Eric Debreuve - debreuve@i3s.unice.fr

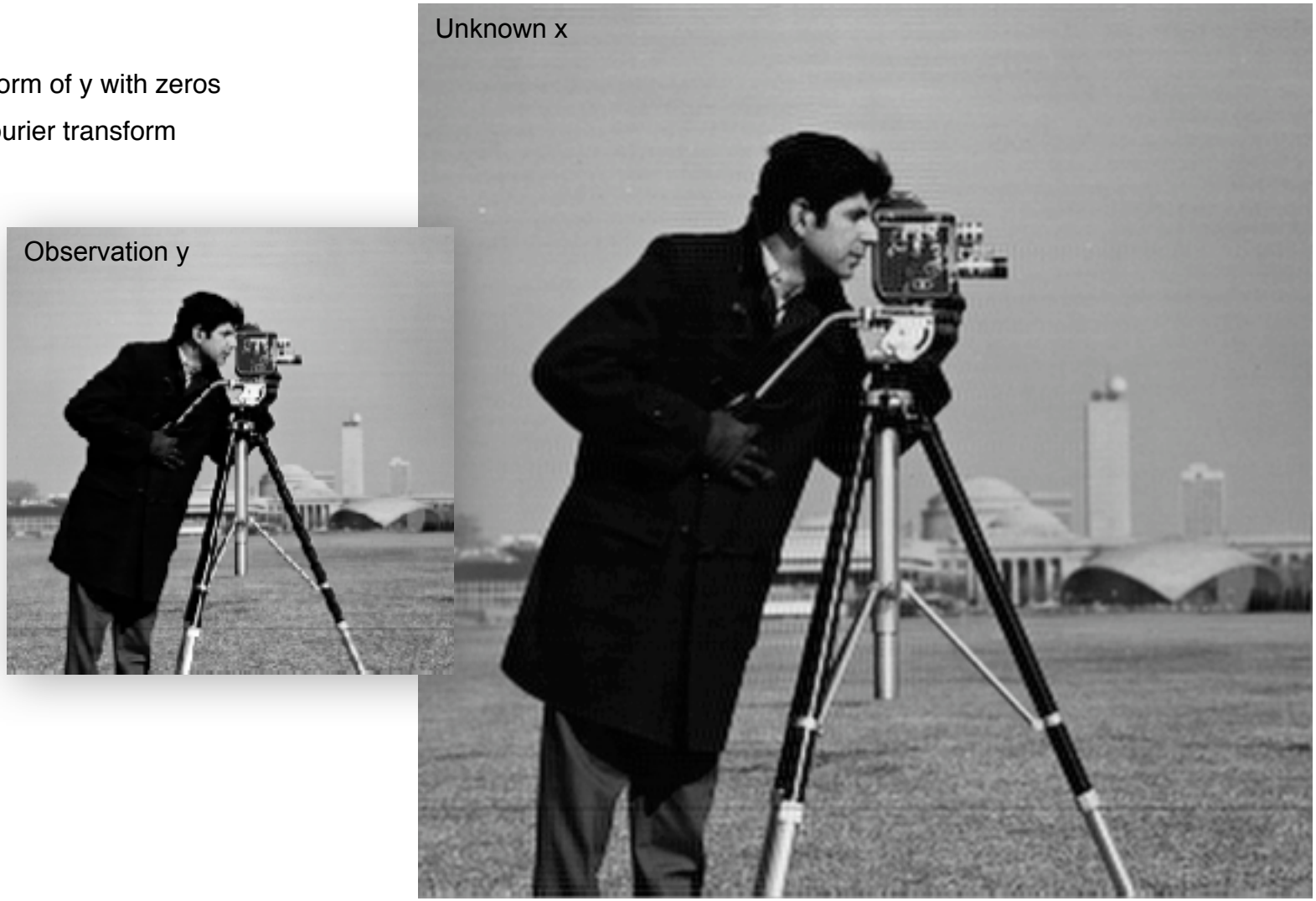
Methods of Inversion

- **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 $\text{observation} = \int_{\text{domain}} \mathsf{M}(\text{unknown})$

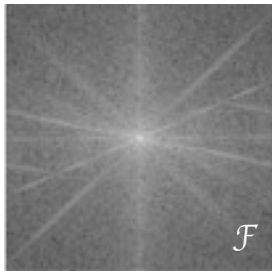
- **Stochastic**
  - The unknown is a realization of a random variable
  - Wiener filter
  - Random walk
  - Maximum likelihood / Maximum a posteriori
  - Maximum entropy / Minimum relative entropy

- **Problem**
  - Observation y: subsampled image
  - Source (unknown) x: original, high-resolution image

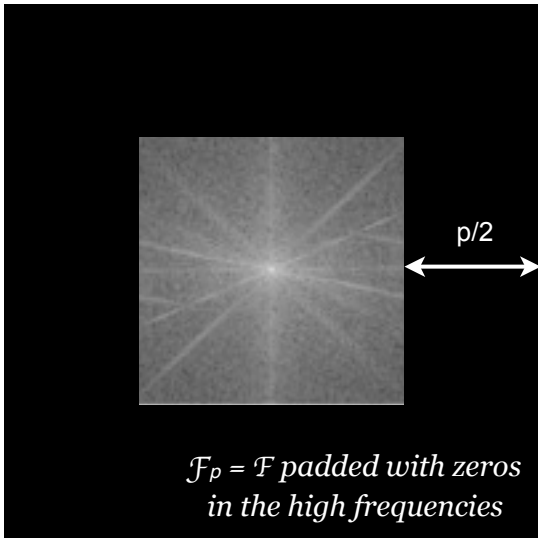
- **Principle**
  - Pad the Fourier transform of y with zeros
  - Perform an inverse Fourier transform



- **Fourier transform of the subsampled image**
  - p is even (see below)



Fourier transform
of the observation y



- **Inverse transform after padding** (in 1 dimension)
  - n even
  - Padding: p = L x n where L is an integer
  - $t \in [0..(n+p-1)]$ (padded)
  - $t' = t\, n / (n+p)$ (unpadded)

- **Model:** $y = M x + b$
  - M: r x c - matrix
    - Possibly overdetermined or underdetermined problem
  - Reformulation into a minimization problem
    - $\min_x |y - Mx|^2$
    - Convex

- **Least mean square solution**
  - Normal equations: $M^T M x = M^T y$ (*)
  - $r \geq c$ & rank of M = c
    - $x = \underbrace{(M^T M)^{-1} M^T}_{\text{Pseudo-inverse (left)}} y$

    - However, if M is badly conditioned
      - $M^T M$ is also badly conditioned
    - If r = c, M is invertible
      - Pseudo-inverse = M^{-1}

  - $r < c$. $M^T M$ is singular

    - Eq. (*) has an infinite number of solutions
    - Rank of M = r
      - Solution with minimal norm: $x = \underbrace{M^T (M M^T)^{-1}}_{\text{Pseudo-inverse (right)}} y$

- **Regularization**
  - $\min_x |y - Mx|^2 + \alpha |x|^2$
  - Least mean square solution
    - $x = \underbrace{(M^T M + \alpha I)^{-1}}_{\text{Better condition number}} M^T y$

- **Model:** $y = m(x) + b$ ➜ $\min_x |y - m(x)|^2$
  - Regularized problem: $\min J_1(x,y) + \alpha J_2(x)$
  - Euler-Lagrange equation (see Typical regularization in image processing)
    - $J'(x) = 0$ ➜ $\frac{\partial J_1}{\partial x}(x,y) - \mathrm{div}\left(\frac{\varphi'(|\nabla x|)}{|\nabla x|}\nabla x\right) = 0$
- **Closed-form solution**
  - Rarely possible in practice
- **Iterative minimization**
  - x is made dependent on a evolution parameter τ

$$\begin{cases} x(\tau = 0) = x_0 \\ \frac{\partial x}{\partial \tau} = -\frac{\partial J_1}{\partial x}(x,y) + \mathrm{div}\left(\frac{\varphi'(|\nabla x|)}{|\nabla x|}\nabla x\right) \end{cases}$$

  - Discretization: $x(\tau + d\tau) = x(\tau) + d\tau\left[-\frac{\partial J_1}{\partial x}(x(\tau),y) + \mathrm{div}\left(\frac{\varphi'(|\nabla x(\tau)|)}{|\nabla x(\tau)|}\nabla x(\tau)\right)\right]$
  - Convergence
    - Stationary point: $\partial x / \partial \tau = 0$
    - Discrete condition: $x(\tau+1) = x(\tau)$

- **Regularized energy**

$$J(x) = J_1(x, y) + \lambda \underbrace{\int \varphi(|\nabla x|)}_{J_2}$$

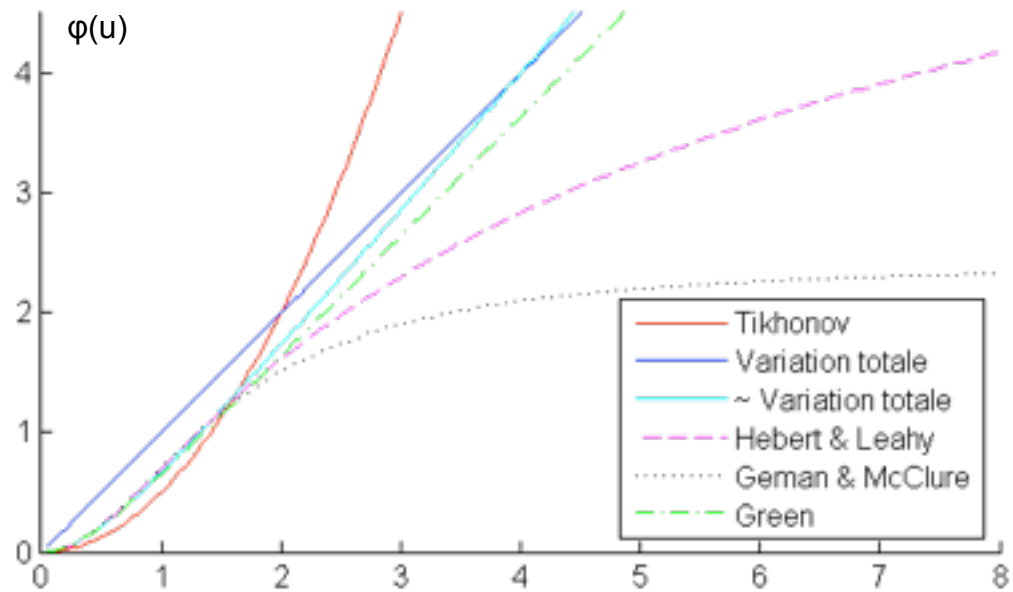
  - Conditions on φ
    - from R to R+
    - even
    - φ'(u) ≥ 0 on R+
    - $\begin{cases} \lim\limits_{u \to 0^+} \dfrac{\varphi'(u)}{2u} = M \\ \lim\limits_{u \to +\infty} \dfrac{\varphi'(u)}{2u} = L \end{cases}$ $\quad L < M$

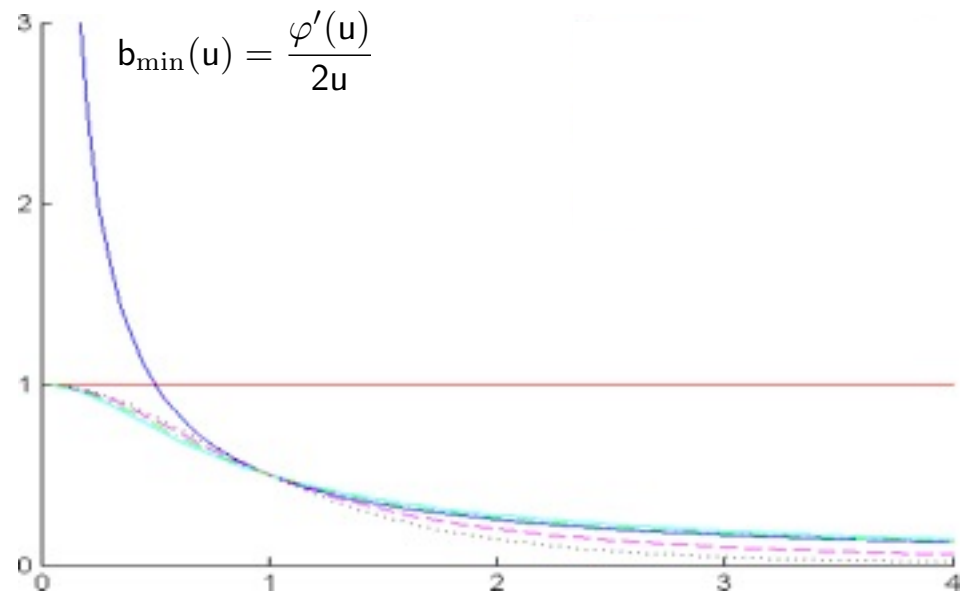
- **There exists a unique function ψ**
  - from [L,M] to R (reminder: usually, M=1 and L=0)
  - strictly convex
  - strictly increasing
  - and such that
  - $\varphi(u) = \min\limits_{b \in [L,M]} \{bu^2 + \Psi(b)\}$
  - $b_{\min}(u) = \dfrac{\varphi'(u)}{2u}$ (continuous representation of edges)

- **Regularized energy**

$$J(x) = J_1(x,y) + \lambda \int \left(b_{\min} |\nabla x|^2 + \Psi(b_{\min}) \right)$$

- **Representation of edges: b**
  - Belongs to [0,1]
  - If the gradient tends toward 0
    - Homogeneous area
    - b tends toward 1
  - If the gradient tends toward +∞
    - Edge
    - b tends toward 0
  - Non binary homogeneity/edge distinction





- **Regularized energy**

$$J(x) = J_1(x,y) + \lambda \int \left(b_{\min} |\nabla x|^2 + \Psi(b_{\min}) \right)$$

- **Interest**
  - φ is not quadratic in general
  - $\int b_{\min} |\nabla x|^2$ seen as a function of $|\nabla x|$ is quadratic if $b_{\min}$ is fixed
    - Linear normal equations

- **Algorithm**
  - Minimization w/r/t b with x constant (*): $b_{\min} = \dfrac{\varphi'(|\nabla x|)}{2|\nabla x|}$
  - Minimization w/r/t x with b constant: $\dfrac{\partial x}{\partial \tau} = -\dfrac{\partial J_1}{\partial x}(x,y) + \mathrm{div}(b_{\min} \nabla x)$ (iteration until convergence)
  - Go back to (*) unless convergence has been reached

Example

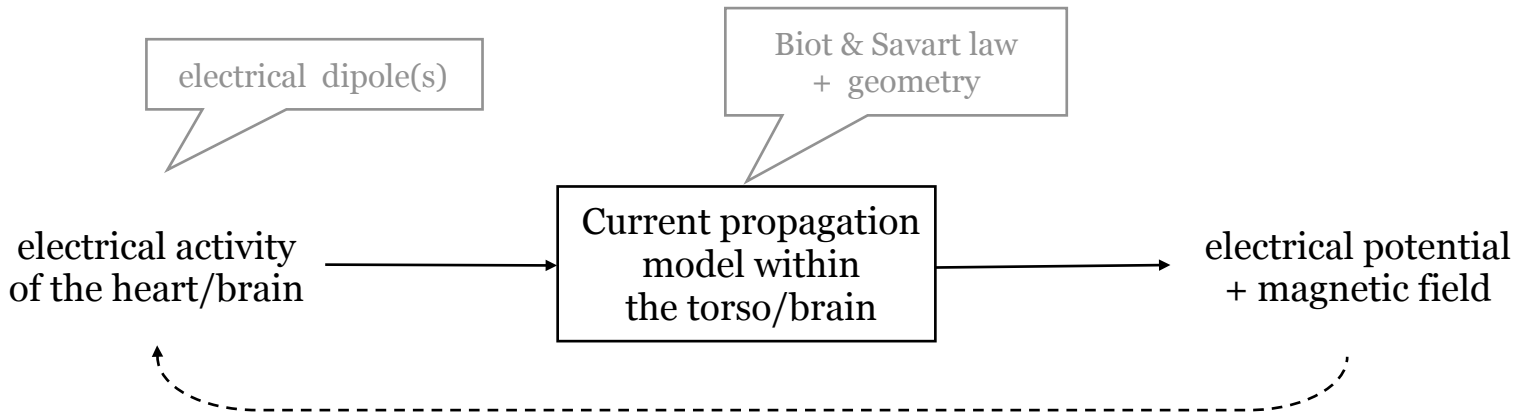
$$\begin{cases} J_1(x,y) = |y - Mx|^2 \\ \dfrac{\partial J_1}{\partial x}(x,y) = 2M^{\mathrm{T}}(Mx - y) \end{cases}$$

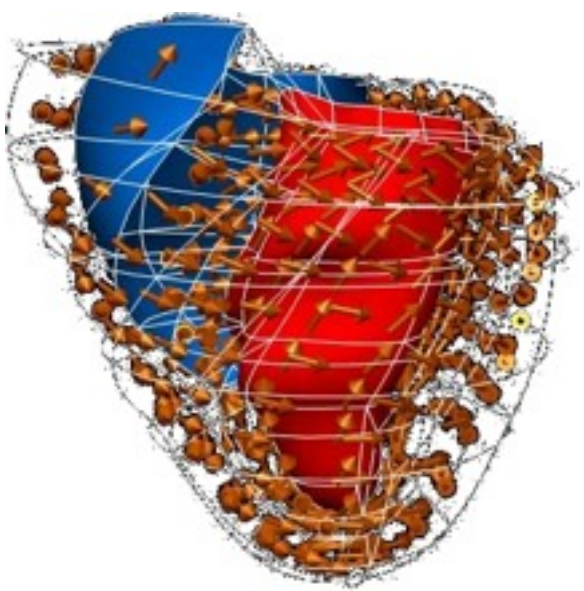
Lecture: Methods of inversion

- **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 $\text{observation} = \int_{\text{domain}} \mathsf{M}(\text{unknown})$

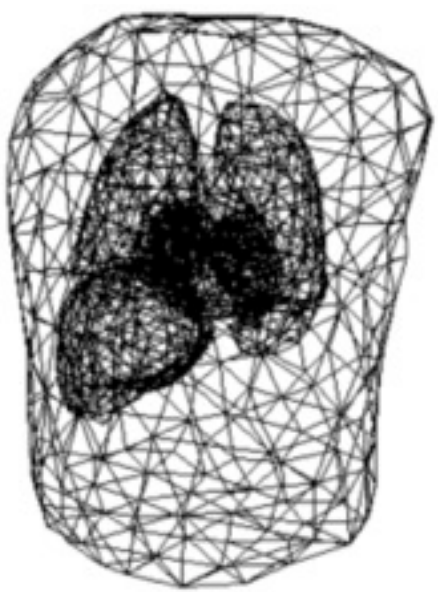
- **Stochastic**
  - The unknown is a realization of a random variable
  - Wiener filter
  - Random walk
  - Maximum likelihood / Maximum a posteriori
  - Maximum entropy / Minimum relative entropy

- **Model**





Bioengineering Research Group,
The University of Auckland, New Zealand



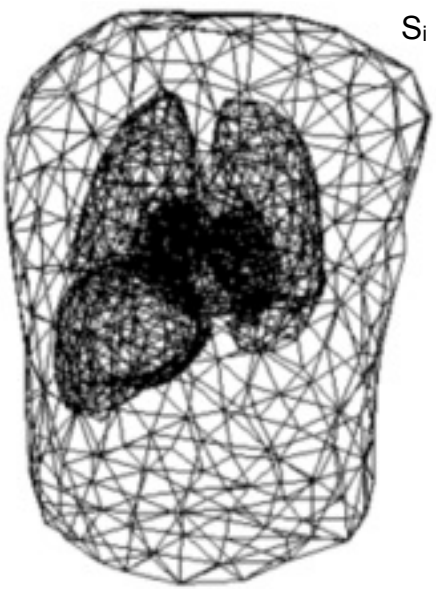
Conductivity model: set of regions
with constant conductivity

Observed electrical potential

- **Computing the electrical potential**
  - Locally at r on S_j

$$v_\infty(r) = \frac{\sigma_j^- + \sigma_j^+}{2}\, v(r) + \frac{1}{4\pi}\sum_{i=1}^{m}(\sigma_i^- - \sigma_i^+)\int_{S_i} v(p)\; n_i(p)\cdot\frac{r-p}{|r-p|^3}\; dp$$

    - Notation: $v_\infty(r) = L(v)(r)$



  - Global residual (over all the surfaces): $\displaystyle\sum_{i=1}^{m}\int_{S_i}\left|v_\infty(p) - L(v)(p)\right|\; dp = 0$

  - Weak formulation
    - Weighting (or test) function w
    - Weighted residual: $\displaystyle\sum_{i=1}^{m}\int_{S_i}\big(v_\infty(p) - L(v)(p)\big)\; w(p)\; dp = 0$
      - Inner product: $\displaystyle\langle f, w\rangle = \sum_{i=1}^{m}\int_{S_i} f(p)w(p)\; dp \quad\Longrightarrow\quad \langle v_\infty, w\rangle = \langle L(v), w\rangle$

- **Weighting functions w**
  - Restricted to functions generated by a basis: Ψ_n, $n \in [1..N]$ (defined on S_i: $R^2 \rightarrow R$)

$$w = \sum_{n=1}^{N} \alpha_n \Psi_n$$

$$\langle v_\infty, w \rangle = \langle L(v), w \rangle \quad \Longleftrightarrow \quad \langle v_\infty, \Psi_n \rangle = \langle L(v), \Psi_n \rangle$$

- **Potential v**
  - Resolution on a basis: Φ_n, $n \in [1..N]$
    - Nodal points or nodes: r_n, $n \in [1..N]$
    - Condition: $\Phi_n(r_i) = \begin{cases} 1 \text{ if } i = n \\ 0 \text{ if } i \neq n \end{cases}$ (does not mean "everywhere but at r_i" !)
  - Potential: $v = \sum_{n=1}^{N} v_n \Phi_n$
  - Unknown: v ➜ nodal parameters v_n

- **For one element of the weighting basis**

$$\begin{cases} \langle v_\infty, \Psi_n \rangle = \langle L(v), \Psi_n \rangle \\ v = \sum_{n=1}^{N} v_n \Phi_n \end{cases}$$

$$\langle v_\infty, \Psi_n \rangle = \sum_{k=1}^{N} v_k \, \langle L(\Phi_k), \Psi_n \rangle$$

$$\langle v_\infty, \Psi_n \rangle = \begin{bmatrix} \langle L(\Phi_1), \Psi_n \rangle & \langle L(\Phi_2), \Psi_n \rangle & \dots & \langle L(\Phi_N), \Psi_n \rangle \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$$

- **For the whole basis**

$$\begin{bmatrix} \langle v_\infty, \Psi_1 \rangle \\ \langle v_\infty, \Psi_2 \rangle \\ \vdots \\ \langle v_\infty, \Psi_N \rangle \end{bmatrix} = \begin{bmatrix} \langle L(\Phi_1), \Psi_1 \rangle & \langle L(\Phi_2), \Psi_1 \rangle & \dots & \langle L(\Phi_N), \Psi_1 \rangle \\ \langle L(\Phi_1), \Psi_2 \rangle & \langle L(\Phi_2), \Psi_2 \rangle & \dots & \langle L(\Phi_N), \Psi_2 \rangle \\ \vdots & \vdots & \dots & \vdots \\ \langle L(\Phi_1), \Psi_N \rangle & \langle L(\Phi_2), \Psi_N \rangle & \dots & \langle L(\Phi_N), \Psi_N \rangle \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$$

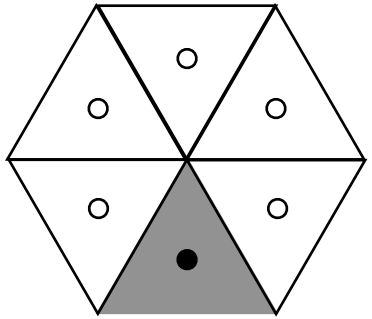
- **Collocation**
  - Weighting basis: characteristic functions of the nodal points

$$\Psi_n(r) = \begin{cases} 1 & \text{if } r = r_n \\ 0 & \text{otherwise} \end{cases}$$

- **Galerkin**
  - Weighting basis = resolution basis: $\Psi_n = \phi_n$

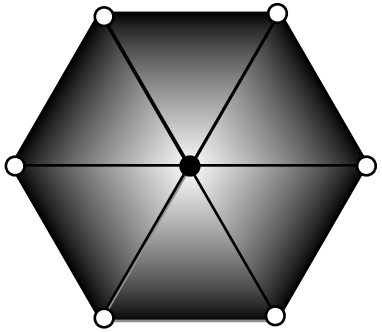
- **Resolution basis**
  - Constant:
    - N triangles
    - Nodal points = center of mass of the triangles

$$\Phi_n(r) = \begin{cases} 1 & \text{if } r \text{ belongs to the } n^{th} \text{ triangle} \\ 0 & \text{otherwise} \end{cases}$$



  - Linear:
    - N vertices
    - Nodal points = vertices

$$\Phi_n(r) = \begin{cases} \dfrac{r \cdot (s_n \times t_n)}{r_n \cdot (s_n \times t_n)} & \text{if } r \text{ belongs to the triangle } (r_n, s_n, t_n) \\ 0 & \text{otherwise} \end{cases}$$



Lecture: Methods of inversion

- **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

- **Convolution:** $y = m * x + b$
  - Least square solution: $\min_x J(x) = \min_x |y - m * x|^2 = \min_x \langle y - m * x, y - m * x \rangle$
    - Definition: $\langle f, g \rangle = \int f(t)\ g(t)\ dt$
  - Derivative: $J(x+h) = J(x) - 2\langle h, \underbrace{n * y - (n * m) * x}_{-\frac{1}{2}\frac{\partial J}{\partial x}} \rangle + |m * h|^2$
    - where n(t) = m(-t)
  - Solution: x such that $\frac{\partial J}{\partial x} = 0$

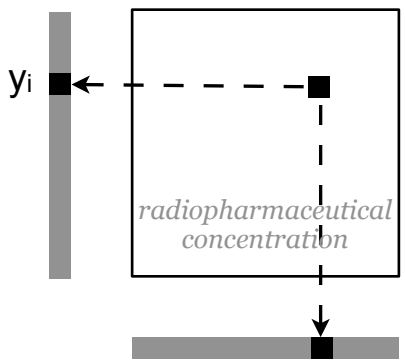
$$n * y = (n * m) * x$$

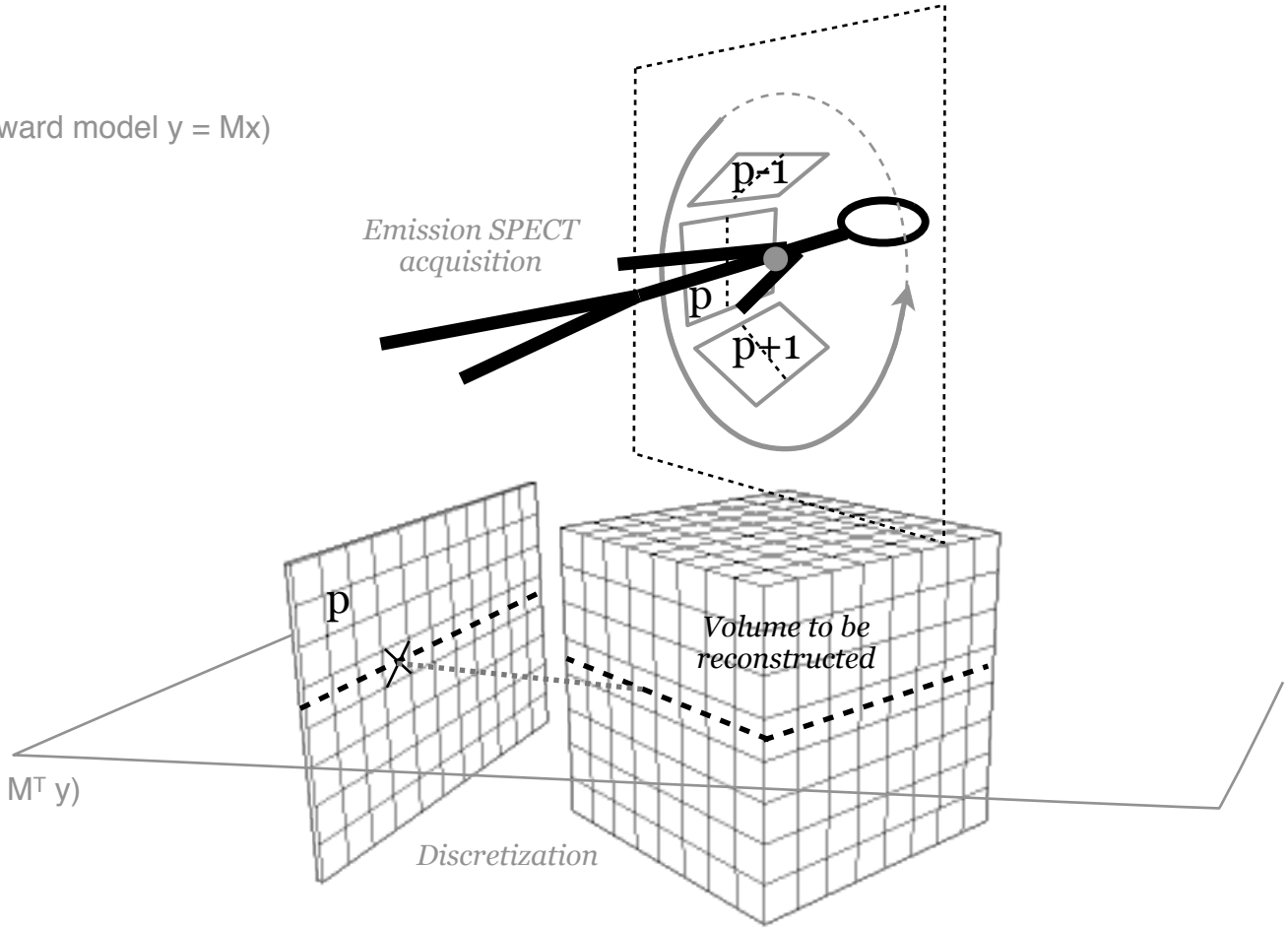
$$\hat{M}^* \ \hat{Y} = (\hat{M}^* \ \hat{M})\ \hat{X} = |\hat{M}|^2\ \hat{X}$$

    - $\hat{X} = \frac{\hat{Y}}{\hat{M}}$: not stable (see slide on stability)

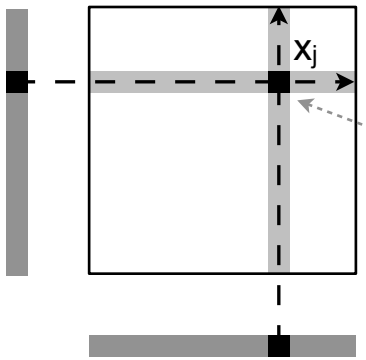
- Regularization: $J(x) = |y - m * x|^2 + \lambda\ |d * x|^2$
  - Derivative: $\frac{1}{2}\frac{\partial J}{\partial x} = [(n * m) + \lambda\ (e * d)] * x - n * y$
    - where e(t) = d(-t)
  - Fourier: $\left[|\hat{M}|^2 + \lambda\ |\hat{D}|^2\right]\ \hat{X} = \hat{M}^*\ \hat{Y} \quad \Leftrightarrow \quad \hat{X} = \frac{|\hat{M}|^2}{|\hat{M}|^2 + \lambda\ |\hat{D}|^2} \frac{\hat{Y}}{\hat{M}}$

- **Emission**
  - Projection: $y_i = \sum_j m_{i,j}\ x_j$ (forward model y = Mx)

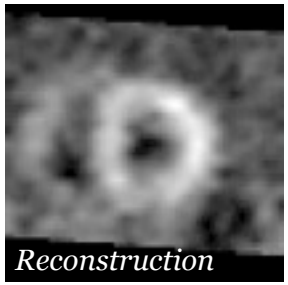




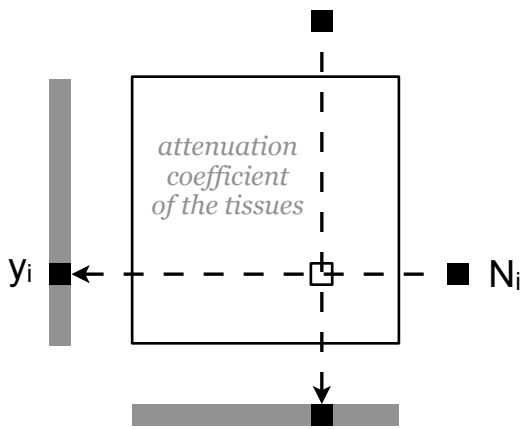
  - Backprojection: $x_j = \sum_i m_{i,j}\ y_i$ ($x = M^T y$)

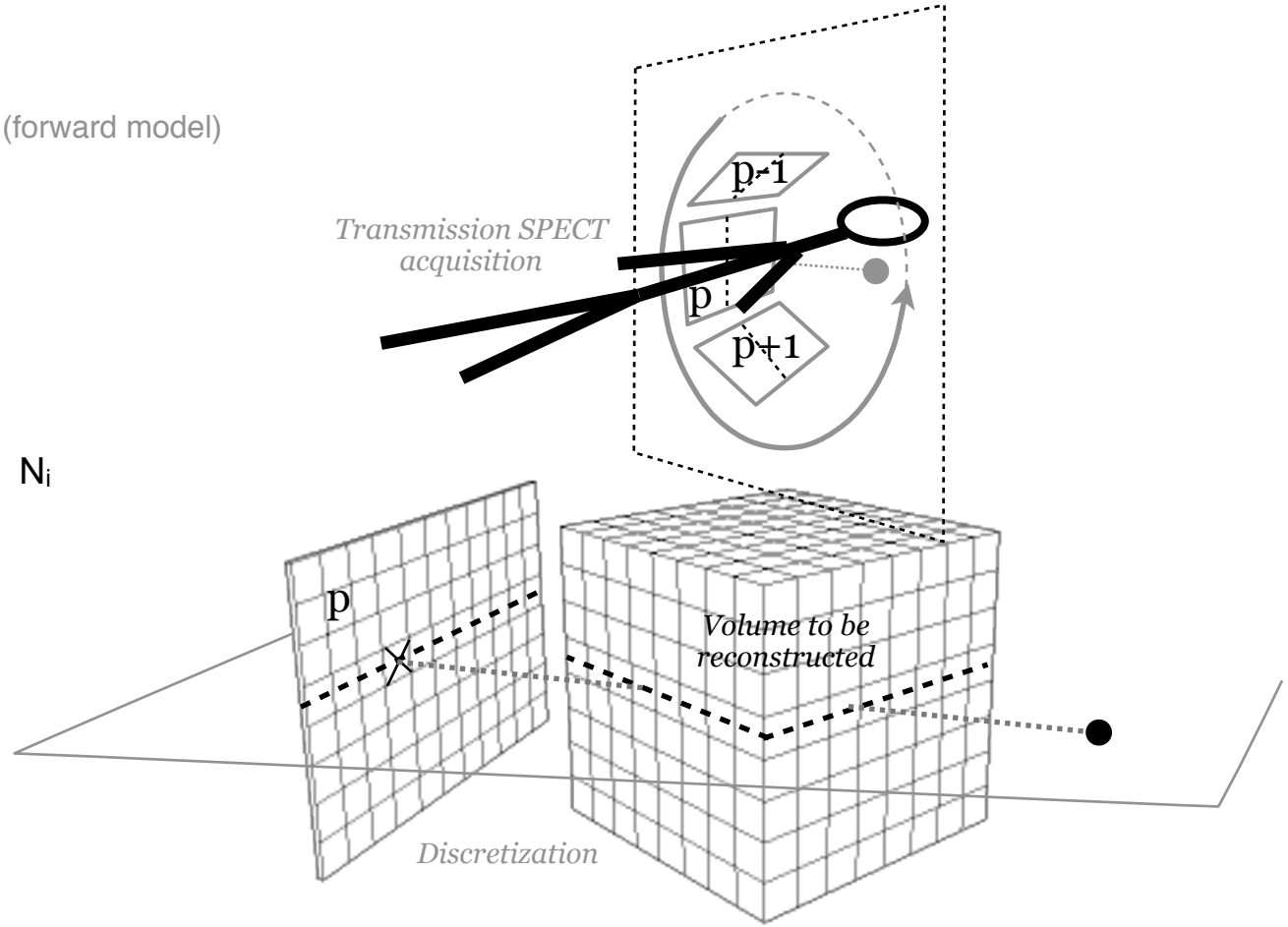


Sum of all the measurements (i.e., projections) that received a contribution from x_j

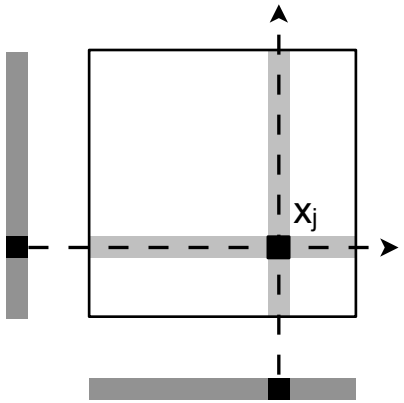


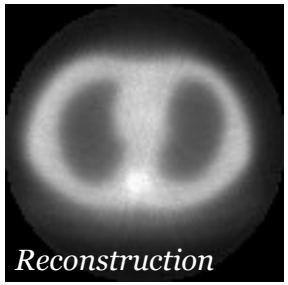
- **Transmission**
  - Projection: $y_i = N_i \, e^{-\sum_j l_{i,j} \, x_j}$ (forward model)





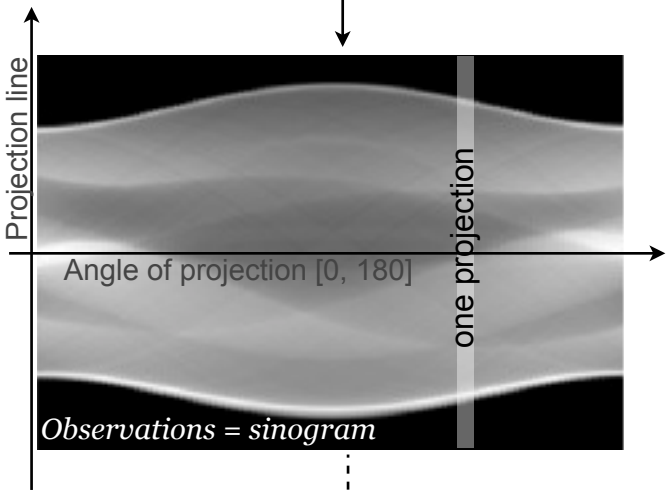
  - Backprojection: $x_j = \sum_i l_{i,j} \, \log \frac{N_i}{y_i}$



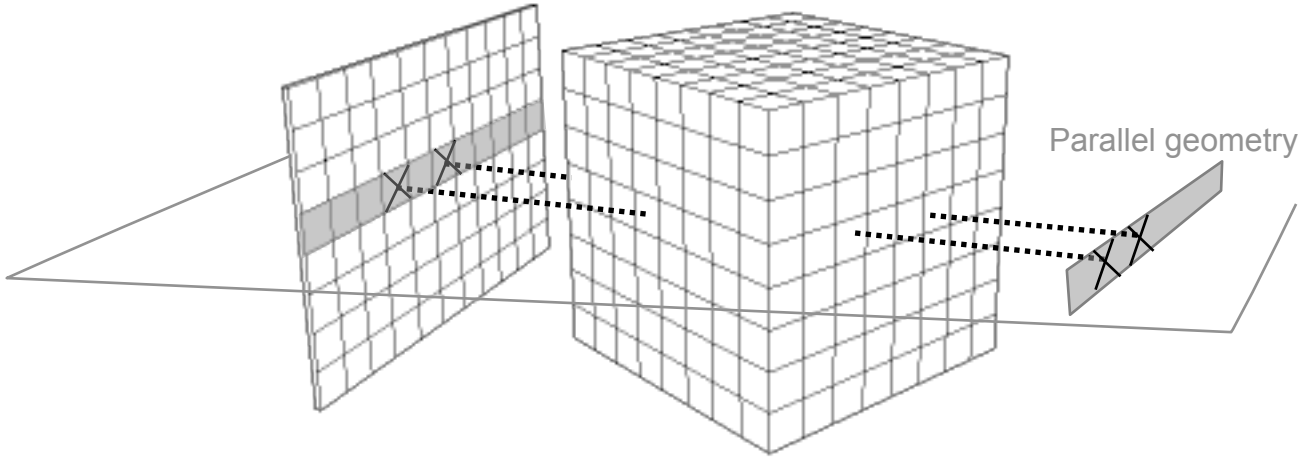


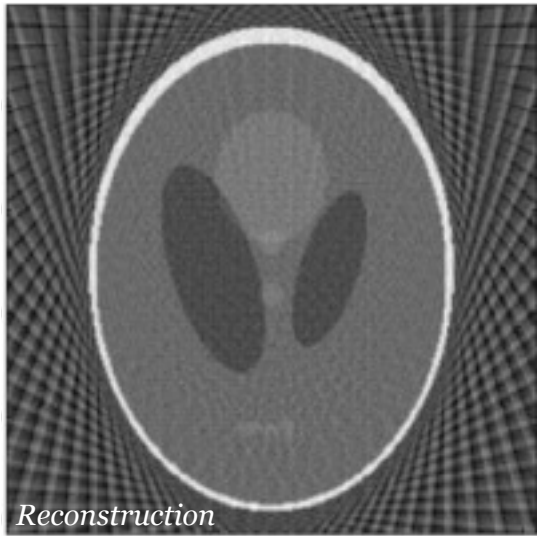
- **Shepp-Logan phantom**
  - Synthetic image *representing* the brain

Radon transform

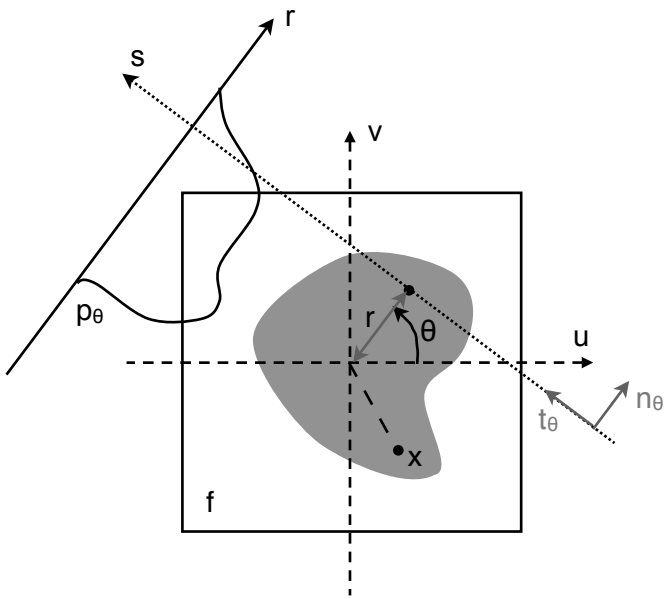


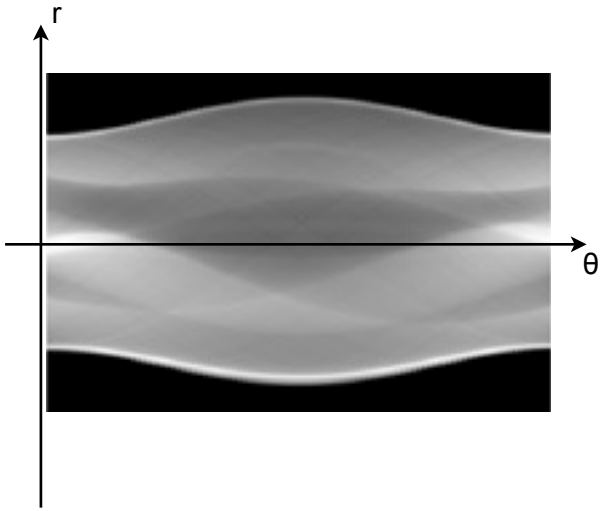
- **Projection** (transmission)
  - Radon transform ➜ set of projections, also called sinogram





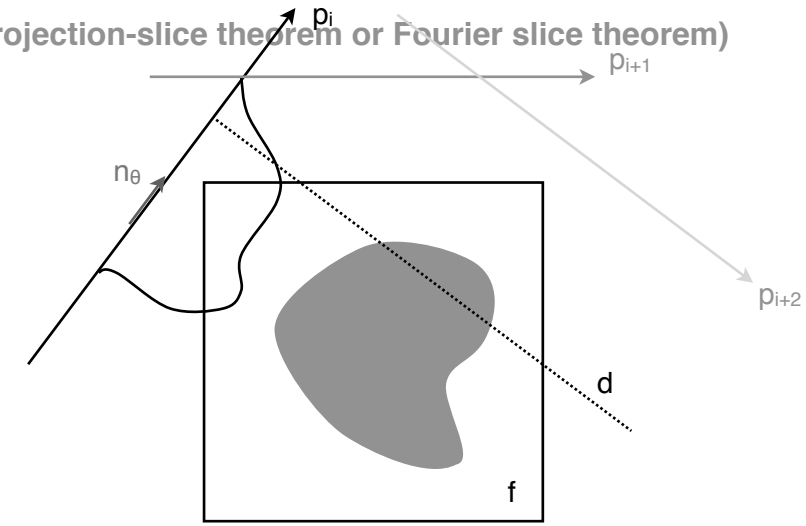
- **Notations**
  - $n_\theta = (\cos\theta, \sin\theta)$
  - $t_\theta = (-\sin\theta, \cos\theta)$

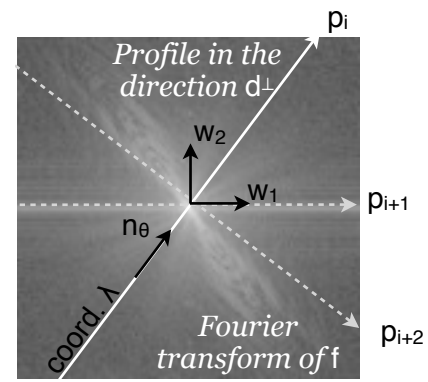




- **Definition** (continuous)
  - Projection: $$p_\theta(r) = \int_R f(s\ t_\theta + r\ n_\theta)\ ds = \int_{R^2} f(x)\ \delta(r - x \cdot n_\theta)\ ds$$ (Radon transform)
  - Backprojection: $$\tilde{f}(x) = \int_0^\pi p_\theta(n_\theta \cdot x)\ d\theta$$ (dual transform)

- **Projection theorem** (or projection-slice theorem or Fourier slice theorem)



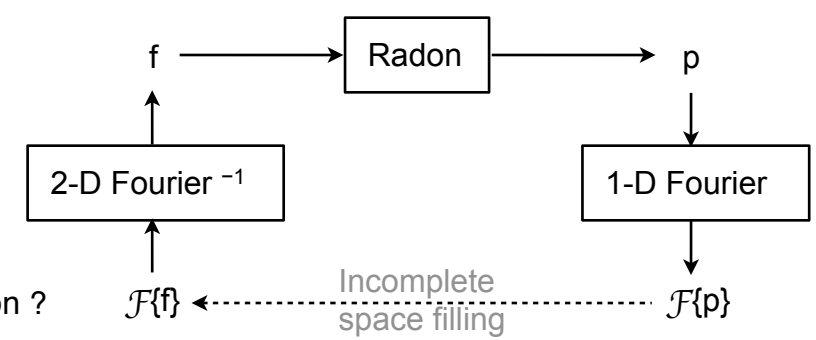


- **Application to the Radon transform**
  - Fourier transform of the projection:
    - with $w = (w_1, w_2) = \lambda n_\theta$

$$\mathcal{F}\{p_\theta\}(\lambda) = \int_R p_\theta(r)\, e^{-2i\pi\lambda r}\, dr = \int_{R^2} f(x)\, e^{-2i\pi w\cdot x}\, dx$$

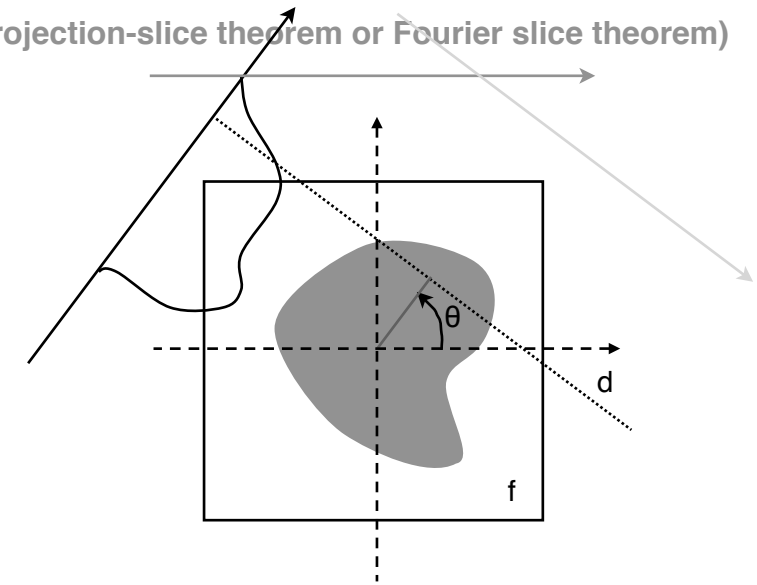
  - Reconstruction: f is the inverse transform of $\mathcal{F}\{f\}$:

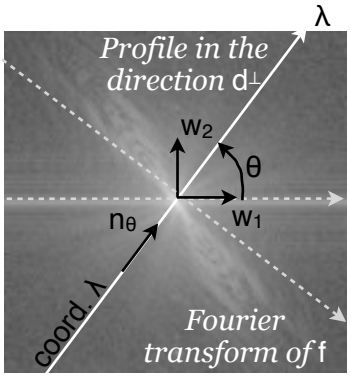
$$f(x) = \int_{R^2} \mathcal{F}\{f\}(w)\, e^{2i\pi w\cdot x}\, dw$$



- Finite number of projections ➜ interpolation ?

- **Projection theorem** (or projection-slice theorem or Fourier slice theorem)





- **Application to the Radon transform**
  - Fourier transform of the projection:
    - with $w = (w_1, w_2) = \lambda n_\theta$

$$\mathcal{F}\{p_\theta\}(\lambda) = \int_R p_\theta(r)\ e^{-2i\pi\lambda r}\ dr = \int_{R^2} f(x)\ e^{-2i\pi w\cdot x}\ dx$$

  - Reconstruction: f is the inverse transform of $\mathcal{F}\{f\}$:

$$f(x) = \int_{R^2} \mathcal{F}\{f\}(w)\ e^{2i\pi w\cdot x}\ dw$$

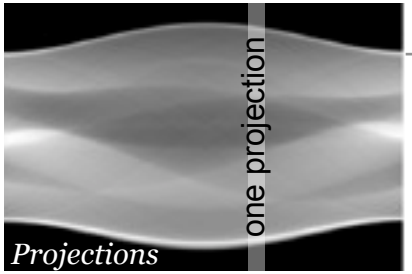
$$= \cdots$$

$$= \int_{\theta=0}^{\pi} (p_\theta * \mathcal{R})(n_\theta \cdot x)\ d\theta \qquad \text{(backprojection)}$$

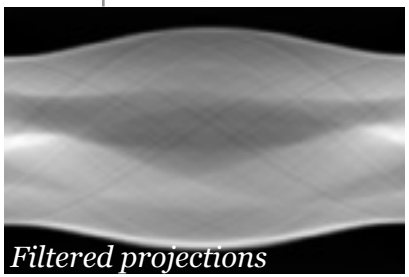
    - with $\mathcal{R} = \mathcal{F}^{-1}\{\ |\lambda|\ \}$
    - Convolution and integral can be discretized

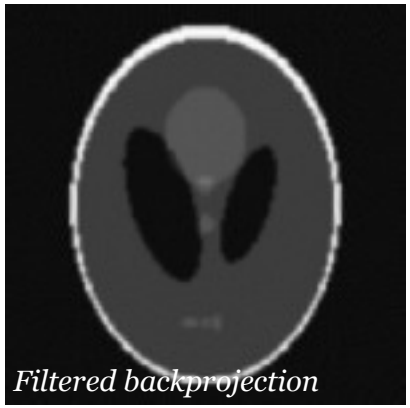
- **Unknown source**

- **Acquisition**

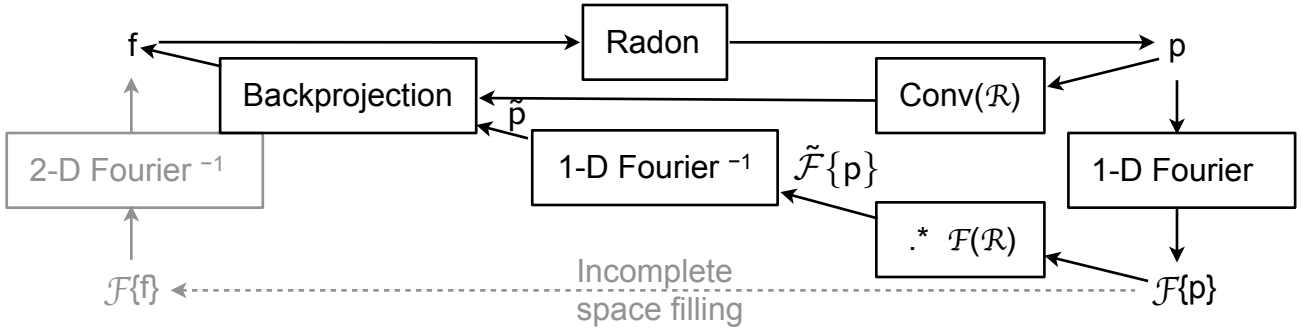


- **Algorithm**
  - 1-D convolution of each projection by $\mathcal{R}$ ……..

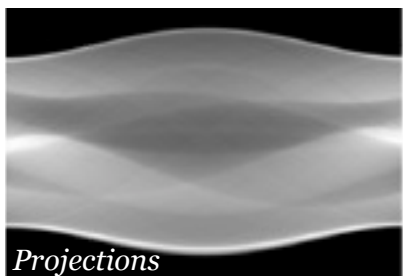


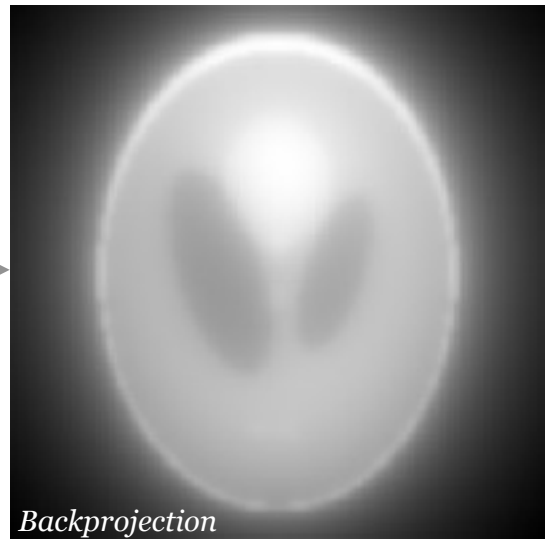


  - Backprojection of the filtered projections ……..

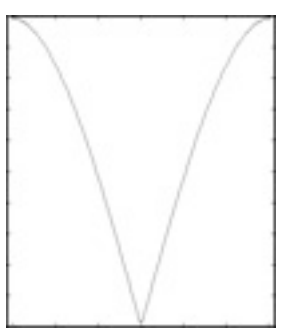


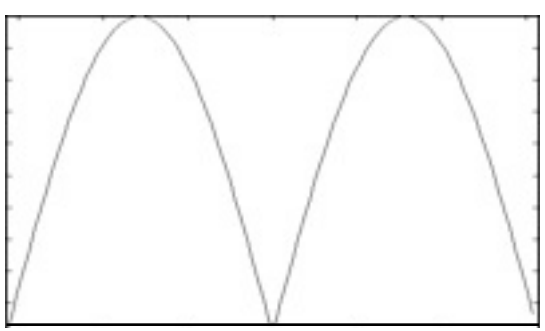
- **Backprojection without filtering**
  - Not a reconstruction





- **Filter $\mathcal{R}$**
  - 1-D Fourier: absolute value function = ramp filter
    - Amplification of high frequencies
      - Necessary (see backprojection)
      - However, amplification of the noise
  - Truncated, smoothed ramp





Algebraic vs. analytical: Brief comparison

- **Algebraic**
  - Specific discrete forward model M / Generic inversion
  - Complex phenomena (e.g., in nuclear medicine imaging, projections degrading factors (attenuation, Compton scattering, loss of resolution with the depth…)) can be modeled in M; the inversion automatically applies the necessary corrections
  - Possibility to introduce regularization
  - Computationally intensive: one iteration of the inversion ⇔ one inversion using an analytical method

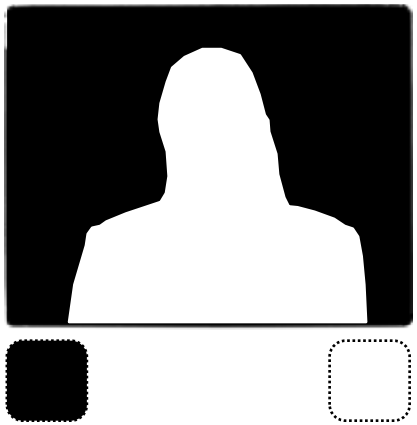
- **Analytical**
  - Rigorous, problem-specific development
  - Dealing with complex phenomena
    - Continuous model
    - Combination with a first-approximation model
    - Analytical inversion of the global formula

    Difficult

    or
    - Pre- and/or post-corrections (not necessarily optimal)
  - Regularization difficult to introduce

- **Segmentation**
  - Image or video partitioning. Ex. objects of interest and background

4 types of region



Object/background

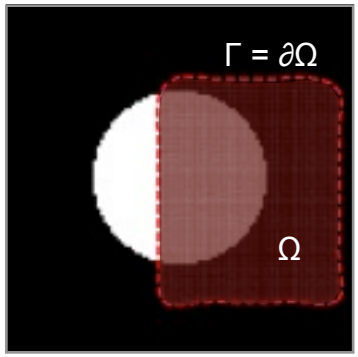
- **Interpretation as an inverse problem**

- Background image
- Object texture or average color
- Object shape

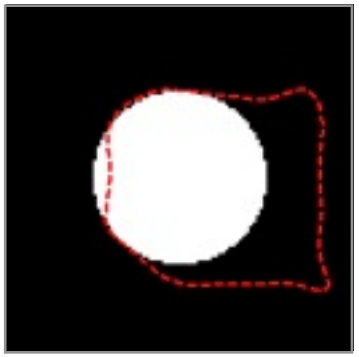
→ Compositing → Observed image

Segmentation

- **Solution using active contours**



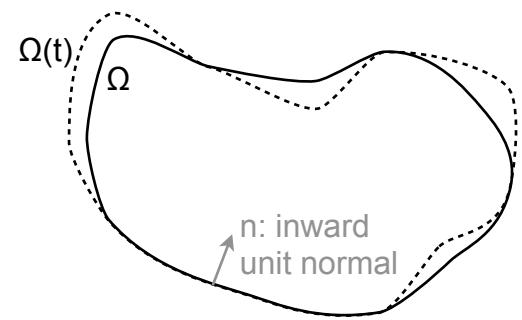
Initial contour



... Iterative deformation based on a contour evolution equation ...

Segmentation

- **Notations**
  - Ω: domain, $\Gamma = \partial\Omega$: contour of Ω
  - $\phi(\Omega, x)$ or $\phi(\Gamma, x)$: function of a domain or a contour, and a position
  - V: smooth vector field (velocity of deformation)
  - T_t: mapping $T_t(x) = X(t)$ such that $\begin{cases} X(0) = x \\ X'(t) = V(X(t)) \end{cases}$ (transforms Ω into $\Omega(t)$)
  - Material derivative: $\dot{\phi}(\Omega, V) = \lim_{t \to 0} \frac{1}{t}(\phi(t) - \phi(0))$ (also called Lagrangian derivative)
  - Shape derivative: $\phi'(\Omega, V) = \dot{\phi}(\Omega, V) - \nabla\phi \cdot V$



- **Segmentation as the minimizer of a…**
  - Contour-based energy: $E(\Gamma) = \int_\Gamma \phi(\Gamma, \Gamma(s))\, ds$ with $\phi(\Gamma) = \phi(\Omega)\big|_\Gamma$
    - Derivative: $dE(\Gamma, V) = \int_\Gamma \phi'\big|_\Gamma\, ds - \int_\Gamma \left(\phi\kappa - \frac{\partial\phi}{\partial n}\right) n \cdot V\, ds$ $\quad dE(\Gamma, V) = \lim_{t \to 0} \frac{1}{t}\big(E(\underbrace{\Gamma(t)}_{\textit{Depends on } V}) - E(\Gamma)\big)$
  - Region-based energy: $E(\Omega) = \int_\Omega \phi(\Omega, x)\, dx$
    - Derivative: $dE(\Omega, V) = \int_\Omega \phi'\, dx - \int_\Gamma \phi\, n \cdot V\, ds$

- **Region-based energy**
  - Classical example: $E(\Omega) = \int_\Omega |f(x) - \mu(\Omega)|^2 \, dx$ with $\mu(\Omega) = \frac{1}{|\Omega|} \int_\Omega f(x) \, dx$
    - Homogeneous object
  - Derivative: $dE(\Omega, V) = \cancel{\int_\Omega \phi' \, dx} - \int_\Gamma \phi \, n \cdot V \, ds = \int_\Gamma (-\phi \, n) \cdot V \, ds = \langle \underbrace{-\phi \, n}_{Gradient}, V \rangle$

    Classical L^2 inner product on Γ
- **Classical evolution**
  - In the direction opposite to the gradient
  - Evolution equation: $\frac{\partial \Gamma}{\partial \tau} = \phi \, n$ (discrete version(*): $\Gamma(\tau+1) = \Gamma(\tau) + d\tau \, \Phi(\tau) \, n(\tau)$)
    - No shape constraint
- **Introducing a shape constraint: regularization**
  - Ex. the contour must be a circle

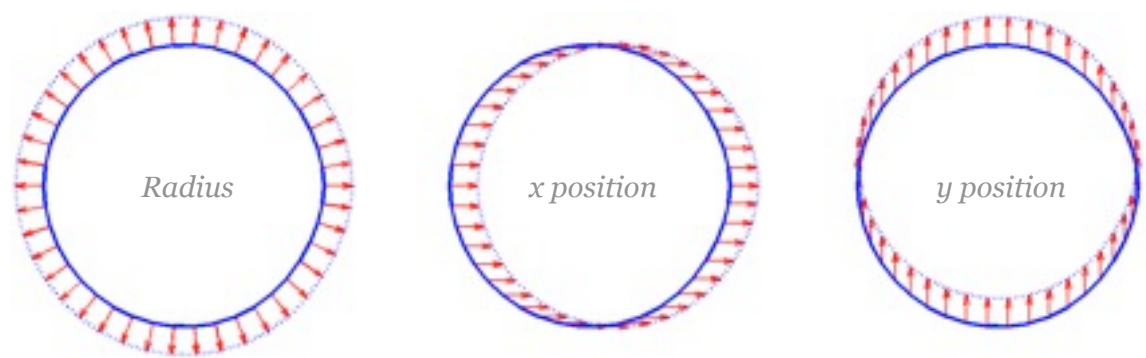
  1. Projection
     - After each iteration of evolution (*), replace Γ with the best circular fit
     - Convergence?
  2. Constraint
     - Minimize: $E_{\text{circle}}(\Omega) = E(\Omega) + \lambda \underbrace{\mathcal{D}(\Omega, \text{Circle}(\Omega))}_{\text{“Distance” between } \Omega \text{ and the best circular fit}}$
     - Not a circle
  3. <u>Parameterization</u>

- **Interpretation of the derivative**
  - Let V be a vector field defined on Γ

$$dE(\Gamma, V) = \int_\Gamma \phi'|_\Gamma \, ds - \int_\Gamma \left(\phi\kappa - \frac{\partial\phi}{\partial n}\right) n \cdot V \, ds \qquad dE(\Omega, V) = \int_\Omega \phi' \, dx - \int_\Gamma \phi \, n \cdot V \, ds$$

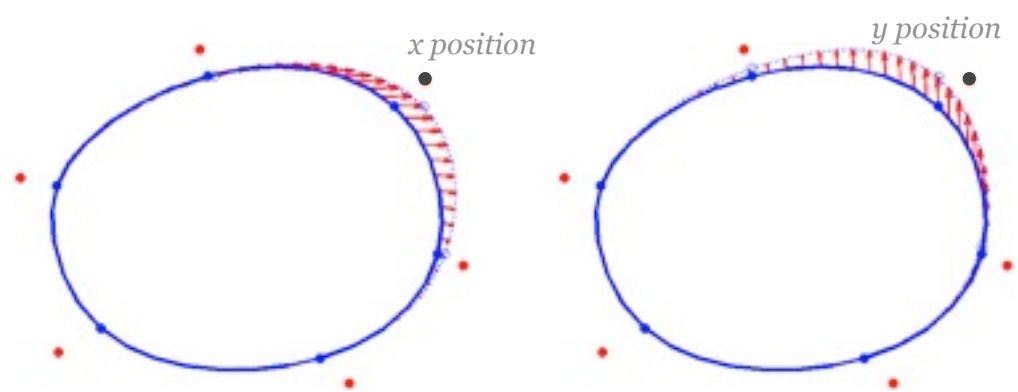
    - The sign of dE(Ω,V) tells us whether the contour deformation will increase or decrease the energy

- **Γ is a parametric curve**
  - Parameters p in $\mathbb{R}^n$
    - Energy: E(Ω or Γ) ➜ E(p)



  - Admissible velocities: $\dfrac{\partial\Gamma}{\partial p_i} = V_i$
    - Ex. circle / spline with 6 control points
  - Derivative: $\dfrac{\partial E}{\partial p_i} = dE\left(\Omega(p), \dfrac{\partial\Gamma}{\partial p_i}\right)$

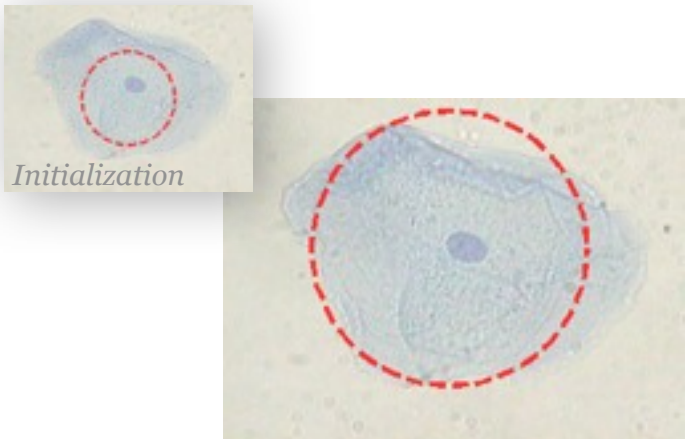
    Derivative from free-form (i.e., unconstrained) deformation



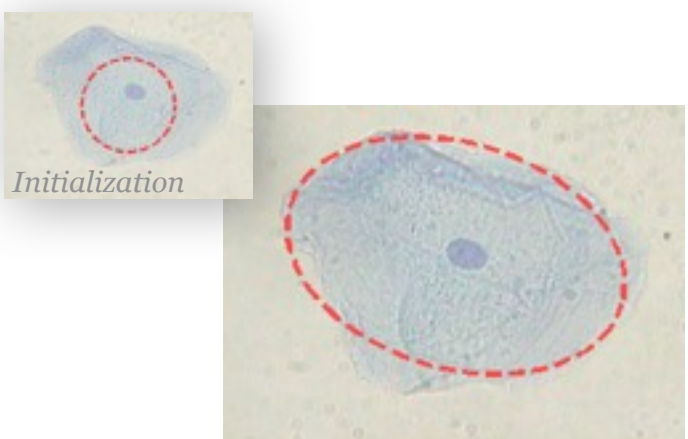
  - Segmentation ➜ minimization in $\mathbb{R}^n$
    - Ex.

$$\begin{cases} p^0 \\ p^{+1} = p - \alpha \, \nabla_p E(p) \\ \alpha = \arg\min\limits_{a \geq 0} E(p - a \, \nabla_p E(p)) \end{cases}$$

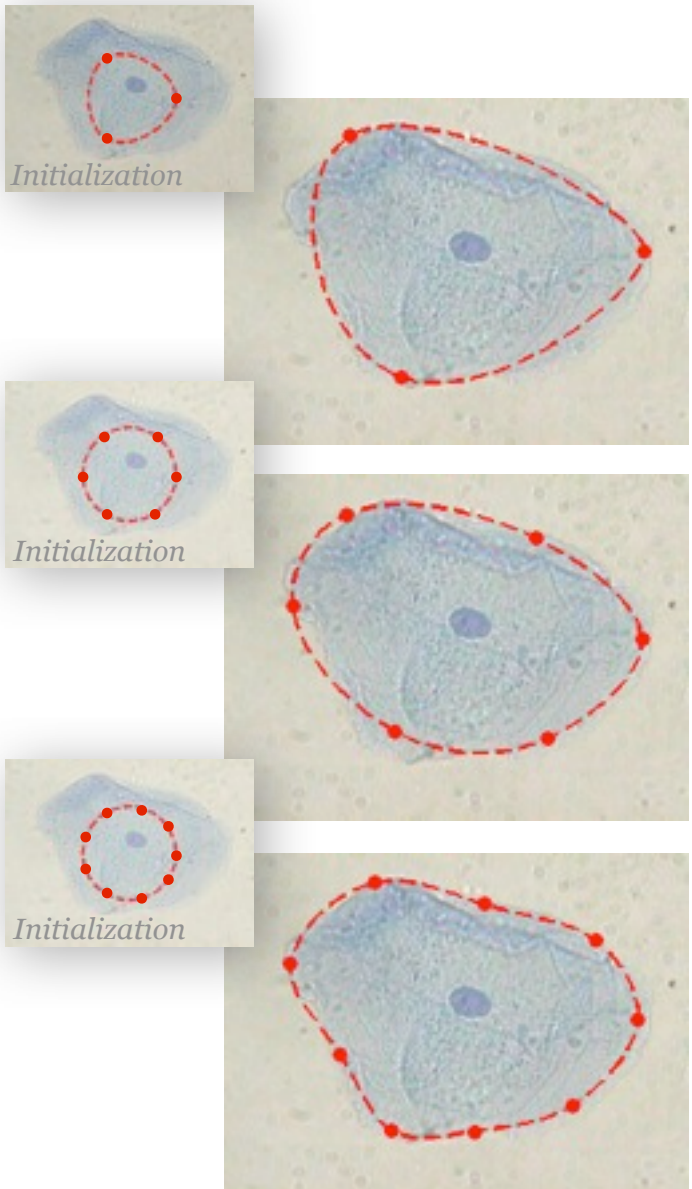
- **Circle**
  - 3 parameters



- **Ellipse**
  - 5 parameters



- **Spline with 3, 6, and 9 control points**
  - 6, 12, and 18 parameters, respectively



Cheek cell (1000x) from the "Biology 101 – Principles of Biology" course, Division of Biological Sciences, University of Montana, Missoula.

Lecture: Methods of inversion

- **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

- **Restoration of a blurred, noisy image**
  - Notations
    - Unknown x
    - Estimation s
    - Observed, blurred, noisy image y
  - Minimization of the mean-square error: $\min_{s} E[\ (x - s)^2\ ]$
    - Assumption: linear solution: $\hat{S} = \hat{G}\ \hat{Y}$ (linear ➜ convolution: s = g * y)

- **Filter**
  - Fourier domain: $\hat{G} = \dfrac{\hat{M}^{\star}\ \hat{R}_x}{|\hat{M}|^2\ \hat{R}_x + \hat{R}_b}$
    - $\hat{R}_x$: Fourier transform of the signal autocorrelation
    - $\hat{R}_b$: Fourier transform of the noise autocorrelation

    Positive, real functions
  - Spatial domain: $G = \left(M^T\ M + \dfrac{R_b}{R_x}\right)^{-1}\ M^T$
    - Similarity with the regularized pseudo-inverse (least square + minimal norm constraint): $(M^T\ M + \alpha\ I)^{-1}\ M^T$
    - For a white Gaussian noise: $R_b = \sigma^2\ I$

- **Fourier domain**
  - Mean-square error: $\epsilon = \mathrm{E}\left[\ \left|\hat{X} - \hat{S}\right|^2\ \right]$

$$= \mathrm{E}\left[\ \left|\hat{X} - \hat{G}\hat{Y}\right|^2\ \right]$$

$$= \mathrm{E}\left[\ \left|\hat{X} - \hat{G}\ (\hat{M}\hat{X} + \hat{B})\right|^2\ \right]$$

$$= \mathrm{E}\left[\ \left|\ \left(1 - \hat{G}\hat{M}\right)\hat{X} - \hat{G}\hat{B}\right|^2\ \right]$$

  - Notation: $\hat{H} = 1 - \hat{G}\hat{M}$

$$\epsilon = \mathrm{E}\left[\ \left|\hat{H}\hat{X} - \hat{G}\hat{B}\right|^2\ \right]$$

$$= \mathrm{E}\left[\ \left(\hat{H}\hat{X} - \hat{G}\hat{B}\right)\ \left(\hat{H}^*\hat{X}^* - \hat{G}^*\hat{B}^*\right)\ \right]$$

$$= \left|1 - \hat{G}\hat{M}\right|^2\ \mathrm{E}\left[\ \left|\hat{X}\right|^2\ \right] - \hat{H}\hat{G}^*\ \cancel{\mathrm{E}\left[\ \hat{X}\hat{B}^*\ \right]} - \hat{H}^*\hat{G}\ \cancel{\mathrm{E}\left[\ \hat{X}^*\hat{B}\ \right]} + \left|\hat{G}\right|^2\ \mathrm{E}\left[\ \left|\hat{B}\right|^2\ \right]$$

$$= \left|1 - \hat{G}\hat{M}\right|^2\ \hat{R}_x + \left|\hat{G}\right|^2\ \hat{R}_b$$

Assumption: the noise is not correlated with the signal

  - Solution: $\hat{G}$ such that ε is minimal

- **Fourier**

$$\hat{G} = \frac{\hat{M}^\star \ \hat{R}_x}{|\hat{M}|^2 \ \hat{R}_x + \hat{R}_b} = \frac{1}{\hat{M}} \underbrace{\frac{|\hat{M}|^2}{|\hat{M}|^2 + \hat{R}_b/\hat{R}_x}}_{\text{Weighting of the naive inversion}}$$

- **Interpretation**
  - Inverse signal-to-noise ratio: $\frac{\hat{R}_b}{\hat{R}_x}$
  - Strong signal compared to noise: $\frac{\hat{R}_b}{\hat{R}_x} \sim 0$
  - Filter $\sim \hat{M}^{-1}$
  - Weak signal compared to noise: $\frac{\hat{R}_b}{\hat{R}_x}$ *tends* toward infinity
  - Filter ~ 0

- **Consequences**
  1. Impossible to recover frequencies degraded by noise (they are *suppressed*)
  2. Lost frequencies ($\hat{M} = 0$) remain equal to zero (no interpolating behavior)
  3. Difficult estimation of R_x (and R_b)

Lecture: Methods of inversion

- **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

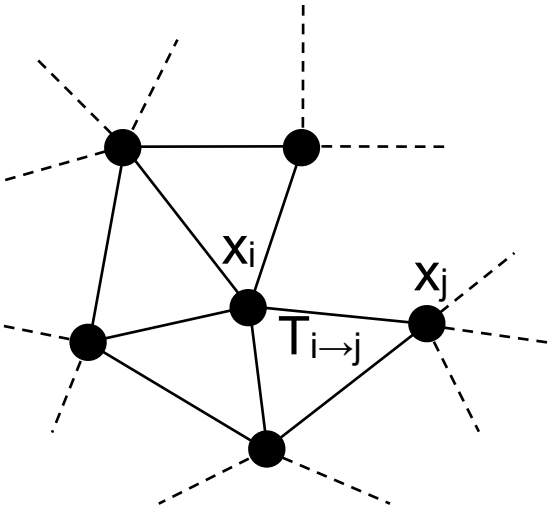
- **Study of a regularized inverse problem**
  - A priori information

    > Estimated distribution of the source: p(x)
    >
    > or
    >
    > Set of rules allowing to generate a source sample

  - Data fidelity
    - Likelihood: $L(y|x) = L_y(x)$ (or L(x) for convenience)
  - A posteriori information: $q(x) = \underbrace{\frac{1}{K}}_{\text{Normalization}} L(x)\ p(x)$
  - Purpose: sampling the a posteriori distribution
- **Bayesian approach**
  - Definition of an a priori distribution
  - Modification ➜ a posteriori distribution
- ***Independent* of the problem complexity**
  - Ex. non linearity

- **Sampling**
  - Generate independent samples of the a posteriori distribution
    - samples = function(source)
  - Analysis of the samples
    - Properties of the source
- **Notion of random walk**
  - Space of sources
  - Possible modifications of a source: possible transitions toward another source
    - Discretization: graph
  - Given a source
    - Choice of a transition
    - Approval or rejection of the transition
      - New sample (possibly identical)

- **Graph of the space of sources**
  - Potential sources
  - Probability of transition $x_i \rightarrow x_j$: $T_{i \rightarrow j}$
    - including the transition $x_i \rightarrow x_i$



- **Initialization**
  - Probability p_i that the random walk is initialized in x_i

- **Equilibrium probability distribution**
  - When the number of transitions tends toward $+\infty$
    - Probability q_i that the *walker* is in x_i

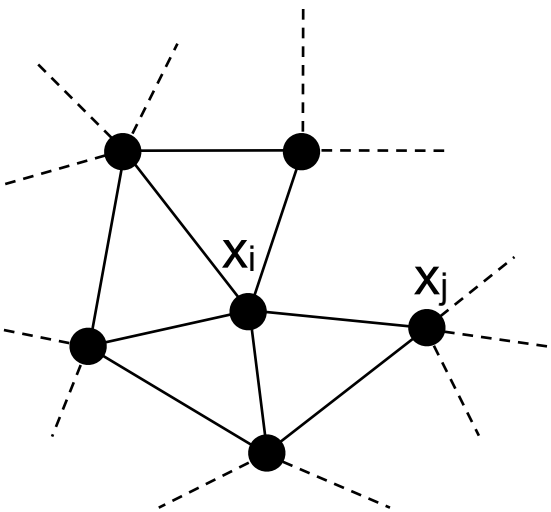
  - Uniqueness: $\forall\ p_i$, the equilibrium distribution is equal to q_i
    - Condition: the graph is connected (but not necessarily complete)

- **At some time instant**
  - Source x_i with n_i neighbors (including x_i)
  - Choose a neighbor x_j at random with uniform probability
  - Transition $x_i \rightarrow x_j$ accepted…

    if $n_j \leq n_i$

    or

    with probability n_i / n_j (< 1) if $n_j > n_i$



- **Equilibrium probability distribution**
  - Uniform distribution: the sources have equal probability

- **Assumption**
  - Random walk with equilibrium distribution p_i

- **How to modify the random walk to have an equilibrium of the form:** $q_i = \frac{1}{\sum_j L_j \; p_j} L_i \; p_i$ **?**

  Normalization

  - At some time instant
    - Source x_i
    - Transition $x_i \rightarrow x_j$ accepted by the original random walk

  - To accept the transition unconditionnally means to keep the equilibrium distribution p_i
    - *Modification*: confirm the transition…

      > if $L_j \geq L_i$ (j has a higher likelihood)
      >
      > or
      >
      > with probability L_j / L_i (< 1) if $L_j < L_i$

- **Link with the Metropolis-Hastings algorithm**
  - p_i: uniform distribution

- **Source space**
  - Possible sources
  - Possible transitions

- **A priori law**

  > Random walk defined by experimental rules
  >
  > or
  >
  > Modification of the uniform random walk if the a priori distribution is known

  $$\begin{cases} \text{pi} = \text{uniform distribution} \\ \text{Li} = \text{a priori distribution} \end{cases}$$

- **A posteriori distribution**
  - Modification of the a priori distribution according to the likelihood

- **Independence of the samples**
  - Use one sample every N iterations

- **A posteriori distribution sampling**
  - Statistical measures
    - Mean or median
    - Variance or mean error
    - More generally, properties of potential solutions

- **Interests**
  - *Independent* of the problem complexity
  - Early riddance of candidate solutions if $L_i = L_i^1 \, L_i^2 \ldots L_i^m$
  - More information on the potential solutions (ex. multimodal law)

- **Difficulties**
  - Independence of the generated samples
  - Sampling of high-dimensional source spaces (too many samples required for statistical significance)

Lecture: Methods of inversion

- **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

- **Likelihood**
  - Probability $P(y|x)$ to observe y if the source is x
- **Estimator:** $\max_x P(y|x)$
  - Solution x such that y has the highest probability of being the result of a measure
- Example - **Model:** $y = M x + b$
  - Distribution $P(y|x)$ = noise distribution
  - Maximum likelihood ⇔ least mean squares?
    - b: white Gaussian noise
    - $\max_x P(y|x) \Leftrightarrow \min_x |y - M x|^2$
  - Not a stable solution

- **Expectation-Maximization / Richardson-Lucy in astrophysics**
  - Adapted to so-called incomplete data
  - Complete data
    - Inaccessible/unobservable
    - Intermediate, more detailed data involved in the formation of the observations
    - Observations can be computed from the complete data but the contrary is false in general
    - Specific to a given problem
  - 2-step iterative algorithm
    1. *Definition of complete data*
    2. Conditional expectation of the log likelihood of these data
    3. Maximization of the conditional expectation
    4. *Back to step 2.*

- **Notations**
  - x: unknown
  - x_n: estimation of x at iteration n
  - y: observations
  - ξ: complete data such that: $y = F(\xi)$
    - F: non injective function ~~$(a \neq b \Rightarrow F(a) \neq F(b))$~~ (distinct elements can have equal images)
  - ξ_n: complete data at iteration n
  - $P(\xi \mid x)$: likelihood of the complete data (instead of $P(y \mid x)$)

- **Initialization**
  - $n \leftarrow 0$
  - $x_0 \leftarrow$ initial estimation
  - $\xi_0 \leftarrow$ initial complete data such that: $F(\xi_0) = y$

- **Repeat**
  - E-step: $f(x) = E[\log P(\xi_n \mid x) \mid y, x_n]$
  - M-step: $x_{n+1} = \arg\max_x f(x)$
  - $n \leftarrow n + 1$
- **Until convergence w/r/t** x

- **Poisson process**
  - Average number of photons reaching the ith detector

$$(M\ x)_i = \sum_j m_{i,j}\ x_j \qquad \text{(on average)}$$

    - Average number of photons emitted by the jth source: x_j
    - Probability that a photon emitted by the jth source reaches the ith detector: $m_{i,j}$
  - An observation y represents a set of independent realizations of the Poisson process

- **Likelihood:** $P(y|x) = \prod_i e^{-(Mx)_i} \frac{(Mx)_i^{y_i}}{y_i!}$

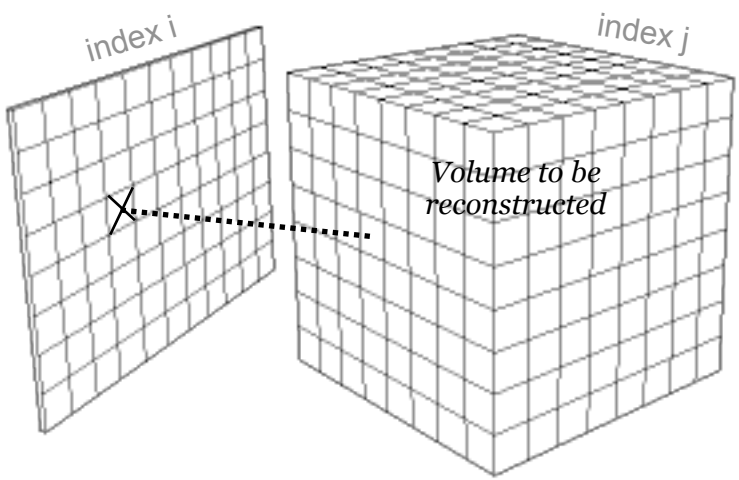
  independence ➜ product

  - Maximum likelihood
  - or minimum of: - log(likelihood)

$$\min_x \sum_i \left[\left(\sum_j m_{i,j}\ x_j \right) - y_i\ \log \left(\sum_j m_{i,j}\ x_j \right) \right]$$

  - Energy
    - Non quadratic
    - Strictly convex

- **Poisson process (definition)**
  - Stochastic process of event/occurrence counting
  - Properties
    - N(t), t ≥ 0: number of events up to time t
    - N(0) = 0
    - No simultaneous events
    - If]t1,t2] ∩]t3,t4] = ∅, then N(]t1,t2]) and N(]t3,t4]) are independent
      - Notation: N(]t1,t2]) = N(t2) - N(t1)
    - N(]t1,t2]) = F(t2 - t1)
    - Distribution: $P[N(]t, t+\tau]) = k] = \frac{e^{-\lambda\tau}(\lambda\tau)^k}{k!}$
      - λ: expected number of events per unit of time



- **Normal equations**

$$\sum_i y_i \frac{m_{i,j}}{\sum_k m_{i,k}\ x_k} = \sum_i m_{i,j}$$

  - Difficult to find the closed-form solution x_j

- **Introduction of complete data**
  - Unknown: radiopharmaceutical concentration $\{x_j, j\}$
  - Observation: projections $\{y_i, i\}$
  - Complete data: contribution $m_{i,j}\ x_j$ of the jth voxel x_j to the ith observation y_i
    - F: $\sum_j$

- **Algorithm**

$$x_j^{n+1} = x_j^n \underbrace{\frac{1}{\sum_i m_{i,j}}}_{Normalization} \overbrace{\sum_i \left[m_{i,j} \underbrace{\frac{y_i}{\underbrace{\sum_k m_{i,k}\ x_k^n}_{Projection}}}_{Error} \right]}^{Backprojection}$$

projection estimated by the model and the current solution for x

backprojection of 1 (no error)

$$\sum_i m_{i,j} \times 1$$

  - Interpretation
    - Projection: $y_i = \sum_j m_{i,j}\ x_j$
    - Backprojection: $x_j = \sum_i m_{i,j}\ y_i$

- **Algorithm**

$$x_j^{n+1} = x_j^n \; \frac{1}{\sum_i m_{i,j}} \; \sum_i \left[m_{i,j} \; \frac{y_i}{\sum_k m_{i,k} \; x_k^n} \right]$$

- **At convergence**
  - Fixed point

$$x_j = x_j \; \frac{1}{\sum_i m_{i,j}} \; \sum_i \left[m_{i,j} \; \frac{y_i}{\sum_k m_{i,k} \; x_k} \right]$$

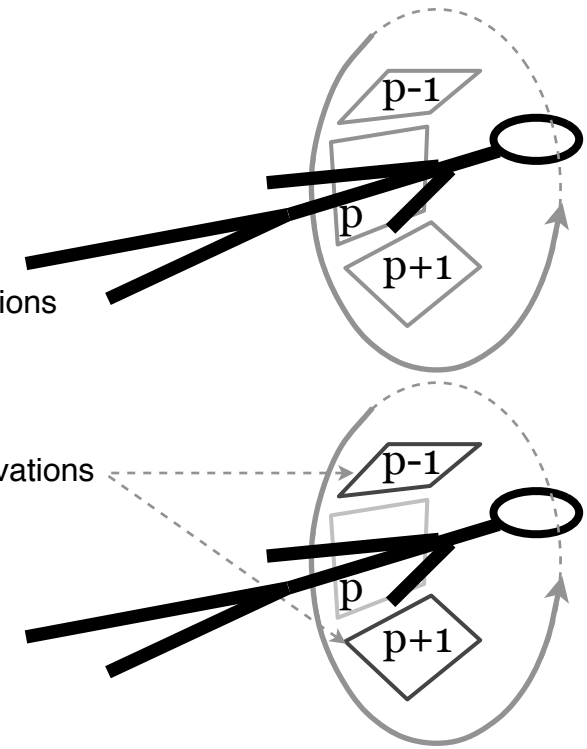
$$\Longleftrightarrow \; 1 = \frac{1}{\sum_i m_{i,j}} \; \sum_i \left[y_i \; \frac{m_{i,j}}{\sum_k m_{i,k} \; x_k} \right]$$

  - Equivalent to the normal equations

$$\sum_i y_i \; \frac{m_{i,j}}{\sum_k m_{i,k} \; x_k} = \sum_i m_{i,j}$$

- **Expectation-Maximization: slow convergence**

- **Ordered subsets**
  - Inspired by block-based linear system resolution
  - Normally
    - At each iteration, all the unknowns are updated taking into account all the observations
  - Acceleration of convergence
    - Sub-iteration: update of all the unknowns taking into account a subset of the observations
    - Iteration = set of all the sub-iterations (covering all the observations)
  - Generally
    - Subsets of observations: disjoint and of identical size



- **In theory**
  - Acceleration of the convergence by a factor = number of subsets

- **Regularization: ML ➜ MAP**

- **Bayes theorem**
  - $P(y \mid x)\, P(x) = P(x \mid y)\, P(y)$

    likelihood — a priori — a posteriori — Independent of the unknown

- **Interpretation**
  - P(x): a priori information (without accounting for the observations)
    - E.g.: Gibbs distribution
  - P(x | y): probability of the unknown a posteriori (knowing the observations)

- **Maximum a posteriori**
  - $\max_x \underbrace{P(y \mid x)}_{\textit{Likelihood}}\, \underbrace{P(x)}_{\textit{A priori}} = \max_x P(x \mid y)$
  - or minimization of: – log P(x | y)

- Example: **EM for emission SPECT**

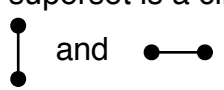
$$x_j^{n+1} = x_j^n \; \frac{1}{\sum_i m_{i,j} + \mathbf{regul.}} \; \sum_i \left[m_{i,j} \; \frac{y_i}{\sum_k m_{i,k} \; x_k^n} \right]$$

- **Model:** y = M x + b
  - b: white Gaussian noise

- **Unregularized**
  - Maximum likelihood ⇔ least mean squares [$\max_x P(y|x) \Leftrightarrow \min_x |y - M x|^2$ (see slide on ML)]

- **Regularization**
  - Hammersley–Clifford theorem

Context

- Graph G = (V, E)
  - E.g.: Vertices V = image pixels;
  - Edges E = links between pixels to represent neighborhood relations (such as 4-connectivity)
- Clique = complete subgraph of G
  - Maximal clique: no strict superset is a clique
    - Ex. 4-connectivity:  and

- Equivalence between 1. and 2.:
  1. P(x) > 0 and represents a Markov random field over G
  2. P can be factorized over the set C_m of maximal cliques: $P(x) = \frac{1}{Z} e^{-\sum_{c \in C_m} V(c)}$ (Gibbs distribution)
     - V(c): potential or energy of the clique c
     - $\sum_{c \in C_m} V(c)$: potential of the configuration x
     - Z: normalizing constant (referred to as partition function here)

- **Regularization**
  - Markov random field prior: P(x(u,v) | x(i,j), (i,j) ≠ (u,v)) = P(x(u,v) | N(x(u,v)))
    - Image smoothness: neighboring pixels tend to have similar colors (expect across edges)
    - Ex.
      - V(c) = α (x(u+1,v) - x(u,v))^2 (horizontal clique)
      - V(c) = α (x(u,v+1) - x(u,v))^2 (vertical clique)
    - MAP: $\min_x$ – log P(x | y)
      - with – log P(x | y) – log(P(y))= – log[P(y | x) P(x)] = – log P(y | x) – log P(x)
      - and $P(x) = \frac{1}{Z}\, e^{-\sum_{c\in C_m} V(c)}$ After "- log"
    - Image energy: $\sum_{c\in C_m} V(c) = \alpha \sum_{(u,v)} (x(u+1,v)-x(u,v))^2+(x(u,v+1)-x(u,v))^2 \simeq \alpha \sum_{(u,v)} |\nabla x(u,v)|^2$
      - Tikhononv regularization on the norm of the gradient ➜ isotropic smoothing
    - Edge-preserving behavior: α ➜ α(u,v)

Lecture: Methods of inversion

- **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

- **Shannon entropy**
  - Random variable X with realizations in $x = \{x_1, x_2 \ldots x_n\}$ and respective probabilities $p_1, p_2 \ldots p_n$
  - $\log(1 / p_i)$: amount of information carried by the realization x_i
    - If an event is almost certain, its occurrence does not tell much
    - If an event is rare, the information gain due to its occurrence is significant
    - The joint probability of several independent events = product of the individual probabilities
      - log ➜ total gain = sum of the respective information gains
  - Entropy: $H(p) = -\sum_{i=1}^{n} p_i \log p_i$ (expected information gain / measure of dispersion ; also expected value of - log p)
  - Low entropy ➜ less uncertainty about the random variable and conversely
    - One realization with probability 1: entropy = 0; All realizations equally probable: entropy = log(n)

- **Differential entropy**
  - Continuous version: $H(p) = -\int p(x) \log p(x) \, dx$
  - Gaussian distribution: distribution with maximum entropy among all distributions with constrained mean and variance

- **Relative entropy / Information gain / Kullback-Leibler divergence**
  - Discrete: $D_{KL}(p||q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$ ←----- True distribution ←----- Model
  - Continuous: $D_{KL}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} \, dx$

- **Problem**
  - Set of constraints involving a probability mass/density function p (the unknown)
  - Solution with maximal entropy

- **Intuition**
  - The solution p "encodes" the constraints
  - but must be the least possible informative, otherwise
    - A solution with a higher entropy would not "encode" the whole information available
    - A solution with a lower entropy would be more informative than the information known from the constraints

- **Formalization**
  - Maximize: $H(p) = -\sum_{i=1}^{n} p_i \log p_i$
  - Under the constraints: $\sum_{i=1}^{n} p_i \, \varphi_k(x_i) = c_k, \;\; k = 1, 2 \ldots m$
    - Additional constraint: $\varphi_0 = 1$, $c_0 = 1$ (ensures that $\sum_i p_i = 1$)
  - Lagrange multipliers λ_k: $J(p, \lambda) = -\sum_{i=1}^{n} p_i \log p_i + \sum_{k=0}^{m} \lambda_k \left(\sum_{i=1}^{n} p_i \, \varphi_k(x_i) - c_k \right)$
    - Solution: $\begin{cases} p_i = \frac{1}{Z} \, e^{\sum_{k>0} \lambda_k \varphi_k(x_i)} \\ \dfrac{\partial \log(Z)}{\partial \lambda_k} = c_k \end{cases}$ (to be solved for λ_k, $k > 0$)

- **A priori on the solution: q**
  - Purpose: minimize the additional information brought by p with respect to the information carried by q

- **Formalization**
  - Minimize: $D_{KL}(p||q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}$ (or maximize the opposite)
  - Under the constraints: $\sum_{i=1}^{n} p_i \varphi_k(x_i) = c_k, \; k = 1, 2 \ldots m$
    - Additional constraint: $\varphi_0 = 1$, $c_0 = 1$
  - Lagrange multipliers λ_k: $J(p, \lambda) = -\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{k=0}^{m} \lambda_k \left(\sum_{i=1}^{n} p_i \varphi_k(x_i) - c_k \right)$
    - Solution: $\begin{cases} p_i = \frac{1}{Z} q_i \, e^{\sum_{k>0} \lambda_k \varphi_k(x_i)} \\ \frac{\partial \log(Z)}{\partial \lambda_k} = c_k \end{cases}$

Deterministic vs. stochastic: Brief comparison

- **Sometimes equivalent**

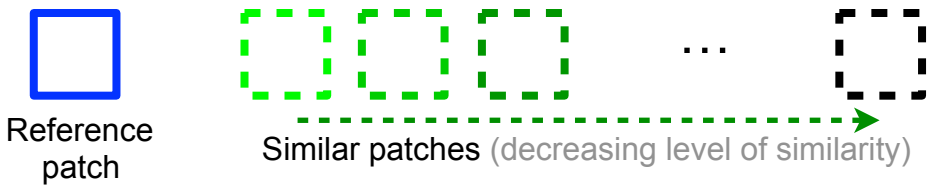
- **Deterministic**
  - Can account for the nature of the noise
  - Numerous tools from algebra and functional analysis

  - Deterministic model regardless of the process
  - Computation of a unique solution (although other good/better solutions might exist)

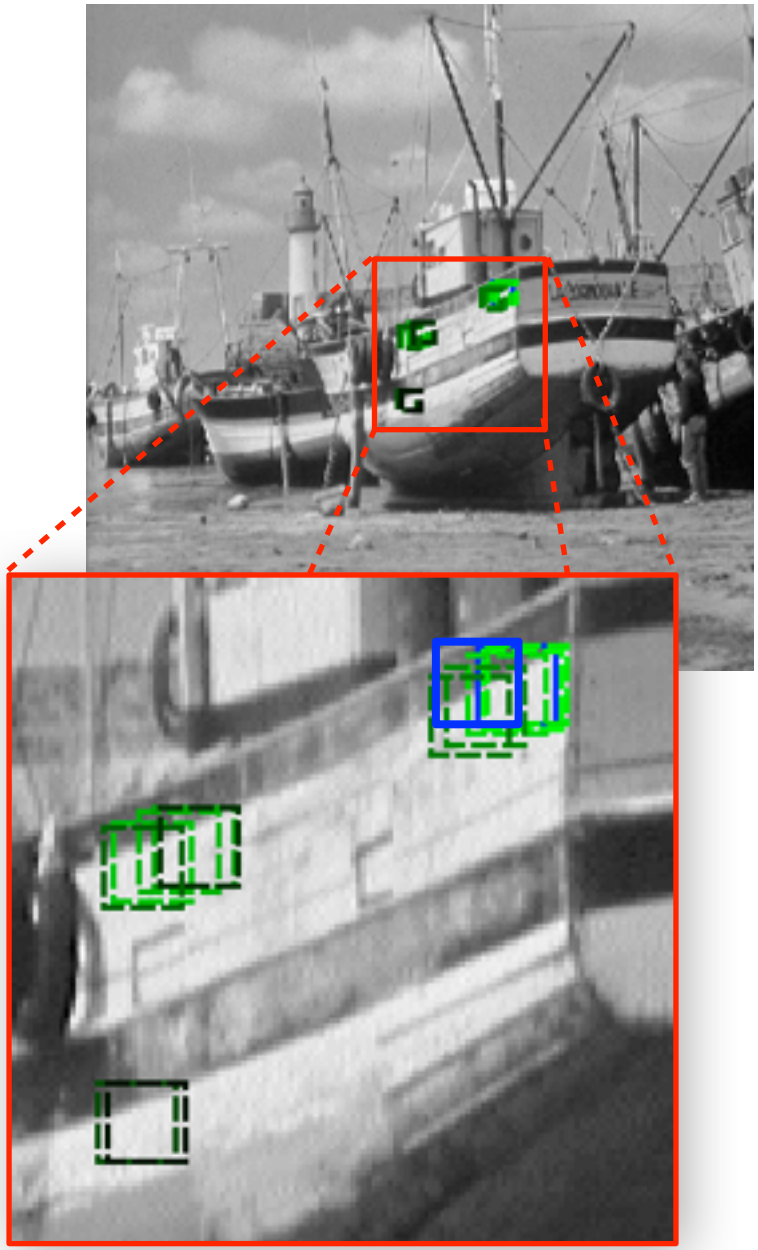
- **Stochastic**
  - Accounting for the random nature of some processes
  - Accounting for the nature of the noise
  - Possibly better knowledge about (the) solution(s)

  - Potentially more difficult to apprehend

- **Principle**
  - Patch: small piece of image (classically, a nxn-square)
  - Assumption: independent, identically distributed noise
  - For a given patch
    - There certainly exist similar patches in the image
      - ▸ Set of similar patches / class of patch



    - These similar patches are corrupted by a ≠ realization of the noise
      - Strong correlation of the noiseless signal *behind*
      - ▸ Denoising should be possible
- **Stochastic formulation**
  - For a patch in a given class ……(example)……
    - If we know the central pixel
    - Then the uncertainty on the knowledge of the neighborhood must be low
      - ▸ Low conditional entropy
    - However, noise ➜ higher uncertainty ➜ higher conditional entropy
  - Denoising: conditional entropy minimization
  - In practice: From “class of patch” to “set of similar patches”
    - ▸ Probability ➜ statistics (samples)



- **Notations**
  - D: image domain (set of all the pixel sites)
  - X: random variable of pixel color
  - C: structure of a pixel neighborhood (defines the patch geometry)
  - Y: pixel neighborhood = $\{X(t), t \in C\}$
  - Z: patch = $\{X, Y\}$
  - h: conditional entropy of a pixel color knowing its neighborhood

- **Stochastic formulation**
  - Denoising = conditional entropy minimization
    - $$\begin{aligned} X^* &= \arg\min_X h(X|Y) \\ &= \arg\min_X \int_y h(X|Y=y)\ f_y(y)\, dy \\ &= \arg\min_X E_Y[h(X|Y)] \\ &\approx \arg\min_X \frac{1}{|D|} \sum_{s \in D} h(X|Y=y_s) \end{aligned}$$

      where f_y: probability density function of Y

      law of large numbers
    - Problem: $Y = y_s$ corresponds to (unknown) noiseless information
  - Practical formulation
    - $$\begin{aligned} X^* &= \arg\min_X h(X|\tilde{Y}) \\ &\approx \arg\min_X \frac{1}{|D|} \sum_{s \in D} h(X|\tilde{Y}=\tilde{y}_s) \end{aligned}$$

      where $\tilde{Y}$ and $\tilde{y}_s$ are the noisy versions of Y and y_s

- **Noisy vs. noiseless**
  - Notations: $\tilde{X} = X + N$ and $\tilde{Y} = Y + M$ with N independent of X (hence, M is independent of Y)
  - By definition: $h(\tilde{X}|\tilde{Y}) = H(\tilde{X}) - I(\tilde{X};\tilde{Y})$ where H is the entropy and I is the mutual information

    $$h(X|Y) = H(X) - I(X;Y)$$

    - $\tilde{X} = X + N$ ➜ $H(\tilde{X}) \geq H(X)$ (1)
    - $P(\tilde{Y}|X,Y) = P(Y+M|X,Y)$ ➜ $X \to Y \to \tilde{Y}$: forms a Markov chain

      $$\begin{aligned} &= P(M|X,Y) \\ &= P(M|Y) \\ &= P(Y+M|Y) \\ &= P(\tilde{Y}|Y) \end{aligned}$$

      - Data processing inequality: $I(X;Y) \geq I(X;\tilde{Y}) = I(\tilde{Y};X)$
    - $P(\tilde{X}|X,\tilde{Y}) = P(X+N|X,\tilde{Y})$ ➜ $\tilde{Y} \to X \to \tilde{X}$: forms a Markov chain

      $$\begin{aligned} &= P(N|X,\tilde{Y}) \\ &= P(N|X) \\ &= P(X+N|X) \\ &= P(\tilde{X}|X) \end{aligned}$$

      - Data processing inequality: $I(\tilde{Y};X) \geq I(\tilde{Y};\tilde{X})$

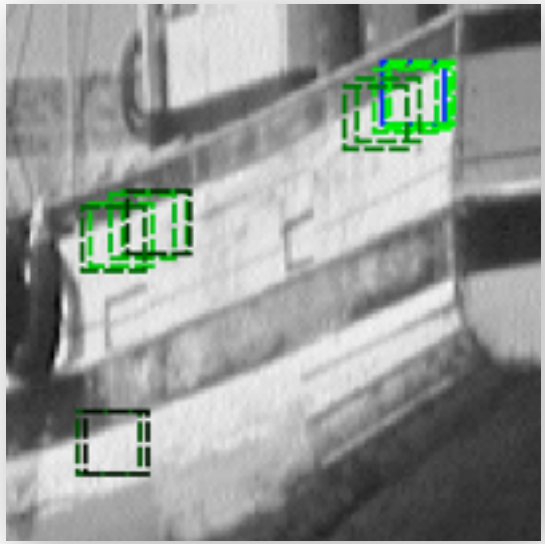
    $I(X;Y) \geq I(\tilde{X};\tilde{Y})$ (2)
  - (1) and (2) ➜ $h(\tilde{X}|\tilde{Y}) \geq h(X|Y)$

- **Noisy vs. *mixed* (practical formulation)**
  - Practical formulation
    - $$X^* = \arg\min_X h(X|\tilde{Y}) \approx \arg\min_X \frac{1}{|D|} \sum_{s \in D} h(X|\tilde{Y} = \tilde{y})$$
  - Previous slide: $I(\tilde{Y}; X) \geq I(\tilde{Y}; \tilde{X})$
  - $$\begin{aligned} & I(\tilde{Y}; X) \geq I(\tilde{Y}; \tilde{X}) \\ \Leftrightarrow\ & H(X) - h(X|\tilde{Y}) \geq H(\tilde{X}) - h(\tilde{X}|\tilde{Y}) \\ \Leftrightarrow\ & h(\tilde{X}|\tilde{Y}) \geq h(X|\tilde{Y}) + \underbrace{H(\tilde{X}) - H(X)}_{\text{Positive}} \\ \Leftrightarrow\ & h(\tilde{X}|\tilde{Y}) \geq h(X|\tilde{Y}) \end{aligned}$$

- **Gradient descent**

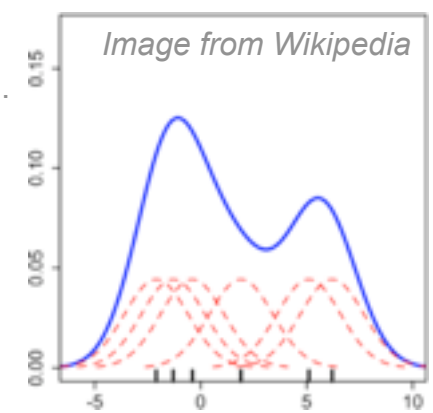
$$\begin{cases} x_s^*(0) \leftarrow \tilde{x}_s \\ x_s^*(n+1) \leftarrow x_s^*(n) - d\tau \, \dfrac{\partial h(X|\tilde{Y})}{\partial x_s}(x_s^*(n)) \end{cases}$$

  - Energy derivative: $\dfrac{\partial h(X|\tilde{Y})}{\partial x_s}$?



- **Notations**
  - D: image domain (set of all the pixel sites)
  - $D_{\tilde{y}_i}$: subset of D of the sites of the pixels whose neighborhood is equal to $\tilde{y}_i$

- **Derivative**
  - $$\begin{aligned} h(X|\tilde{Y}=\tilde{y}_i) &= -\int_x f_{x|y}(x|\tilde{y}_i) \log\left(f_{x|y}(x|\tilde{y}_i)\right) dx \\ &= -E_X[\log(f_{x|\tilde{y}_i})] \\ &\approx -\frac{1}{|D_{\tilde{y}_i}|} \sum_{s \in D_{\tilde{y}_i}} \log f_{x|y}(x_s|\tilde{y}_i) \end{aligned}$$ (Ahmad-Lin entropy estimator)
    - where: $f_{x|y}(x|y) \approx \dfrac{1}{|D_y|} \sum_{t \in D_y} K(x - x_t)$ … (kernel-based PDF estimator) …



  - It can be shown that:

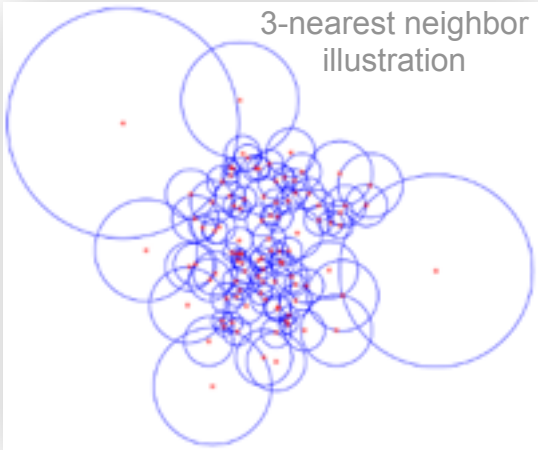
$$\frac{\partial h(X|\tilde{Y}=\tilde{y}_i)}{\partial x_j} = \begin{cases} -\dfrac{1}{|D_{\tilde{y}_i}|} \dfrac{\nabla f_z}{f_z}(z_j) \cdot \dfrac{\partial z_j}{\partial x_j} & \text{if } j \in D_{\tilde{y}_i} \\ 0 & \text{otherwise.} \end{cases}$$

where z_j is the concatenation of x_j and $\tilde{y}_j$

- **Derivative**
  - Previous slide: $\dfrac{\partial h(X|\tilde{Y}=\tilde{y}_i)}{\partial x_j} = \begin{cases} -\dfrac{1}{|D_{\tilde{y}_i}|}\dfrac{\nabla f_z}{f_z}(z_j)\cdot\dfrac{\partial z_j}{\partial x_j} & \text{if } j \in D_{\tilde{y}_i} \\ 0 & \text{otherwise.} \end{cases}$
  - Energy: $h(X|\tilde{Y}) \approx \arg\min_X \dfrac{1}{|D|}\sum_{s\in D} h(X|\tilde{Y}=\tilde{y})$
  - Derivative: $\dfrac{\partial h(X|\tilde{Y})}{\partial x_s}(x_s) \approx -\dfrac{1}{|D|}\dfrac{\nabla f_z}{f_z}(z_s)\cdot\dfrac{\partial z_s}{\partial x_s}$

    Approximated by the mean-shift

    Projection of the patch z_s onto the central pixel x_s



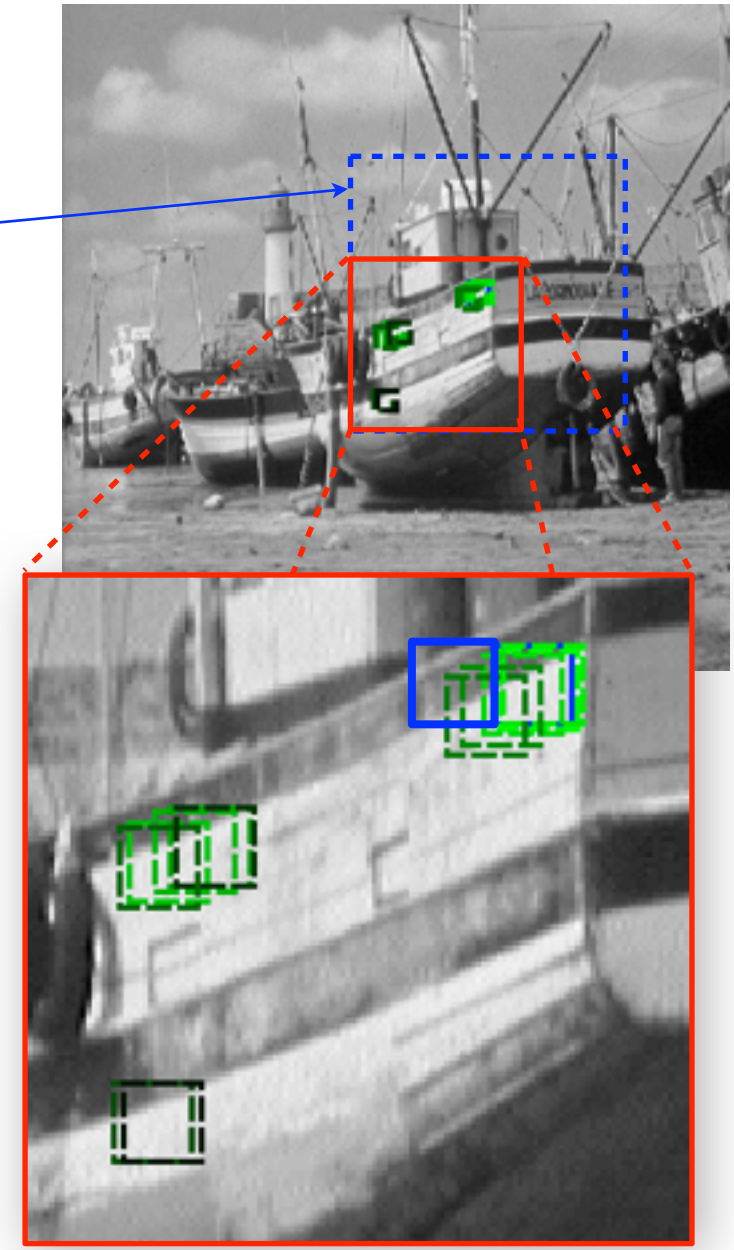
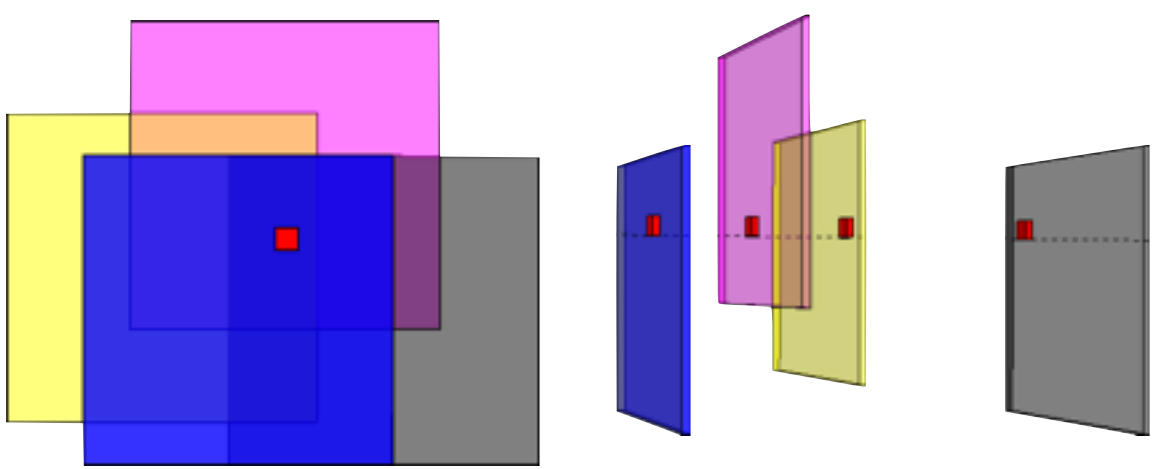
- **Mean-shift approximation**
  - In the k-nearest neighbor (kNN) framework: $\dfrac{\nabla f_z}{f_z}(z_s) \approx \dfrac{d+2}{\rho^2}\left(\dfrac{1}{k}\sum_{\substack{z_i\in\{z\}\\ |z_i-z_s|\le\rho}} z_i - z_s\right)$
- **Gradient descent**

$$\begin{cases} x_s^*(0) \leftarrow \tilde{x}_s \\ x_s^*(n+1) \leftarrow x_s^*(n) - d\tau\,\dfrac{\partial h(X|\tilde{Y})}{\partial x_s}(x_s^*(n)) \end{cases} \xrightarrow{\quad d\tau = \dfrac{\rho^2|D|}{d+2} \quad} \begin{cases} x_s^*(0) \leftarrow \tilde{x}_s \\ x_s^*(n+1) \leftarrow \left(\dfrac{1}{k}\sum_{\substack{z_i\in\{z\}\\ |z_i-z_s^*(n)|\le\rho}} z_i\right)\cdot\dfrac{\partial z_s}{\partial x_s} \end{cases}$$

  - One iteration is enough
  - Refinement: weighting of z_i as a function of its distance to z_s : $w_i = \exp\left(-\dfrac{|z_i - z_s^*(n)|^2}{\alpha}\right)$
  - Alternative: remove the projection and combine the denoised patches

- **Summary**
  - For a given noisy patch
    1. Look for similar patches (e.g., the k closest patches in the L^2 sense)…
       - within the entire image or…
       - within some rectangular area around the patch
    2. Take the (weighted) average of those patches ➜ denoised patch
    3. (a) Either replace the central pixel of the noisy patch with the central pixel of the denoised patch (with the projection onto the central pixel)

       (b) Or accumulate denoised patches (without projection)
  - Repeat for all the noisy patches



- **Accumulation/aggregation/combination of denoised patches**