nabla \mathsf{x}| - \varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|^2}$

Reminder: $\mathsf{b}(|\nabla \mathsf{x}|) = \dfrac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}$



- **Finally:** $$\mathrm{div}\left(\frac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}\ \nabla \mathsf{x}\right) = \varphi''(|\nabla \mathsf{x}|)\ \mathsf{x}_{\mathsf{nn}} + \frac{\varphi'(|\nabla \mathsf{x}|)}{|\nabla \mathsf{x}|}\ \mathsf{x}_{\mathsf{tt}}$$

- **Iterative minimization scheme**

$$\frac{\partial x}{\partial \tau} = -\frac{\partial J_1}{\partial x}(x, y) + \mathrm{div}\left(\frac{\varphi'(|\nabla x|)}{|\nabla x|}\ \nabla x\right)$$

$$\mathrm{div}\left(\frac{\varphi'(|\nabla x|)}{|\nabla x|}\ \nabla x\right) = \varphi''(|\nabla x|)\ x_{nn} + \frac{\varphi'(|\nabla x|)}{|\nabla x|}\ x_{tt}$$

- **Regularization only**

$$\frac{\partial x}{\partial \tau} = \varphi''(|\nabla x|)\ x_{nn} + \frac{\varphi'(|\nabla x|)}{|\nabla x|}\ x_{tt}$$

- Tikhonov: $\varphi(u) = u^2$ ➜ $\varphi'(u) / u = 2$ et $\varphi''(u) = 2$
  - Geometrical heat equation: $\frac{\partial x}{\partial \tau} = \Delta x$
    - Isotropic smoothing
    - Edges are not preserved

- **Half-quadratic regularization**

$$\frac{\partial x}{\partial \tau} = \varphi''(|\nabla x|)\, x_{nn} + \frac{\varphi'(|\nabla x|)}{|\nabla x|}\, x_{tt}$$

  - Conditions on φ
    - from R to R+
    - even
    - φ'(u) ≥ 0 on R+
    - $\begin{cases} \lim\limits_{u \to 0^+} \dfrac{\varphi'(u)}{2u} = M \\ \lim\limits_{u \to +\infty} \dfrac{\varphi'(u)}{2u} = L \end{cases}$ L < M
  - In practice: M = 1, L = 0
  - Additional conditions

$$\begin{cases} \lim\limits_{u \to 0} \varphi''(u) = 2 \\ \lim\limits_{u \to +\infty} \varphi''(u) = 0 \\ \lim\limits_{u \to +\infty} \varphi''(u) \,/\, \frac{\varphi'(u)}{u} = 0 \end{cases}$$

Small gradient

|∇x| tends toward 0
Homogeneous area

Minimization equation = geometrical heat equation (Tikhonov)

$$\frac{\partial x}{\partial \tau} = \Delta x$$

Laplacian

Large gradient

|∇x| tends toward +∞
Edge

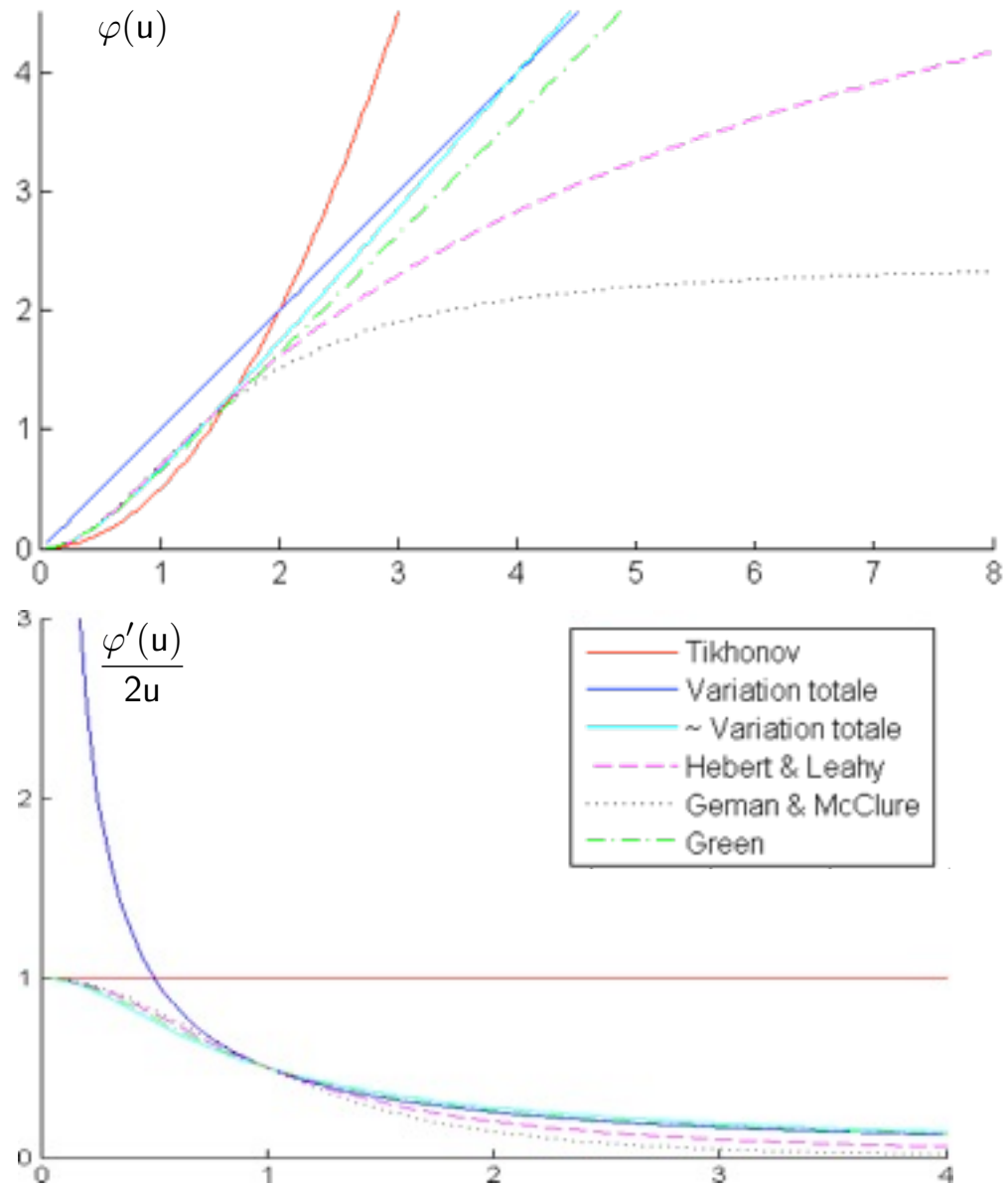
Minimization equation

$$\frac{\partial x}{\partial \tau} = \beta\, (\alpha\, x_{nn} + x_{tt}) \qquad \begin{cases} \alpha \to 0 \\ \beta \to 0 \end{cases}$$

Weighted Laplacian

Lower and anisotropic smoothing
Softer ⊥ to an edge
Preservation of edges

- **Unknown image:** x
- **Regularization**
  - Applied to the image gradient
    - $|\nabla x| = u$
- **Smooth solution**
  - Tikhonov: u^2
- **Piecewise smooth solution**
  - Total variation: $|u|$
  - Half-quadratic regularization
    - Total variation approximation: $\sqrt{(u^2 + \varepsilon^2)} - \varepsilon$
    - Hebert & Leahy: $\log(1 + u^2)$
    - Geman & McClure: $u^2 / (1 + u^2)$
    - Green: $\log(\cosh(u))$
    - …



- **Context: 1-D**
  - Minimization of $J(x) = \underbrace{(y-x)^2}_{\text{data fidelity}} + \underbrace{\lambda\,|x|}_{L_1-\text{regularization}}$ ..... with parameter $\lambda \geq 0$
    - Unregularized solution: $\lambda = 0 \Rightarrow \hat{x} = y$
    - Regularized solution: $\hat{x} = \arg\min_x J(x)$? (Note that $J(x)$ is strictly convex [sum of strictly convex functions])
  - Let us look for positive solutions: $\hat{x} > 0$
    - Minimization of $J(x) = (y-x)^2 + \lambda x$

$$J'(x) = -2\,(y-x) + \lambda$$

$$J'(\hat{x}) = 0 \Leftrightarrow \boxed{\hat{x} = y - \frac{\lambda}{2}} \text{ with the following necessary condition: } \boxed{y > \frac{\lambda}{2}}$$

    - Does it exist $\hat{x}_2 \leq 0$ such that $J(\hat{x}_2) \leq J\Big(y - \frac{\lambda}{2}\Big)$?

$$(y-\hat{x}_2)^2 - \lambda\,\hat{x}_2 \leq \lambda\,y - \frac{\lambda^2}{4}$$

$$\Leftrightarrow \hat{x}_2^2 - (2y+\lambda)\,\hat{x}_2 + \underbrace{y^2 - \lambda\,y + \frac{\lambda^2}{4}}_{\frac{1}{4}(2y-\lambda)^2} \leq 0$$

Discriminant: $\Delta = 8\lambda y$ ..... [remember that $y > 0$]

Therefore, the polynomial is less or equal to zero in $\Big[y + \frac{\lambda}{2} - \sqrt{2\lambda y},\, y + \frac{\lambda}{2} + \sqrt{2\lambda y}\Big]$

Such a solution $\hat{x}_2$ exists only if the lower bound $y + \frac{\lambda}{2} - \sqrt{2\lambda y}$ is less or equal to zero

$$
\begin{aligned}
& y + \frac{\lambda}{2} - \sqrt{2\lambda y} \leq 0 \\
\Leftrightarrow\ & y + \frac{\lambda}{2} \leq \sqrt{2\lambda y} \\
\Leftrightarrow\ & y^2 + \lambda y + \frac{\lambda^2}{4} \leq 2\lambda y \ \ldots\ldots \text{since both sides are positive} \\
\Leftrightarrow\ & y^2 - \lambda y + \frac{\lambda^2}{4} \leq 0 \\
\Leftrightarrow\ & \left(y - \frac{\lambda}{2}\right)^2 \leq 0 \ \ldots\ldots \text{contradiction; therefore, there is no negative } \hat{x}_2 \textit{ better} \text{ than } \hat{x}
\end{aligned}
$$

- Summary: if $\boxed{y > \frac{\lambda}{2}}$, then $\boxed{\arg\min_x J(x) = \hat{x} = y - \frac{\lambda}{2}}$

- What happens if $\boxed{0 \leq y \leq \frac{\lambda}{2}}$?
  - Note that $J(x = 0) = y^2$. Can we do better? $\exists ?\ \hat{x} \neq 0$ such that $J(\hat{x}) < y^2$

$$
\begin{aligned}
& (y - \hat{x})^2 + \lambda\,|\hat{x}| < y^2 \\
\Leftrightarrow\ & \hat{x}^2 - 2\hat{x}y + \lambda\,|\hat{x}| < 0
\end{aligned}
$$

    - Check for $\hat{x} > 0$:

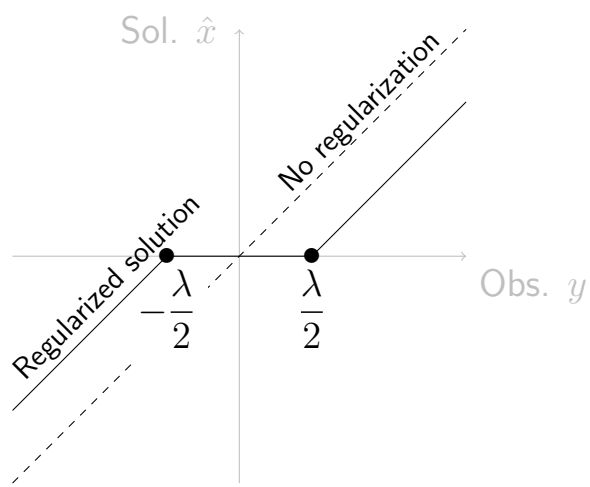
$$
\begin{aligned}
& \hat{x}^2 - (2y - \lambda)\,\hat{x} < 0 \\
\Leftrightarrow\ & \hat{x} - (2y - \lambda) < 0 \ \ldots\ldots \text{since } \hat{x} > 0 \\
\Leftrightarrow\ & 2y - \lambda > \hat{x} \\
\Rightarrow\ & 2y - \lambda > 0 \ \ldots\ldots \text{since } \hat{x} > 0 \\
\Leftrightarrow\ & y > \frac{\lambda}{2} \ \ldots\ldots \text{contradiction}
\end{aligned}
$$

- Check for $\hat{x} < 0$: $\hat{x}^2 - (2y + \lambda)\,\hat{x} < 0$
  $\Leftrightarrow \hat{x} - (2y + \lambda) > 0$ ..... since $\hat{x} < 0$
  $\Leftrightarrow \hat{x} > 2y + \lambda$ ..... contradiction since $\hat{x} < 0$ and $2y + \lambda \geq 0$

- Summary: if $\boxed{0 \leq y \leq \frac{\lambda}{2}}$, then $\boxed{\arg\min_x J(x) = \hat{x} = 0}$

- Extension to $\boxed{y < -\frac{\lambda}{2}}$ and $\boxed{-\frac{\lambda}{2} \leq y \leq 0}$
  - Note that $J_{-y}(-x) = J_y(x)$ [$J(x)$ should actually be written $J_y(x)$]

    Therefore, $\arg\min_x J_{-y}(-x) = \arg\min_x J_y(x)$
    Moreover, $\arg\min_x J_{-y}(-x) = -\arg\min_x J_{-y}(x)$
    Finally, $\arg\min_x J_{-y}(x) = -\arg\min_x J_y(x)$

  - Hence, the solutions for negative values of y are simply the opposite of the ones for positive values

- **Illustration**
  - Known as soft thresholding in the context of the wavelet transform (used for denoising)
  - Note that the unregularized solution is found by setting $\lambda = 0$