Inverse Problems in images by variational methods

Laure Blanc-Féraud

DR CNRS
Laboratoire I3S
Morpheme group (I3S - INRIA Sophia Antipolis – iBV)

Image Construction

$$g = H(u) \;\text{☺}\; n$$

- g : observation = Physical quantities : optics, radar, laser, IR, magnetic field, X rays, ultrasons, …
- H : operator which links the observation to the quantity we are looking for, we want to image u, through the measuring instrument and possibly a reconstruction process.
- u : image we want to obtain
- n : random part in the observation process (noise)

Observation model: Gaussian noise case

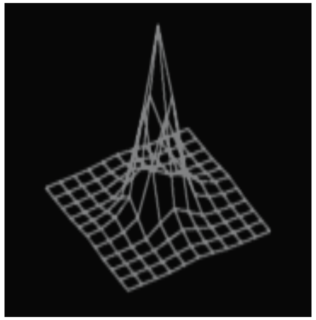
Observed images are degraded :

$$g = h * u + n$$

Noise
white Gaussian

Observed Image

Convolution kernel (PSF)

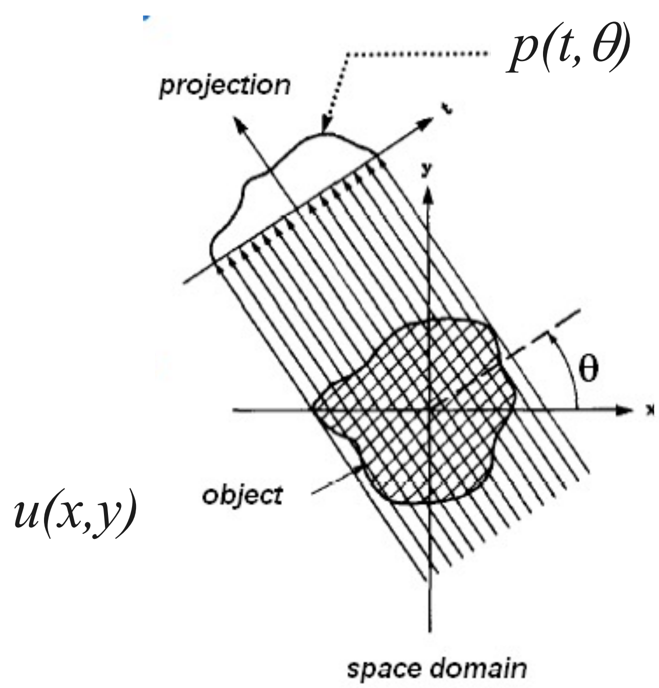


Original Image

- Restoration : retrieve u from g
- Inverse $g = h * u + n$ is an ill posed problem

Examples

- Linéaires : $H(u) = H.u$
  - Tomographic Reconstruction, in medical imaging or geosciences,….
  - Reconstruct the volume of an object (the human body in the case of medical imaging), based on a series of measurements made outside the object.
    Example of X-ray tomography

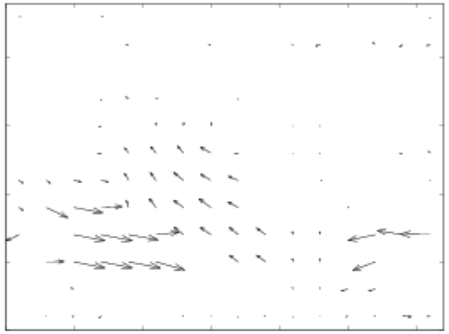


H= Radon Matrix = 1D projection according to the angle θ

Examples

- Non Linear example :

  - Optical flow
    we observe $I(x,t)$ and $I(x,t+1)$ and we're looking for the apparent movement $v(x,t)$ such that $I(x+v,t) = I(x,t+1)$, or
  - $I(x+v(x,t),t) - I(x,t+1)=0$



Linearization by derivation: optical flow equation

$$\nabla I\left(x,t+1\right).v\left(x,t\right)+I_t\left(x,t\right)=0$$

  - Microwave reverse diffraction, non-destructive control: Maxwell equations …

- Direct Problem $g = H(u)$ ☺ n

  It is the equation of image construction, the mathematical modeling of the physical phenomena of acquisition.

It defines g from H, u and n.

- Inverse Problem

  From the observed data g, and modeling of the direct problem, find u assuming H known and the law parameters of n known.

  → H could be partially only known,

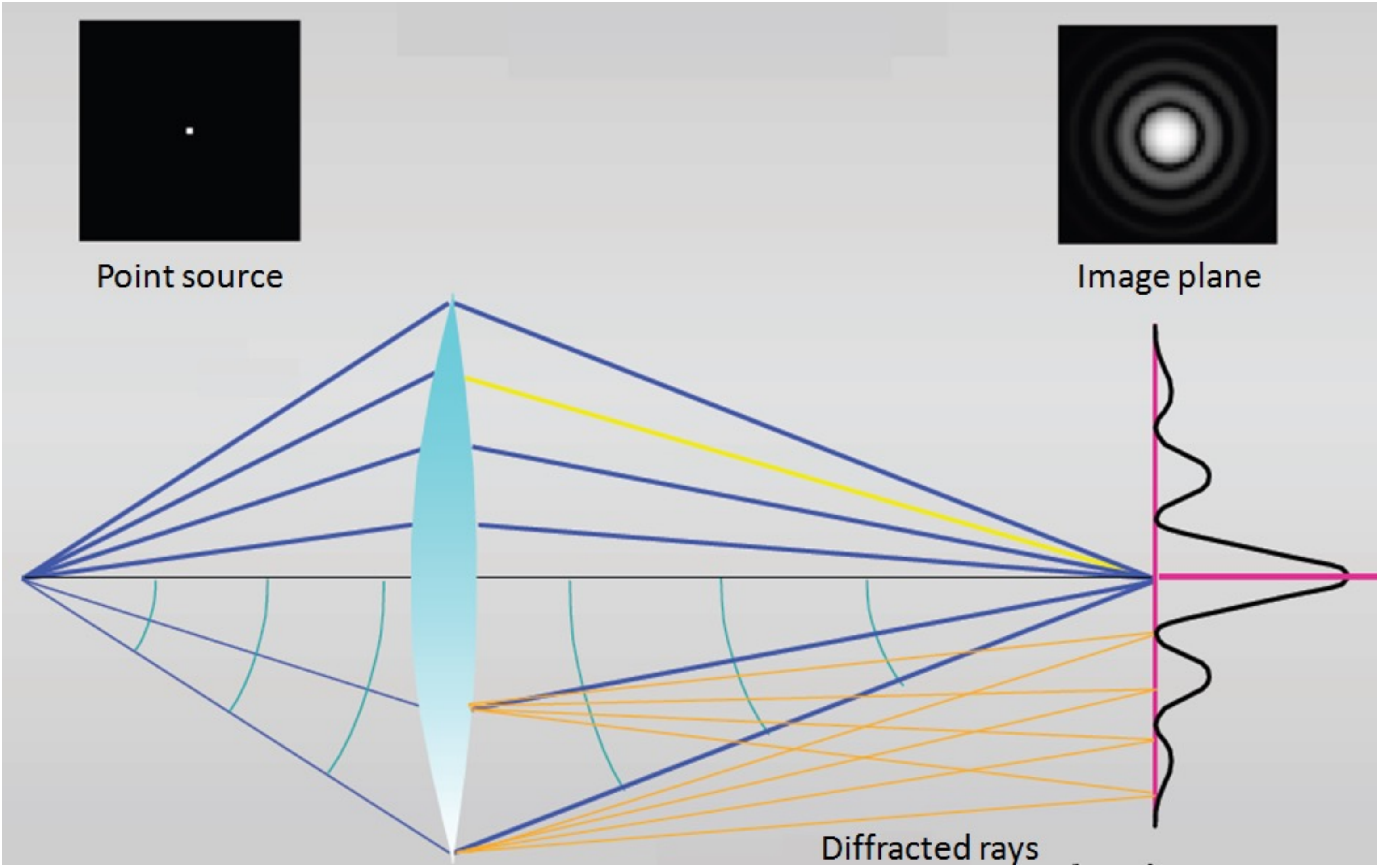
  → Noise parameters must be estimated before or in the same time.

Image Restoration: deblurring, denoising

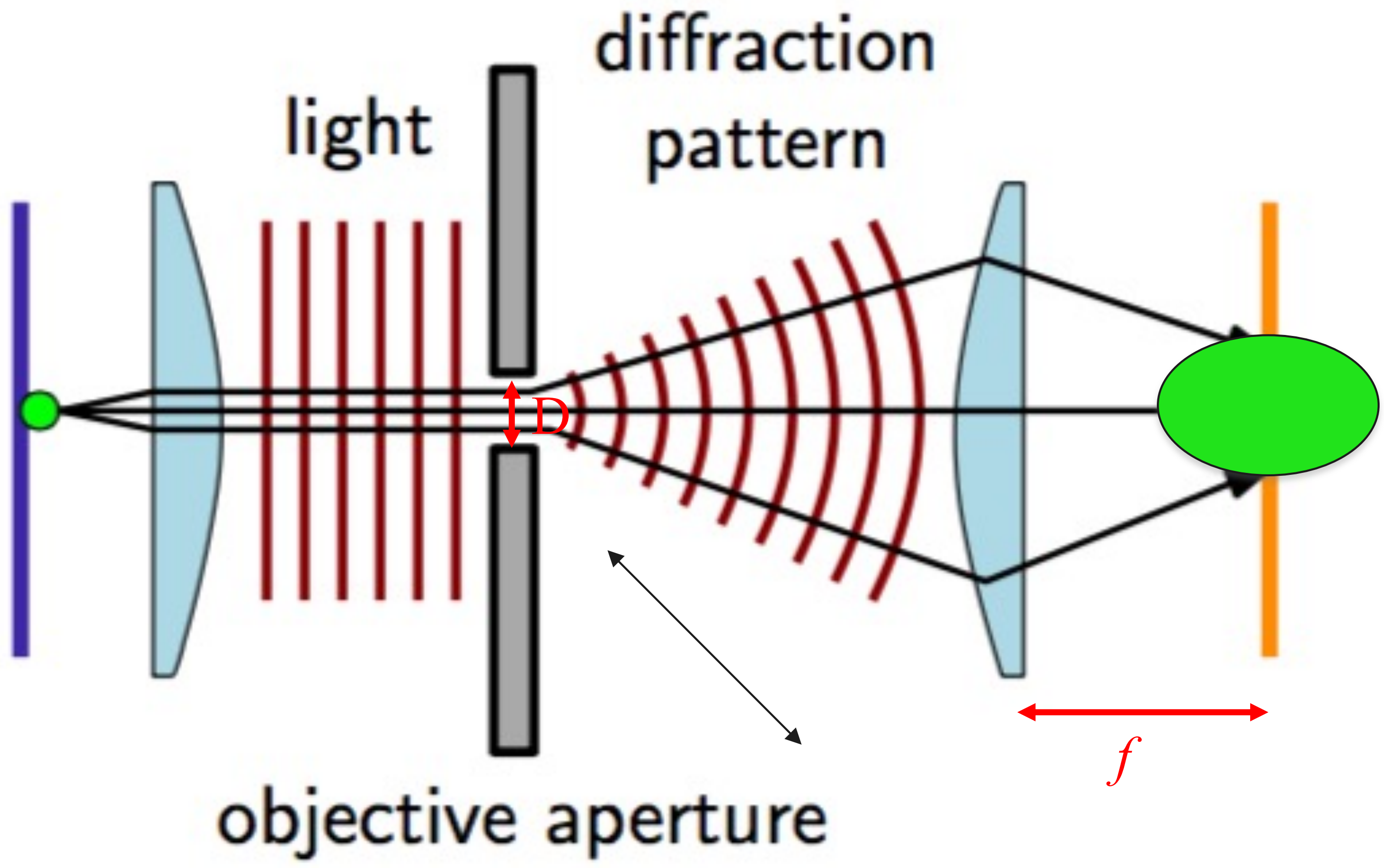
- Optical images have limited resolution due to diffraction limit
- Model the diffraction effect on images : blur is modeled by convolution
- Simple Model:
  - Convolution is linear
  - But difficult to inverse: ill-posed problem.

Diffraction

The image of a point source is the Airy patch

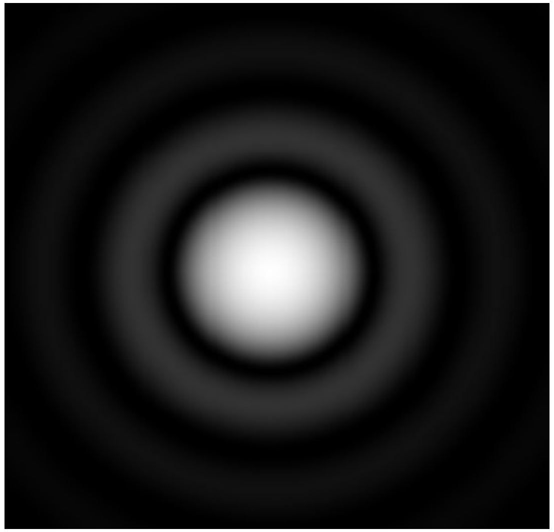


Diffraction



Optical system blurring : diffraction (continuing)

- Objective aperture D, λ is the wavelength (in visible light $\lambda=0.6\mu m$), f is the focal length

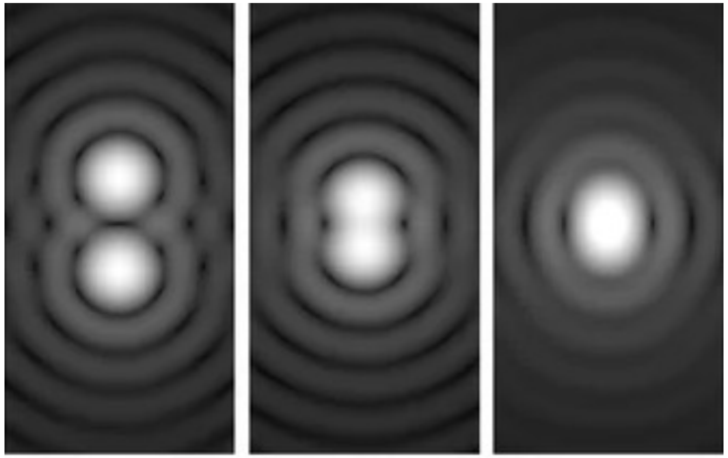


- The image of a point is the Airy patch → bright circular spot with attenuated annulus.

The radius of the Airy patch brings an idea of the dimension of the smallest details which can be view with an ideal optic. The radius is given by

$$r_A = 1.22\frac{\lambda f}{D}$$

It also gives information on the resolution of the image. If two points are closer than r they cannot be resolved.



PSF

- All image formation system is not perfect and introduces blur in the observed image.
- The degree of spreading (blurring) of a single point like (Sub Resolution) object is a measure for the quality of an optical system. The 2D or 3D blurry image of such a single point light source is called the **Point Spread Function** (PSF).
- In general, the blurring is largely due to diffraction limited imaging by the instrument (in x,y directions)
- It could also be due to out of focus.
- The convolution is the mathematical model which explains the formation of an image that is degraded by blurring.

Spatial Dispersion: IR or PSF

We call impulse response (IR) or Point Spread Function (PSF), the imaging system response to a punctual intensity distribution (Dirac distribution). The PSF function is expressed in the continuous setting image coordinates. It is normalized and positive; It is usually denoted by h variable.

$$PSF \quad h : \Omega \subset R^2 \rightarrow R^+$$

$$x \rightarrow h(x) \quad \text{and} \quad \int_\Omega h(x)dx = 1$$

In the most general case, h is depending on the position and the spatial intensity distribution of the scene.

This is the PSF: Point Spread Function

Optical system blurring

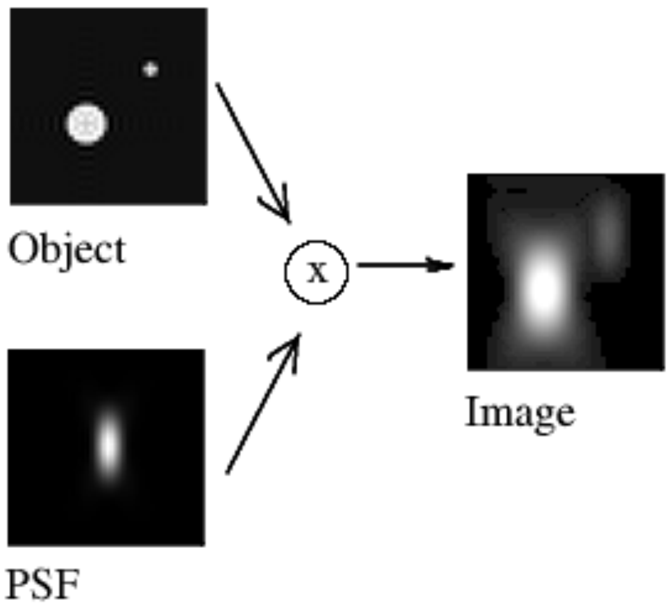
- Diffraction is mathematically described by a convolution equation of the form $$g_{x,y} = \left(h * u\right)_{x,y}$$

- where the image g arises from the convolution of the real light sources u (the specimen) and the PSF h. The convolution operator $*$ implies an integral all over the space: $$g_{x,y} = \int_{\Omega} h(x-s, y-t)\, u(s,t)\, ds\, dt$$

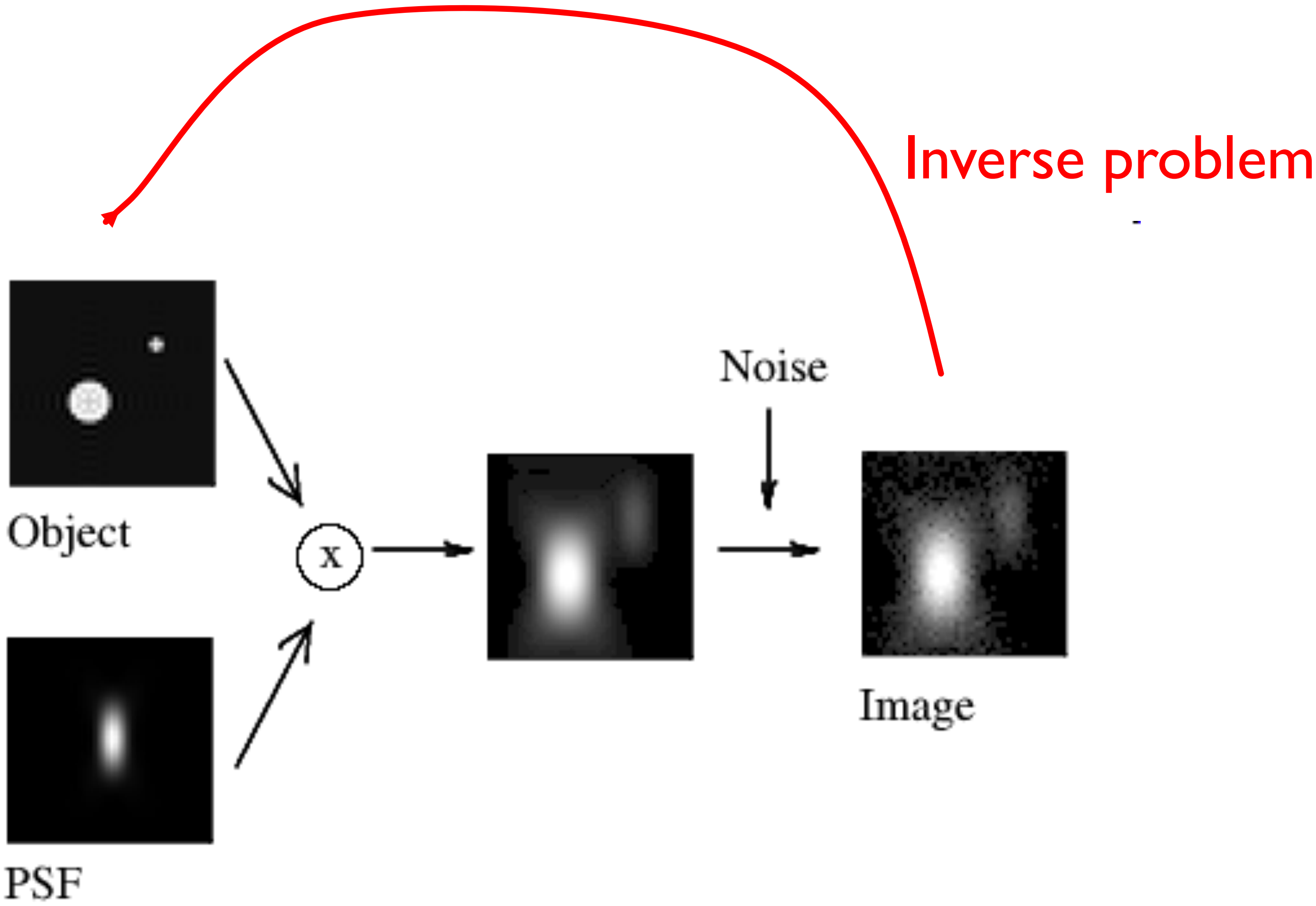
- **Interpretation** You can interpret the convolution equation as follows: the recorded intensity in a voxel located at point (x,y) of the image g arises from the contributions of all points of the specimen u, their real intensities weighted by the PSF h depending on the distance to the considered point.

Optical system blurring

- The image formation model is a convolution btw the object and the PSF. It gives for example:



Restoration : inversion of the image formation model (blur + noise)



Notations, assumptions

$\Omega \subset \mathrm{R}^2$ Open bounded subset

Continuous variables: $u(x)$

$$u: \quad \Omega \to \mathrm{R}$$

$$x \to u(x)$$ Grey level at point $x=(x_1,x_2)$

$\Omega \subset \mathrm{N}^2$ Bounded subset of discrete points

Discrete variables : pixel i,j

$$\mathrm{u}_{i,j} = \mathrm{u}(i\Delta x, j\Delta y), i, j = 0 \ldots N$$

g :observed image, degraded from u

Discrete Fourier Transform (recall)

- let u be a discrete signal of finite support :

  $u_0, u_1, \ldots\ldots u_{N-1}$

- Its Dicrete Fourier Transform is

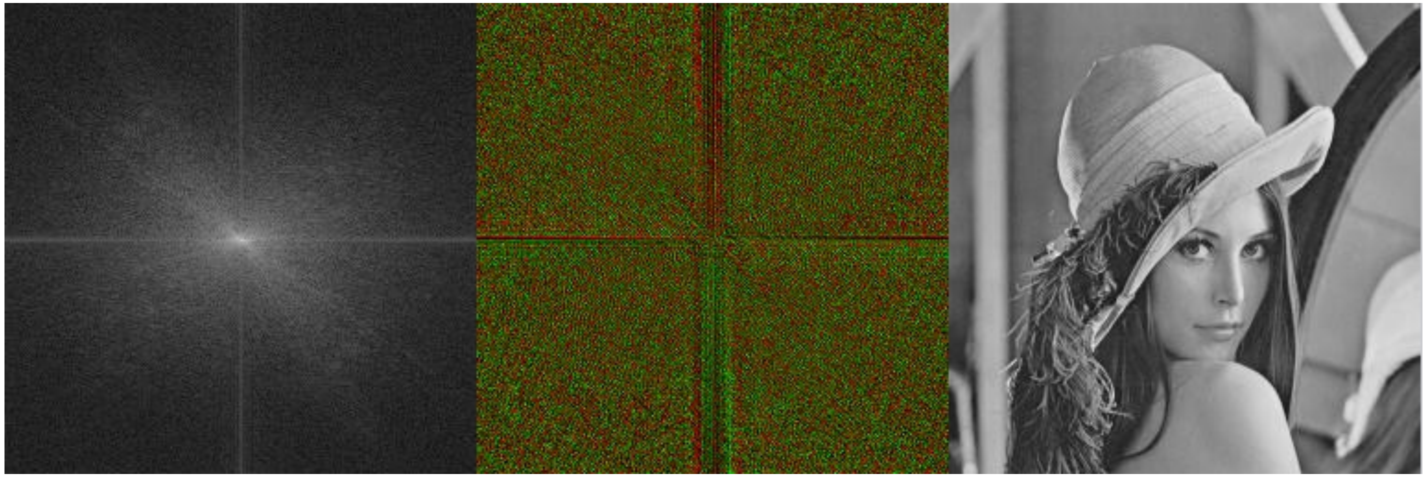
$$\hat{u}_k = \sum_{n=0}^{N-1} u_n \exp\left(\frac{-2i\pi kn}{N}\right)$$

- The Inverse Discrete Fourier transform is

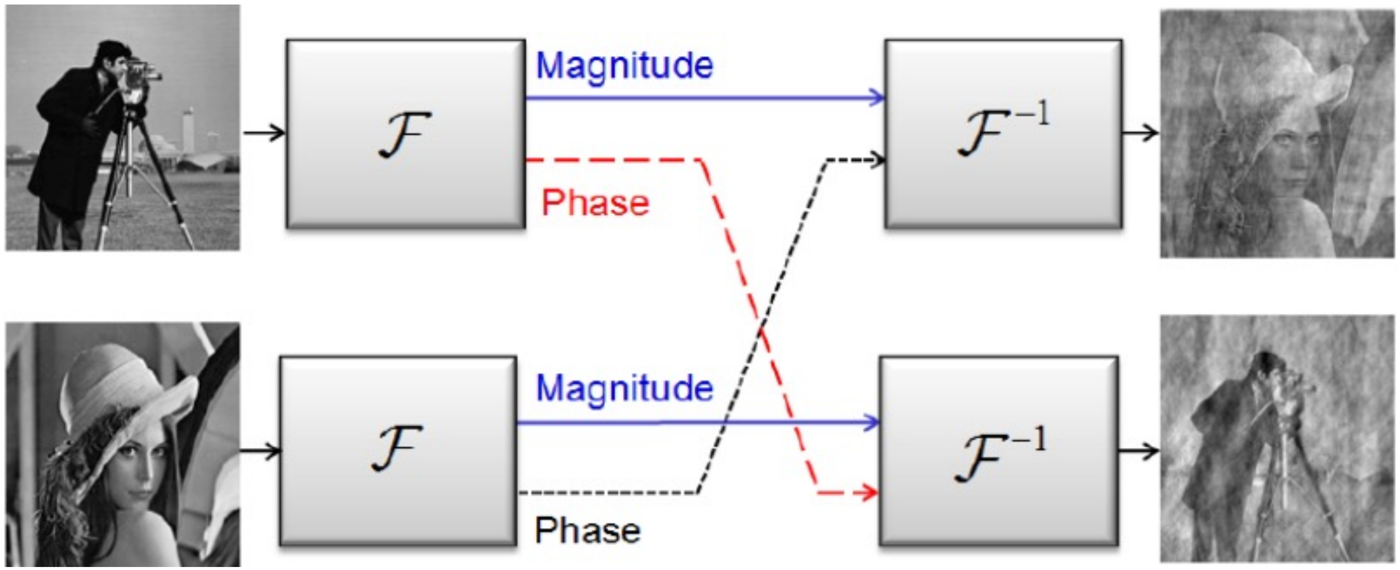
$$u_n = \frac{1}{N} \sum_{k=0}^{N-1} \hat{u}_n \exp\left(\frac{2i\pi kn}{N}\right)$$

- Fast algorithm : FFT

Discrete Fourier Transform: magnitude/phase



magnitude phase



Continuous Fourier Transform (recall)

- Let u be a function from $\mathrm{R}^d \to \mathrm{R}$ (for 2D images, d=2). We assume that $u \in L^1(\mathrm{R}^d)$, i.e.

$$\int_{R^d} |u(x)|\, dx < \infty$$

- The Fourier Transform (F) of $u \in L^1(\mathrm{R}^d)$, is the continuous function defined by

$$F(u) = \hat{u} \quad \text{and} \quad \forall \zeta \in R^d,\ \hat{u}(\zeta) = \int_{R^d} u(x) \exp\left(-i\left\langle \zeta, x \right\rangle\right) dx \qquad (1)$$

where $< \zeta, x >$ is the standard real scalar product $\left\langle \zeta, x \right\rangle = \sum_{i=1}^{d} \zeta_i . x_i$

$\hat{u}$ is continuous and $\hat{u}(\zeta) \to 0$ when $|\zeta| \to +\infty$

When $\hat{u} \in L^1$, we can retrieve the initial function u with the inverse Fourier transform

$$u(x) = \frac{1}{(2\pi)^d} \int_{R^d} \hat{u}(\zeta) \exp\left(i\left\langle \zeta, x \right\rangle\right) d\zeta \qquad (2)$$

Equations (1) and (2) are written in the L^1 sense, that is u (or $\hat{u}$) equals a.e. the continuous function defined by the right-hand side term.

Fourier Transform (recall)

- If $u \in S$ where S is the Schwartz space of functions $u \in C^\infty$ quickly decreasing that is $x^\alpha \partial^\beta u(x) \to 0$ when $|x| \to \infty$ $\forall (\alpha,\beta) \in N^2$, then $\hat{u} \in S$ too.

The Fourrier transform $F : u \to \hat{u}$ is an isomorphism of S and can be continuously extended to an isomorphism on L^2.

- Parseval : $\boxed{\|u\|_2 = \|\hat{u}\|_2}$

- Some properties: let u and v two functions in S, we have

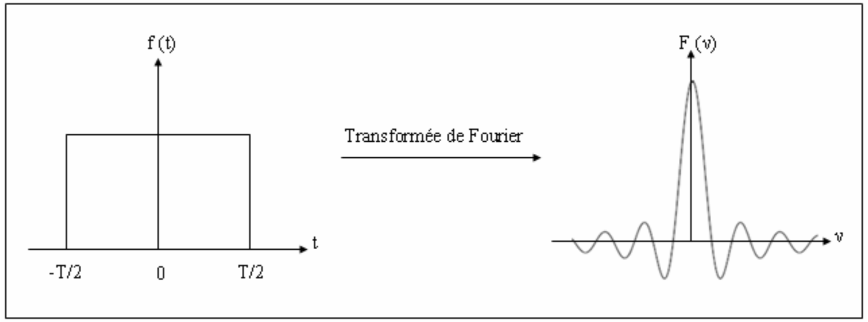
$$\boxed{\widehat{u * v} = \hat{u}.\hat{v}} \quad \text{and} \quad \widehat{u.v} = \frac{1}{(2\pi)^d}\hat{u} * \hat{v} \qquad (3)$$

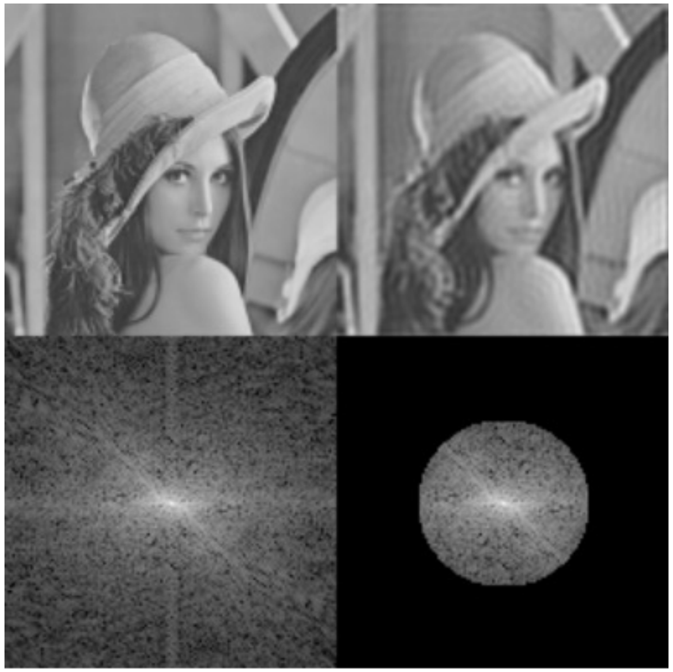
NB : making the change of variable $\zeta=2\pi f$ we have:

$$u(x) = \int_{R^d} \hat{u}(f)\exp\left(2i\pi\langle f, x\rangle\right) df$$

and $\boxed{\widehat{u.v} = \hat{u} * \hat{v}}$

Fourier Transform and convolution





spatial coordinates:
convolution by a sinus cardinal

frequency coordinates:
cut-off high frequencies

Spatial Dispersion: FTM

Under assumption of blur kernel is stationary and independent of the scene:

Blur = convolution

$$g(x,y) = (h * u)_{x,y} = \int_{\Omega} h(x-s, y-t)u(s,t)\,ds\,dt$$

In Fourier space

$$F(g)_{u,v} = F(h)_{u,v} . F(u)_{u,v}$$

From the scene to the image on the sensor:

Optical system, sensor integration

Discrete Convolution

- Let U a discrete signal (finite length) U(0)…U(N-1)
- Let h be the discrete PSF

$$g(k) = \sum_{n=0}^{N-1} h(n)u(k-n) = \sum_{n=0}^{N-1} h(k-n)u(n)$$

- With centered PSF h(-K)….h(K) $\quad g(n) = \sum_{k=-K}^{K} h(k)u(n-k)$

- We have for the PSF $\sum_{k=-K}^{K} h(k) = 1 \qquad h(k) \geq 0 \quad \forall k$

- Matrix/vector writting $\mathrm{g} = \mathrm{Hu}$

  H is a band Toeplitz matrix if boundary conditions are zero

  H is a circulante matrix if boundary conditions are periodic.

  In the circulante case H can be diagonalized by DFT.

Sensor

- **Integration**
  - Each sensor is an integrator. If we have a matrix of sensors, each one is modeled by a rectangular cell of size p_x x p_y : photosensible area, which are distributed in a grid p_{ex} , p_{ey}: pixel step size.

$$u_{k,l} = \int_{[-p_x,p_x]\times[-p_y,p_y]} u\left(kp_{ex} + x, lp_{ey} + y\right) dxdy$$

  - It is also a convolution, the following PSF:

$$PSF_{int} = \sum_{k,l} 1_{[-p_x,p_x]\times[-p_y,p_y]}\delta_{kp_{ex},lp_{ey}} \quad \text{et} \quad u_{k,l} = \left(PSF_{int} * u\right)\left(kp_{ex}, lp_{ey}\right)$$

  - Associated FTM is

$$\left(FTM_{int}\right)_{u,v} = \frac{\sin\left(\pi u \, {p_x}/{p_{ex}}\right)}{\pi u \, {p_x}/{p_{ex}}} \frac{\sin\left(\pi u \, {p_y}/{p_{ey}}\right)}{\pi u \, {p_y}/{p_{ey}}}$$

Spatial Dispersion

- *PSF* is (at least)

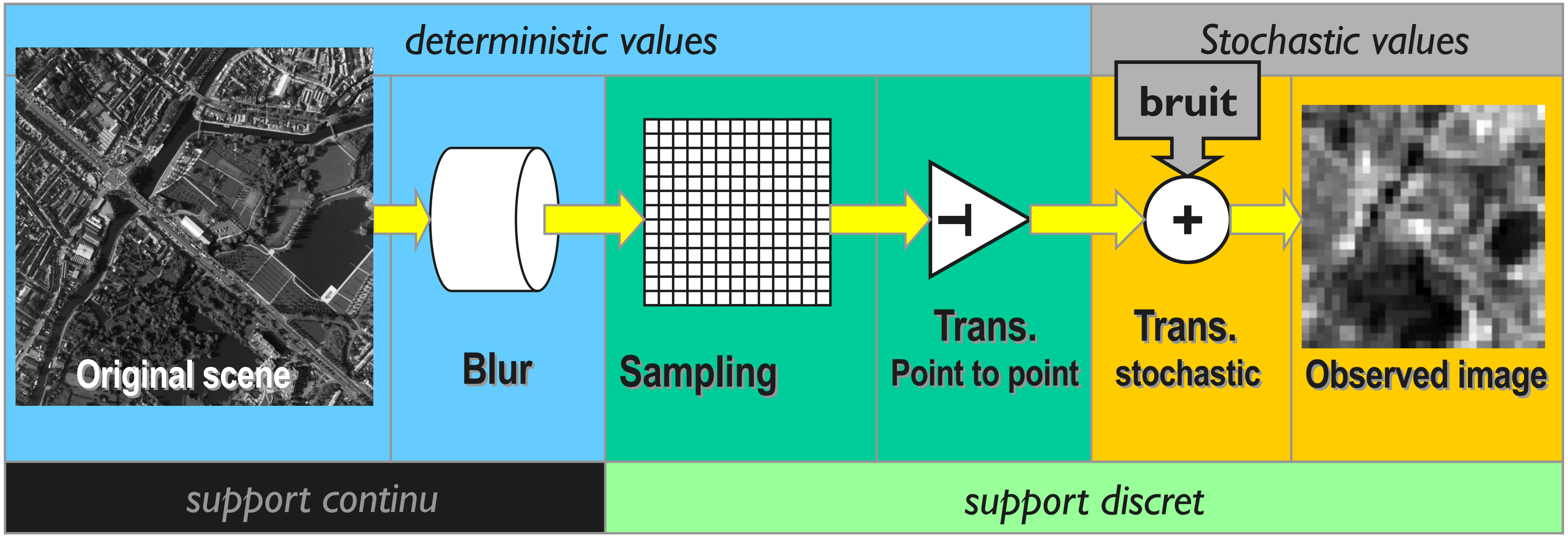
$$PSF = PSF_{optics} * PSF_{sensors}$$

Which correponds to the *FTM*

$$FTM = FTM_{optics} \,.\, FTM_{sensor}$$

- We use simplified models with few parameters, and usually PSF is approximated by a Gaussian function.

Observation model for satellite imaging



- Physical degradation (diffraction blur) is applied in the continuous setting, then the sensor integration makes the variables discrete. The true model is discrete/continuous:

$$g_{i,j} = (h_s * h_b * u)_{i,j} + n_{i,j}$$

$$u: \quad \Omega \subset R^2 \to \mathrm{R}$$
$$(x_1, x_2) \to u(x_1, x_2)$$

$$h_b: \quad \Omega \subset R^2 \to \mathrm{R}$$
$$(x_1, x_2) \to u(x_1, x_2)$$

$$h_s: \quad \Omega \subset R^2 \to \mathrm{R}$$
$$(x_1, x_2) \to h_s(x_1, x_2) = \sum_{k,l \in Z^2} \delta_{kp_{ex}, lp_{ey}} . 1_{[-p_x, p_x] \times [-p_y, p_y]}(x_1, x_2)$$

- Let X be a random sequence. If X models an image it is a random field $X = X_{i,j}$ $i,j=1,\ldots N$ Each $X_{i,j}$ is a random variable, which is characterized by its density probability, continuous or discrete, denoted $p_X(x,i,j)$
- Stationary assumption: the density probality is the same for all pixels: $p_X(x,i,j) = p_X(x)$ for all (i,j)
- Under stationary assumption, the mean and the variance are given by

$$m_X = \int_{x\in\mathrm{R}} x\, p_X(x)\, dx$$

$$\sigma_X^2 = \int_{x\in\mathrm{R}} (x - m_X)^2\, p_X(x)\, dx$$

Random sequences (stationay assumption)

- The correlation function or autocorrelation needs the joint probability densities

$$R_X(k,l) = E\left[X(i,j)X^*(i+k,j+l)\right] =$$
$$= \int_{x_1}\int_{x_2} x_1 x_2^* \, p_{XX}(x_1,x_2,k,l)\,dx_1\,dx_2, \qquad \forall i,j$$

- The covariance or autocovariance is defined by

$$C_X(k,l) = E\left[\left[X(i,j)-m_X\right]\left[X(i+k,j+l)-m_X\right]^*\right] \qquad \forall i,j$$
$$= \int_{x_1\in\mathrm{R}}\int_{x_2}\left[x_1-m_X\right]\left[x_2-m_X\right]^* p_X(x_1,x_2,k,l)\,dx_1\,dx_2$$

Let remark that $\sigma_X^2 = C_X(0,0)$

- Ergodicity : allows to identify mathematical expectations (over sets) with infinite spatial means. For a stationary random sequence, it means that

$$R_X(k,l) = \lim_{N\to+\infty} \frac{1}{N^2} \sum_{i,j=0}^{N} x(i+k, j+l)\, x^*(i,j)$$

We also have

$$m_X = \lim_{N\to+\infty} \frac{1}{N^2} \sum_{i,j=0}^{N} x(i,j)$$

Random sequences

- Correlation matrix of a real signal $X^T = (X_1, \dots X_N)$
  1D indexes, for 2D images, rank the 2D indexes in a 1D vector by lexicographic ordering (line by line)

$$E\left(X\,X^*\right) = \begin{pmatrix} E(X_1X_1) & E(X_1X_2) & . & . & E(X_1X_N) \\ E(X_1X_2) & E(X_2X_2) & & & \\ & & . & & \\ & & & . & \\ E(X_1X_N) & & & & E(X_NX_N) \end{pmatrix}$$

- In the stationay case:

$$E\left(X\,X^T\right) = \begin{pmatrix} R_X(0) & R_X(1) & . & . & R_X(N-1) \\ R_X(1) & R_X(0) & & & \\ & & . & & \\ & & & . & \\ R_X(N-1) & & & & R_X(0) \end{pmatrix}$$

They are symmetric Toeplitz matrices

Gaussian White Noise

- Let n be the noise. It is a multidimensional variable on a field of pixels (i,j) $i,j=1,\ldots N$.

- If n is a white noise, the random variables $n(i,j)$ are mutually independent, so uncorrelated variables. Then the autocorrelation matrix is diagonal. Moreover if we assume a stationary noise then the autocorrelation matrix written as

$$E\left(nn^t\right)=\sigma^2 Id_N$$

- If the noise is a white Gaussian noise with 0 mean, the joint density probabilities of all pixels is of Gaussian law $N(0_N, \sigma^2 Id_N)$ given by

$$P_n\left(n\right)=\frac{1}{\left[2\pi\right]^{\frac{1}{2}}\sigma^n}\exp-\frac{\left(n\right)^t\left(n\right)}{2\sigma^2}=\frac{1}{\left[2\pi\right]^{\frac{1}{2}}\sigma^n}\exp-\frac{\left\|n\right\|^2}{2\sigma^2}$$

Noise

Several sources of noise

- Quantum noise: electron accumulation, photon count, Poisson statistic.
- Thermal noise and acquisition noise: Gaussian statistic.
- Quantification noise: uniform, small variance wrt other noise sources.
- Compression noise : colored, correlated, non stationary. Difficult to take into account, considered as Gaussian noise in first approximation
- Transmission noise : loss of bits… Difficult to take into account.

Assumptions (realistic) : independence of noises between themselves and independence between pixels (white noise) and stationarity of the distribution (same law in each pixel).

Poisson Noise + Gaussian Noise $\rightarrow$ approximation by white additive Gaussisn noise with zero mean and variance which depends on the intensityt de variance $u_{i,j}$ at pixel (i,j)

$$P(n/u) = \prod_{i,j} \mathcal{N}_2\left(0, \left(A + Bu_{i,j}\right) Id\right)$$

Noise with stationary law and non stationary variance.

At high count rate (real optical scene, long time exposure), Poisson law tends to Gaussian law. The noise is then white Gaussian $\mathcal{N}(0,\sigma^2 Id)$

Gaussian noise assumption

The noise is additive : $g=h*u+n$

In each pixel i, n_i is a random variable with a Gaussian distribution :

$$P\left(n_i=\alpha_i\right)=\frac{1}{Z}\exp-\frac{\alpha_i^2}{2\sigma^2}$$

The random variable N³-dimensionnel $n=(n_1,n_2,,,,n_{N3})$ has a Gaussian distribution with zero mean parameter and variance σ^2 with independence between pixels:

$$P\left(n_1=\alpha_1,n_2=\alpha_2,\ldots,n_{N^3}=\alpha_{N^3}\right)=\prod_{i=1}^{N^3}P\left(n_i=\alpha_i\right)$$

$$P\left(n_1=\alpha_1,n_2=\alpha_2,\ldots,n_{N^3}=\alpha_{N^3}\right)=\frac{1}{Z'}\exp-\frac{\sum_{i=1}^{N^3}\alpha_i^2}{2\sigma^2}=\frac{1}{Z'}\exp-\frac{\left\|\alpha\right\|^2}{2\sigma^2}$$

Gaussian noise assumption

With the model $g=h*u+n$ the probability of observing g if I know that u is the result of the convolution of h with the specimen u is the likelihood:

$$P\left(g/\left(h*u\right)\right)=P\left(g-\left(h*u\right)=n/\left(h*u\right)\right)=P\left(g-\left(h*u\right)=n\right)$$

$$P\left(g/\left(h*u\right)\right)=\frac{1}{Z'}\exp-\frac{\sum_{i=1}^{N^3}\left[\left(h*u\right)-g\right]_i^2}{2\sigma^2}=\frac{1}{Z'}\exp-\frac{\left\|g-h*u\right\|^2}{2\sigma^2}$$

The Maximum Likelihood estimator of u is $\max_u P\left(g/\left(h*u\right)\right)$

Which is equivalent to $\min_u \left\|g-h*u\right\|^2$

Poisson density

The scalar random variable Y has a Poisson distribution with λ parameter

$$Y \propto P(\lambda) \quad \Leftrightarrow \quad P(Y = y / \lambda) = \frac{\lambda^y e^{-\lambda}}{y!}$$

This models the photon shot noise, which is a count measure, modeled by Poisson law.

The λ parameter is the mean and the variance.

For high count (high λ parameter), the Poisson law is well approximated by a Gaussian law

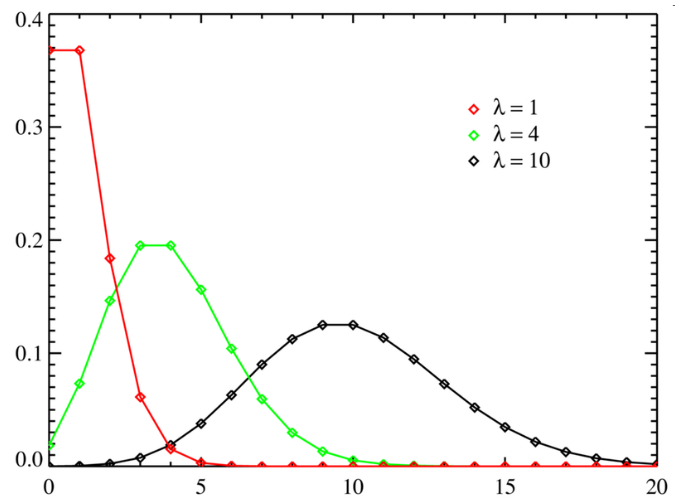
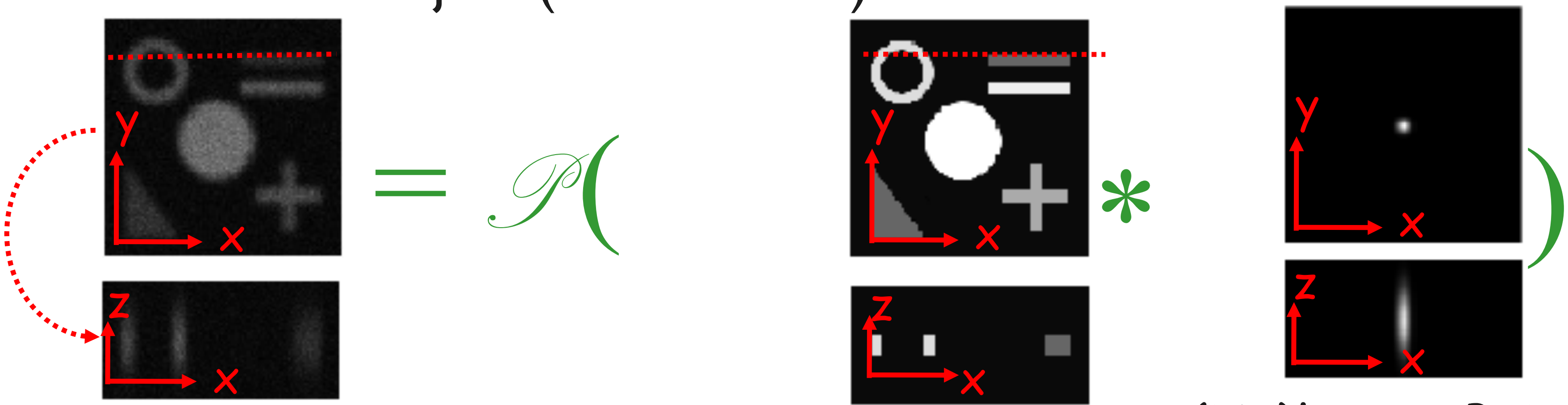


Image observation model (Poisson noise)

Simulated 3D object (128x128x64)



Observed image

Unknown true object

(Un)known Point Spread Function

$$g = \mathscr{P}(u * h)$$

- The image is blurred: the degradation is given by the Point Spread Function h.
- The image is noisy: the noise is usually photon noise, a term that refers to the inherent natural variation of the incident photon flux.

➢ Restoration Goal: Given the observation i, recover the object o

Poisson density

The random variable Y has a Poisson distribution with λ parameter which models the photon count noise :

$$Y \propto \mathsf{P}(\alpha) \quad \Leftrightarrow \quad P(Y = y / \alpha) = \frac{\alpha^y e^{-\alpha}}{y!}$$

Writing $g = \mathscr{P}(u*h)$ means that g has a Poisson distribution with parameter $u*h$. More precisely, in each pixel i, g_i has a Poisson distribution with parameter $(u*h)_i$

$$P\left(g_i / \left[u * h\right]_i\right) = \frac{[h * u]_i^{g_i} e^{-[h*u]_i}}{g_i\,!}$$

Poisson noise assumption

The Maximum Likelihood estimator of u is $\max_u P\left(g/\left(h*u\right)\right)$

In the Poisson noise case, due to independence between pixels:

$$P(g/u,h) = \prod_i \frac{[h*u]_i^{g_i}\, e^{-[h*u]_i}}{g_i\,!}$$

Which is equivalent to the minimization problem of
– log(P(g/u,h))

$$\boxed{\min_u \sum_i [(h*u)_i - g_i.\log(h*u)_i]}$$

Observation model: Gaussian noise case

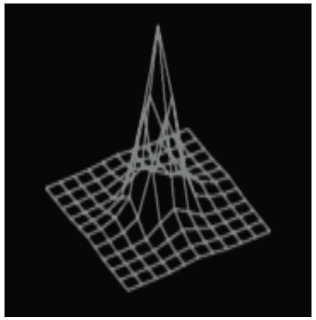
Observed images are degraded :

$$g = h * u + n$$

Noise
white Gaussian

Observed Image

Convolution kernel (PSF)



Original Image

- Restoration : retrieve u from g
- Inverse $g = h * u + n$ is an ill posed problem

Image Restoration

- Retrieve u, from the observed image $g = Hu+n$
- We assume that
  - Operator H is known,
  - Statistics (pdf, mean, standard deviation…) of the noise n are known

- Restoration = deconvolution problem.
- if $H=id$: denoising

- What is the difficulty of this inverse problem?

Difficulty of the inversion

- Assume that we are in the discrete setting, and that u have N points and u_0 have M points.
  - if $M>N$ we observe more points than the number of point we want to compute. The problem is over-determinate. The equations can be compatible or not. In any cases, select N equations among M or compute the least solution:

$$\underset{u}{Min}\left\|Hu-g\right\|^2$$

  - If $M<N$ we observe less points than the number of point we want to compute. The problem is under-determinate. Add constraints on the solution to select one solution among the set of solution.

  - If $M=N$ the H matrix is square (NxN). Then H can be invertible or not. If it is not invertible, with almost one null singular value, then the problem is again under-determinate. Add constraints on the solution to select one solution among the set of solution.

Difficulty of the inversion

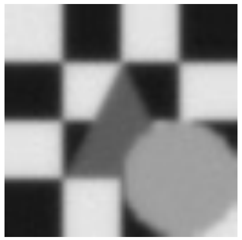
- Assume the matrix H is square ($M=N$) and invertible. Then we can compute

$$H^{-1}g = H^{-1}\left(Hu + n\right)$$

which gives the inverse solution

$$\hat{u} = u + H^{-1}n$$

Example



$\longrightarrow$

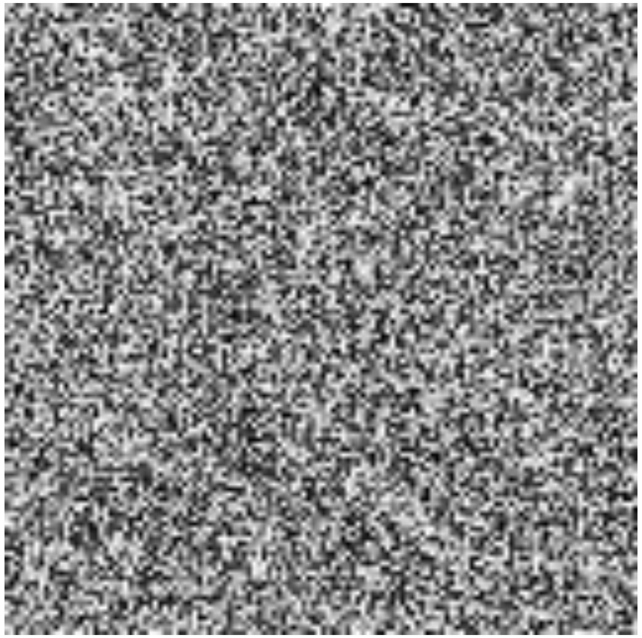
$\neq$

Blurred and noisy image u_0

Inverse solution

Original image

Inversion



Blurred and noisy image. Image given the space french agency which simulate the optics of the french satelitte SPOT5. Resolution 2,5m.
@CNES

Inverse solution

Inverse solution

- We have assumed that H is invertible in the mathematical sense, which means that the solution of an equation

$$v=Hu$$

  - Exists
  - Is unique
- We need one condition more to obtain an acceptable solution: the stability of the solution wrt the data, which means that u depends continuously on v, that is

for any sequence u_n such that $Hu_n \xrightarrow[n\to+\infty]{} Hu$

Then $u_n \xrightarrow[n\to+\infty]{} u$

Well-posed problem Hadamard 1923

Consider the equation

$$v = Hu \quad (1)$$

where u and $v : \Omega \subset R^2 \to R$ and $H : L^2(\Omega) \to L^2(\Omega)$

The inverse problem consists in finding u form a given v. This inverse problem is well-posed if the three following conditions are satisfied:

- Existence: for any v, we can find a u such that (1) is satisfied,
- Uniqueness: the solution *u* is unique,
- Stability: for any sequence u_n such that $\lim\limits_{n \to +\infty} Hu_n = Hu$

Then $\lim\limits_{n \to +\infty} u_n = u$

Deconvolution: an ill-posed problem

- Convolution: $h(x,y)=h(y-x)$
- Riemann-Lebesgue Lemma

If $h \in L^2(\Omega)$ then it can be shown that

$$\lim_{\alpha \to +\infty} \int_\Omega h(y)\sin(\alpha y)dy = 0$$

Then we have $$\lim_{\alpha \to +\infty} \int_\Omega h(x-y)\left[u(y)+\sin(\alpha y)\right]dy = v(x)$$

Any high frequency signal added to u leaves the integral unchanged.

The continuous problem is per se ill-posed.

Fourier analysis

- Back to the discrete problem.
- Circular discrete convolution (convolution with periodic boundary conditions) is a simple product in the Fourier plane.

$$(g)_{i,j} = (h * u)_{i,j} + (n)_{i,j}$$

$$\rightarrow \quad F(g)_{k,l} = F(h)_{k,l} . F(u)_{k,l} + F(n)_{k,l}$$

- Matrix-vector form: under periodic boundary conditions, the matrix H is circular block circular, so eigen vectors are the basis vectors of the 2D Fourier transform

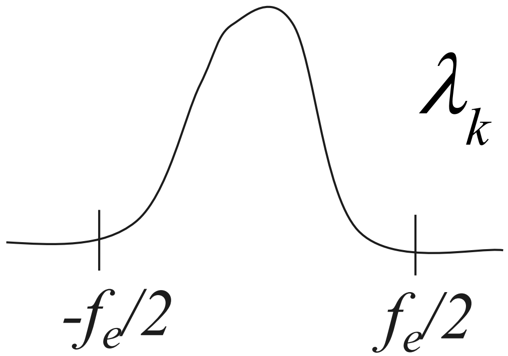
$$g = H u + n$$

$$\rightarrow \quad F(g) = diag\left\{\lambda_{k,l}\right\} . F(u) + F(n)$$

- The eigen values $\lambda_{k,l}$ are the coefficients of the 2D Fourier transform of the kernel h. So it is the MTF.

Fourier Analysis

- As h models a low frequency filter (blur), then the eigen values $\lambda_{k,l}$ of H matrix corresponding to high frequencies are small, may be zero.
- The PSF corresponding to blur attenuate high frequencies of the image (think to the Gaussian blur model)



- We want to inverse $F(u_0) = diag\{\lambda_{k,l}\}.\, F(u) + F(n)$
- If $\exists$ (k,l), $\lambda_{k,l}$ =0, the problem in $F(u)$ has an infinity of solutions
- If $\forall(k,$l), $\lambda_{k,l} \neq$0, the problem in $F(u)$ has a unique solution, but instable.

$$F(u)_{k,l} = \frac{F(g)_{k,l}}{\lambda_{k,l}} + \frac{F(n)_{k,l}}{\lambda_{k,l}}$$

Matrix Conditioning

- Let consider the matrix vector equation $v=Hu$
- Existence and uniqueness are ensured as soon as H is a square non singular matrix.
- Stability is measured by the condition number of the matrix.
- Definition : the condition number is defined, when H is regular, by

$$Cond(H) = \|H\|.\|H^{-1}\|$$

where $\|H\|$ is the matrix norm induced by the vector norm on R^n: $\|H\| = \sup_{x \neq 0} \left\{ \frac{\|Hx\|}{\|x\|} \right\}$

- Properties

$$Cond(H) \geq 1, \qquad Cond(H) = Cond(H^{-1}),$$

$$Cond(I) = 1, \qquad Cond(\lambda H) = Cond(H) \quad \text{for } \lambda \neq 0,$$

with the Euclidian norm, $Cond(H) = \frac{|\mu_{\max}|}{|\mu_{\min}|}$ μ_i : singular values of H

if H is normal $Cond(H) = \frac{|\lambda_{\max}|}{|\lambda_{\min}|}$ λ_i : eigen values of H

What *Cond(H)* measures?

- Let consider the matrix vector equation $v=Hu$, and let δv be a pertubation on v. δv leads to a perturbation δu in u such that

$$v+\delta v=H(u+\delta u)$$

We have $\delta v=H\delta u$ so $\delta u=H^{-1}\delta v$ and we can deduce

$$\|\delta u\|\leq\|H^{-1}\|.\|\delta v\|$$

But we also have $\|v\|\leq\|H\|.\|u\|$ .Then for non null vectors u,v we have

$$\frac{\|\delta u\|}{\|u\|}\leq\|H\|.\|H^{-1}\|\frac{\|\delta v\|}{\|v\|}$$ also wirtten as $$\boxed{\frac{\|\delta u\|}{\|u\|}\leq Cond(H)\frac{\|\delta v\|}{\|v\|}}$$

Then a small condition number (near 1) will ensure stability because a small relative perturbation on the observed data v will produce a small relative perturbation on the solution u

Least square solution

- The least square solution is given by the resolution of the optimisation problem

$$\underset{u \in L^2(\Omega)}{Inf} \int_{\Omega} \left| g - Hu \right|^2 dx$$

If the operator H is such that Ker(H)={0}, then it exists a unique solution, given by the Euler equation

$$\begin{cases} H^*\left(Hu - g\right) = 0 \\ \left.\dfrac{\partial u}{\partial \mathrm{N}}\right|_{\partial\Omega} = 0 \end{cases}$$

H is a linear operator, H^* is its adjoint

The solution can be computed by solving the associated dynamical system, where u is now depending on time t (equivalent to gradient descent with fixed iteration step)

$$\begin{cases} \dfrac{\partial u}{\partial t} = H^*\left(g - Hu\right) \\ \left.\dfrac{\partial u}{\partial \mathrm{N}}\right|_{\partial\Omega} = 0\,, \quad u\left(x, t = 0\right) = g\left(x\right) \end{cases}$$

- Of course if the inverse solution is unstable, so is the least square solution.

Least square solution

- In discrete variables: $\underset{u}{Min}\left\|Hu-g\right\|^2=\underset{u}{Min}\sum_{i,j=1}^{N}\left(Hu-g\right)_{i,j}^2$

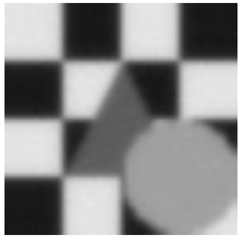
If the matrix H has non null singular values then the problem has a unique solution, given by the resolution of the linear system

$$\left(Hu-g\right)=0 \quad \text{and the solution is} \quad \hat{u}=\left(H^*H\right)^{-1}H^*g$$

- Boundary conditions are included in the construction of the matrix H
- Of course if H is ill-conditioned, so is H^*H
- In the frequency domain, MTF is the Modulation Transfer Function $MTF = F(h)$ if H admits an inverse then it has non null eigen values, so $(MTF)_{k,l}\neq 0$, and the least square solution is given by

$$F\left(\hat{u}\right)=\frac{MTF^*}{\left|MTF\right|^2}F\left(g\right)$$

Least square solution

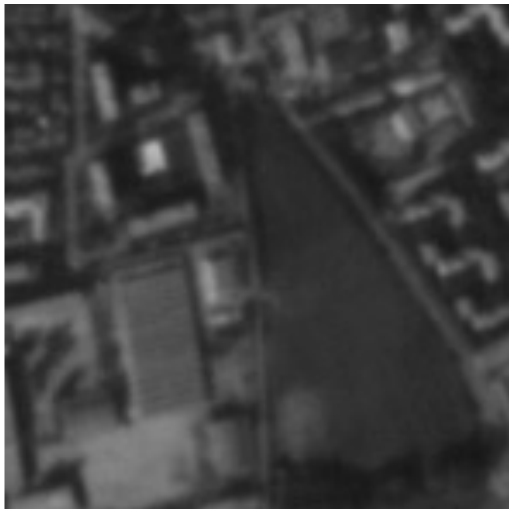


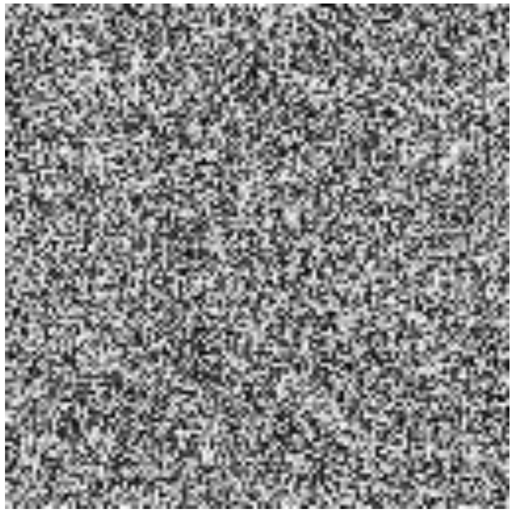
Blurred and noisy image u_0

Least square solution

$\neq$

Original image





Regularisation of ill-posed inverse problems

- Principle :

- do not try to inverse H but search for a solution in a set of admissible solutions:

$$\left\{ u \,/\, \frac{1}{N^2} \| g - Hu \|^2 \leq \sigma^2 \right\}$$

Where N^2 is the number of discrete points and σ^2 the variance of the Gaussian noise

- Reduce this set of solutions by introducing constraints. The goal is to obtain the stability of the inversion i.e a solution close to the ideal solution.

- Basic methods of regularisation are

- TSVD « Truncated Singular Value Decomposition
- stopping iterative inversion methods
- projection on constraints (positivity…)
- a priori model introduction

Tikhonov Regularisation

- Search for a solution in the set of admissible solutions (continuous setting): $\left\{ u \,/\, \frac{1}{|\Omega|} \int_{\Omega} (g - Hu)^2 \, dx \le \sigma^2 \right\}$

- Reduce this set by introducing a constraint in order to obtain a regular solution $\underset{u}{Min} \int_{\Omega} |\nabla u|^2 \, dx$

- ∇u is the gradient of u : $\nabla u(x) = \left(u_{x_1}(x), u_{x_2}(x)\right)$

- Solution can be computed by minimizing a penalized criterion

$$J(u) = \int_{\Omega} |g - Hu|^2 \, dx + \lambda \int_{\Omega} |\nabla u|^2 \, dx$$

Data term **Regularisation term**

- λ is the regularisation parameter, which compounds the influence between the two terms.

Euler-Lagrange Equation

$$J\left(u\right)=\int_{\Omega}\left[\left(g-Hu\right)\left(x\right)\right]^{2}dx+\int_{\Omega}\left|\nabla u\left(x\right)\right|^{2}dx$$

J is convex, the infimum exists (if Ker{H} does not contain constant images) and satisfies $J'\left(u\right)=0$

Calculus of $J'\left(u\right)$

$$\lim_{\varepsilon\to 0^{+}}\frac{J\left(u+\varepsilon v\right)-J\left(u\right)}{\varepsilon}=\left\langle J'\left(u\right),v\right\rangle$$

For the second term of J , we use the Green formula (integration by part):

$$\int_{\Omega}\left\langle \nabla u,\nabla v\right\rangle dx=-\int_{\Omega}v\Delta u\,dx+\int_{\partial\Omega}v\frac{\partial u}{\partial n}ds$$

Equation d'Euler-Lagrange

$$J(u) = \int_\Omega \left[\left(g - Hu \right)(x) \right]^2 dx + \int_\Omega \left| \nabla u(x) \right|^2 dx$$

J est convexe, l'infimum existe et vérifie $J'(u) = 0$

soit

$$\boxed{\begin{cases} H^*(Hu - g) - \lambda \Delta u = 0 \\ \left. \dfrac{\partial u}{\partial n} \right|_{\partial\Omega} = 0 \end{cases}}$$

$$\begin{aligned} H : X &\to X \\ u &\to Hu = h * u \end{aligned}$$

$$\begin{aligned} H^* : X &\to X \\ u &\to H^* u = h^* * u \quad \text{où } h^*(x) = \overline{h}(-x) \end{aligned}$$

Regularisation (Tikhonov)

- If H does not annihilate the constants then the minimization problem has a unique solution in $H^1(\Omega)$., which is computed by solving the Euler equation

$$\begin{cases} H^*(Hu-g)-\lambda\,\Delta u=0 \\ \left.\dfrac{\partial u}{\partial n}\right|_{\partial\Omega}=0 \end{cases}$$

where Δ is the laplacian of u: $\Delta u = u_{x_1x_1} + u_{x_2x_2}$

- Dynamical scheme; $u(i,j,t)$ $\longrightarrow$ $\begin{cases} \dfrac{\partial u}{\partial t}=H^*(g-Hu)+\lambda\,\Delta u \\ \left.\dfrac{\partial u}{\partial n}\right|_{\partial\Omega}=0 \end{cases}$

- Isotropic diffusion (Laplacian, heat equation)
- Linear regularisation, stationnary process, edges of images are not reconstructed.
- ~ Wiener filter

$$\hat{u}=\left(H^*H-R_{uu}^{-1}R_{nn}\right)^{-1}H^*u_0 \quad\longleftrightarrow\quad \hat{u}=\left(H^*H-\lambda\,\Delta\right)^{-1}H^*g$$

Result

Original image @CNES

Blurred and noisy image

(@CNES, simulation SPOT5)

Restored image

Minimisation of regularized criterion

$$\min_{u \in X \cap X_J} \left\| g - Hu \right\|_q^q + J(u)$$

Data term

Norm or semi-norm in a regularizing space

- What are the right regularizing spaces adapted to the images?
- What is the right data term?
- Existence, uniqueness of a solution ?
- Minimization algorithm?
- Evaluation of the numerical results ?

Linear/non linear regularization

- l_2 Regularisation smooth contours

$$\|g - Hu\|_2^2 + \lambda \|\nabla u\|_2^2$$

- Non linear Regularization
  - l_2/l_1 Regularisation

$$\|g - Hu\|_2^2 + \lambda \int_\Omega \varphi(|\nabla u|) dx$$

  - Regularisation by Total Variation

$$\|g - Hu\|_2^2 + \lambda \|\nabla u\|_1$$

- Regularization by wavelet transform (WT) ?
  - Regularisation in the wavelet domain

$$\|g - Hu\|_2^2 + \lambda \sum_{i,j=1}^{M} \left| \langle u, \psi_{i,j} \rangle \right|$$

  -
  - Regularization in a dictionary of waveforms

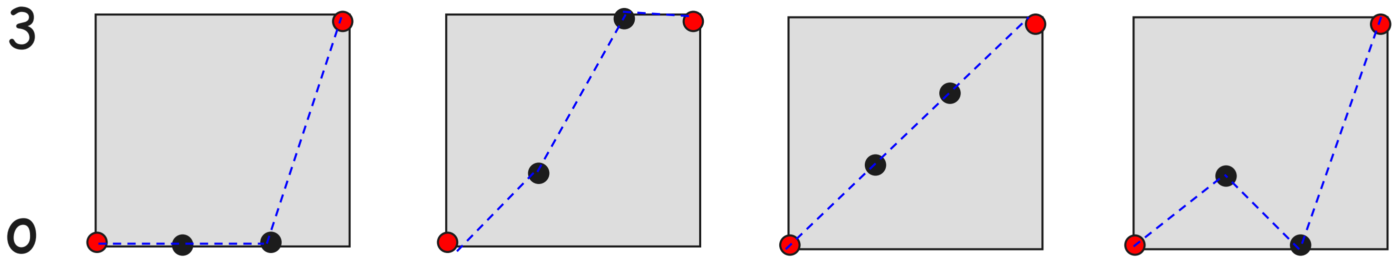
Norm l^1 and l^2

- Reduce the weight of high gradients in the minimization process. In discrete variables:
  - remplace l_2 norm by l_1 norm

$$E(\mathrm{u}) = \|\mathrm{g} - \mathrm{Hu}\|_2^2 + \lambda \|\nabla \mathrm{u}\|_1$$

$$\|\nabla \mathrm{u}\|_1 = \sum_{i,j} |\nabla \mathrm{u}|_{i,j} \quad \textbf{avec} \quad |\nabla \mathrm{u}|_{i,j} = \sqrt{(\mathrm{u}_{i,j} - \mathrm{u}_{i-1,j})^2 + (\mathrm{u}_{i,j} - \mathrm{u}_{i,j-1})^2}$$

- Norms l_2 and l_1: a small example in1D. $\|\nabla u\|_1 = \sum_i |u_i - u_{i-1}|$



$\|.\|_2^2=9$	$\|.\|_2^2=5$	$\|.\|_2^2=3$	$\|.\|_2^2=11$
$\|.\|_1=3$	$\|.\|_1=3$	$\|.\|_1=3$	$\|.\|_1=5$

$$\|\nabla u\|_2^2 = \sum_i |u_i - u_{i-1}|^2$$

$$\|\nabla u\|_1 = \sum_i |u_i - u_{i-1}|$$

l_1 Minimisation Total Variation

- Criterion to minimize is

$$J(u) = \left\| g - Hu \right\|_2^2 + \left\| \nabla u \right\|_1$$

- Considering the l_1 norm rather the l_2 norm as for Tikhonov

- *Total Variation (*TV) regularization is wery well-knomn and used in image processing.

$$\left\| \nabla u \right\|_1 = \sum_{i,j} \left| \left(\nabla u \right)_{i,j} \right|$$

l_1 Regularization

- Algorithm in the discrete setting : search for u minimizing $J(u)$

$$J(\mathrm{u}) = \|\mathrm{g} - \mathrm{Hu}\|_2^2 + \lambda \|\nabla \mathrm{u}\|_1 = \sum_{i,j} \left(g_{i,j} - (Hu)_{i,j} \right)^2 + \lambda \sum_{i,j} \left| (\nabla u)_{i,j} \right|$$

$$(\nabla u)_{i,j} = \left((\nabla u)_{i,j}^1, (\nabla u)_{i,j}^2 \right)$$

$$(\nabla u)_{i,j}^1 = \begin{cases} u_{i+1,j} - u_{i,j} & \text{si } i < N \\ 0 & \text{si } i = N \end{cases}$$

$$(\nabla u)_{i,j}^2 = \begin{cases} u_{i,j+1} - u_{i,j} & \text{si } j < N \\ 0 & \text{si } j = N \end{cases}$$

Discrete variables

- Discrete recommended modulus

$$\left|\nabla u\right|_{i,j} = \sqrt{\left(u_{i+1,j} - u_{i,j}\right)^2 + \left(u_{i,j+1} - u_{i,j}\right)^2}$$

- Difficulty: no differentiability in 0
- Formally the Euler is

$$H^*(Hu - g) - \lambda^2 div\left(\frac{1}{|\nabla u|}\nabla u\right) = 0$$

$$\frac{\partial u}{\partial n} = 0$$

- Anisotropic diffusion

Algorithms

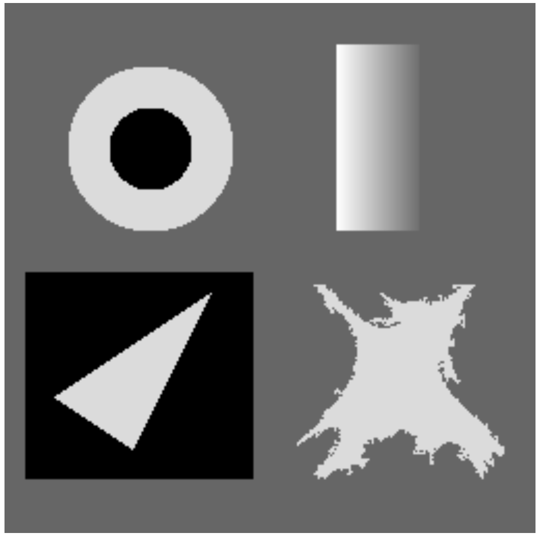
- Non-smooth convex optimization: a lot of algorithms proposed these last few years.
  - Rudin-Oscher-Fatemi (ROF) algorithm Physica D, 1992
  - Projection methods, proximal algorithms [Combettes-Wajs Multiscale Modeling & Simulation 2005]
  - Primal-Dual [Chambolle-Pock Journal of Mathematical Imaging and Vision 2011]
  - Interior Point method [Boyd-Vandenberghe. Cambridge university press, 2004]
  - Alternating Direction Method of Multipliers (ADMM) [Glowinski- Le Tallec, SIAM Studies in Applied and Numerical Mathematics1989.]
  - Fast algorithms…

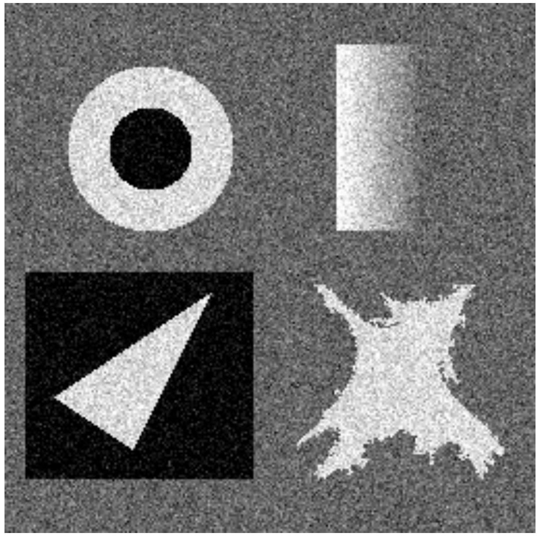
A good review in https://arxiv.org/abs/1412.4237

First Order Algorithms in Variational Image Processing Martin Burger, Alex Sawatzky, and Gabriele Steidl, 2014

Example

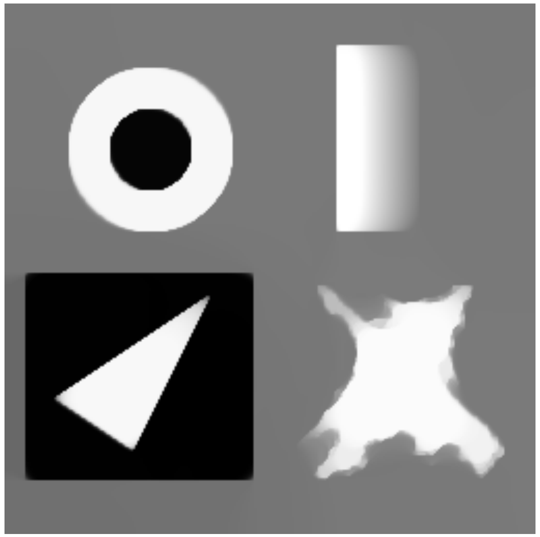
original





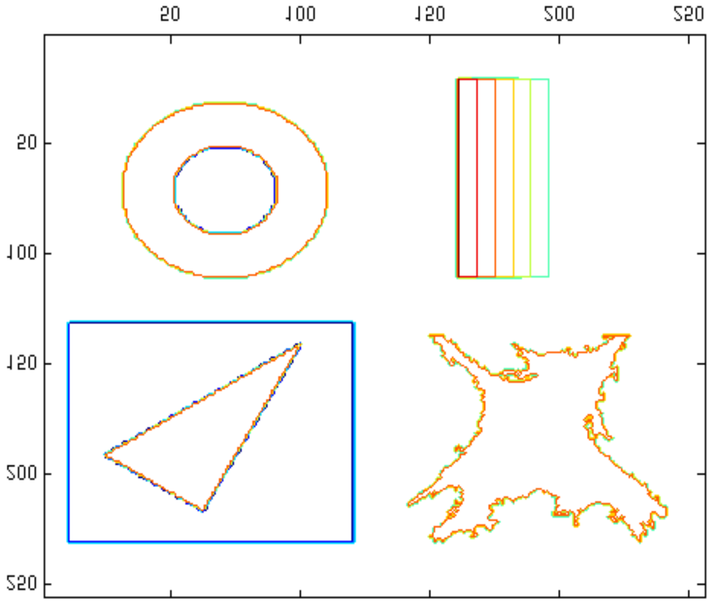
noisy

regularisation l_2

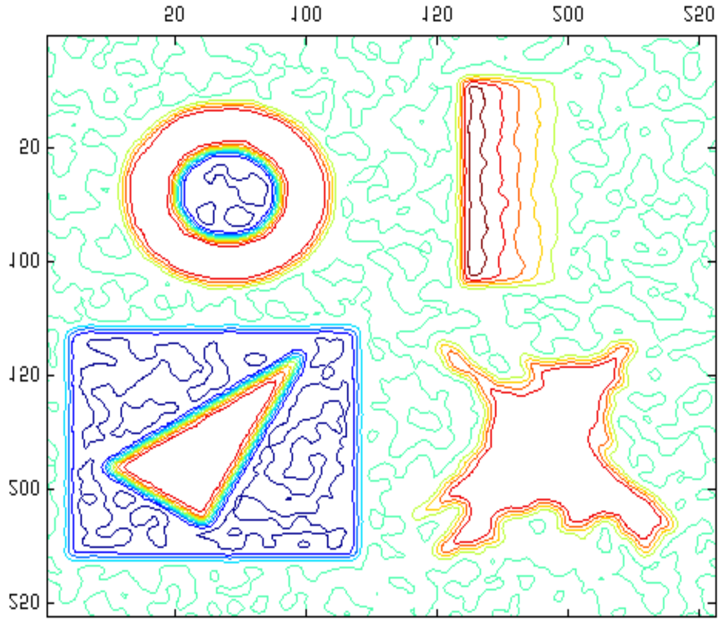


regularisation l_1

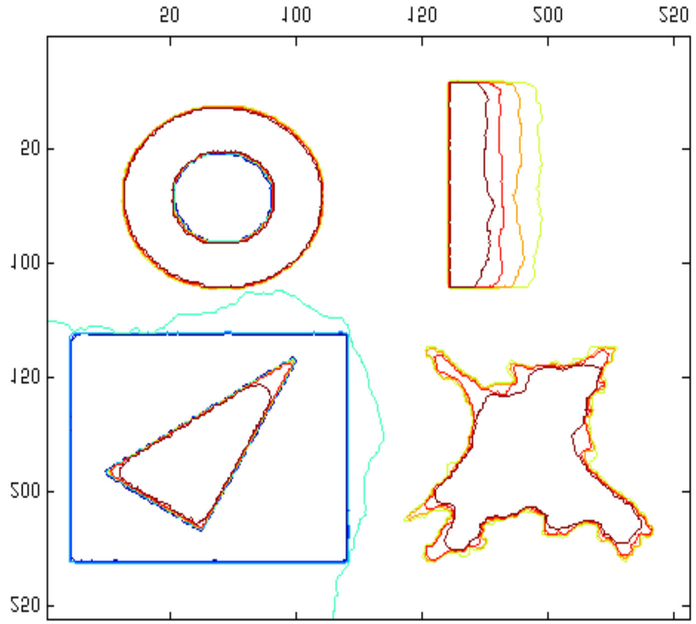
Level lines



original



regularisation l_2



regularisation l_1

Other regularization

- l_2 Regularisation smooth contours

$$\|g - Hu\|_2^2 + \lambda \|\nabla u\|_2^2$$

- Non linear Regularization
  - l_2/l_1 Regularisation

$$\|g - Hu\|_2^2 + \lambda \int_\Omega \varphi(|\nabla u|) dx$$

  - Regularisation by Total Variation

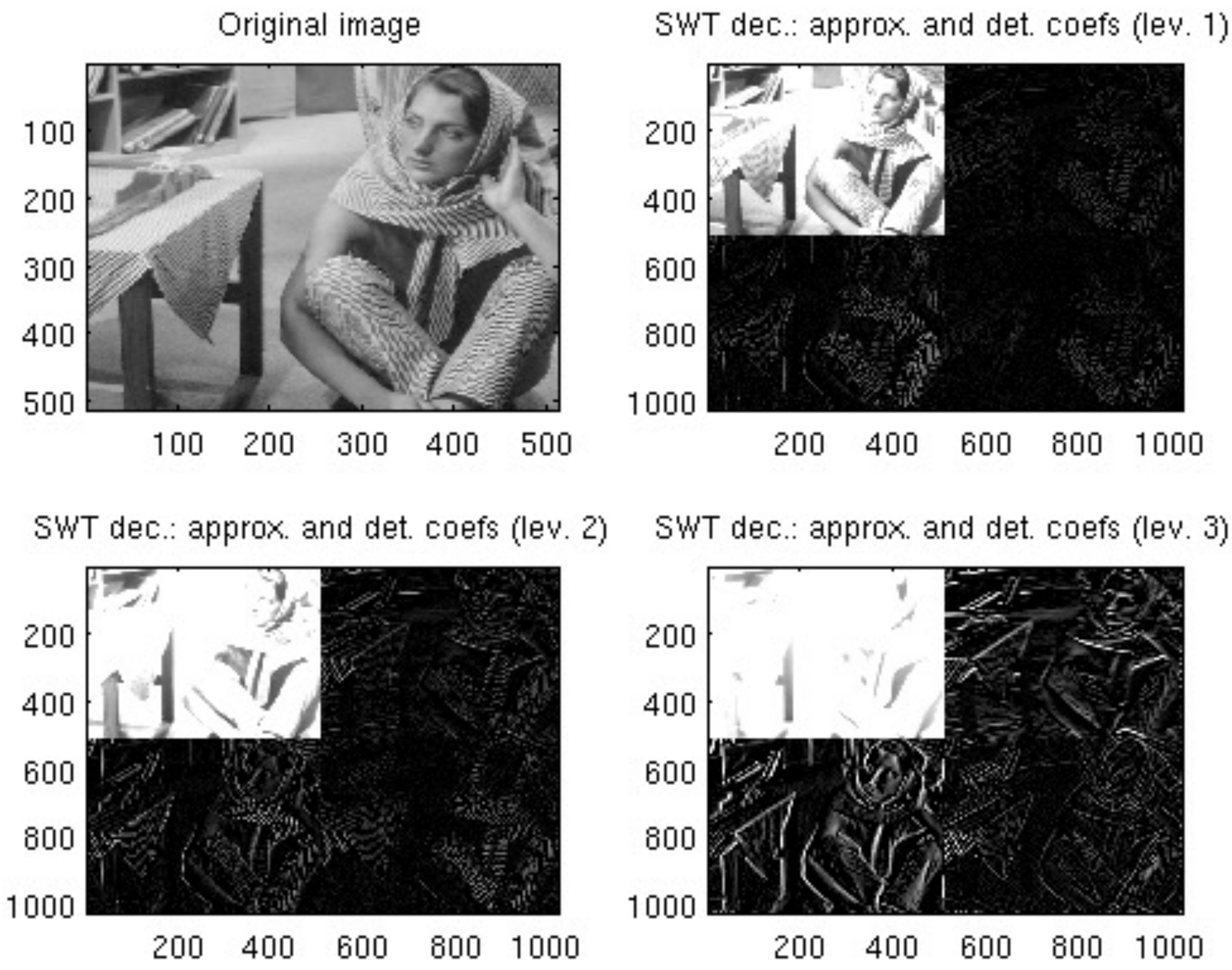
$$\|g - Hu\|_2^2 + \lambda \|\nabla u\|_1$$

- Regularization by wavelet transform (WT) ?
  - Regularisation in the wavelet domain

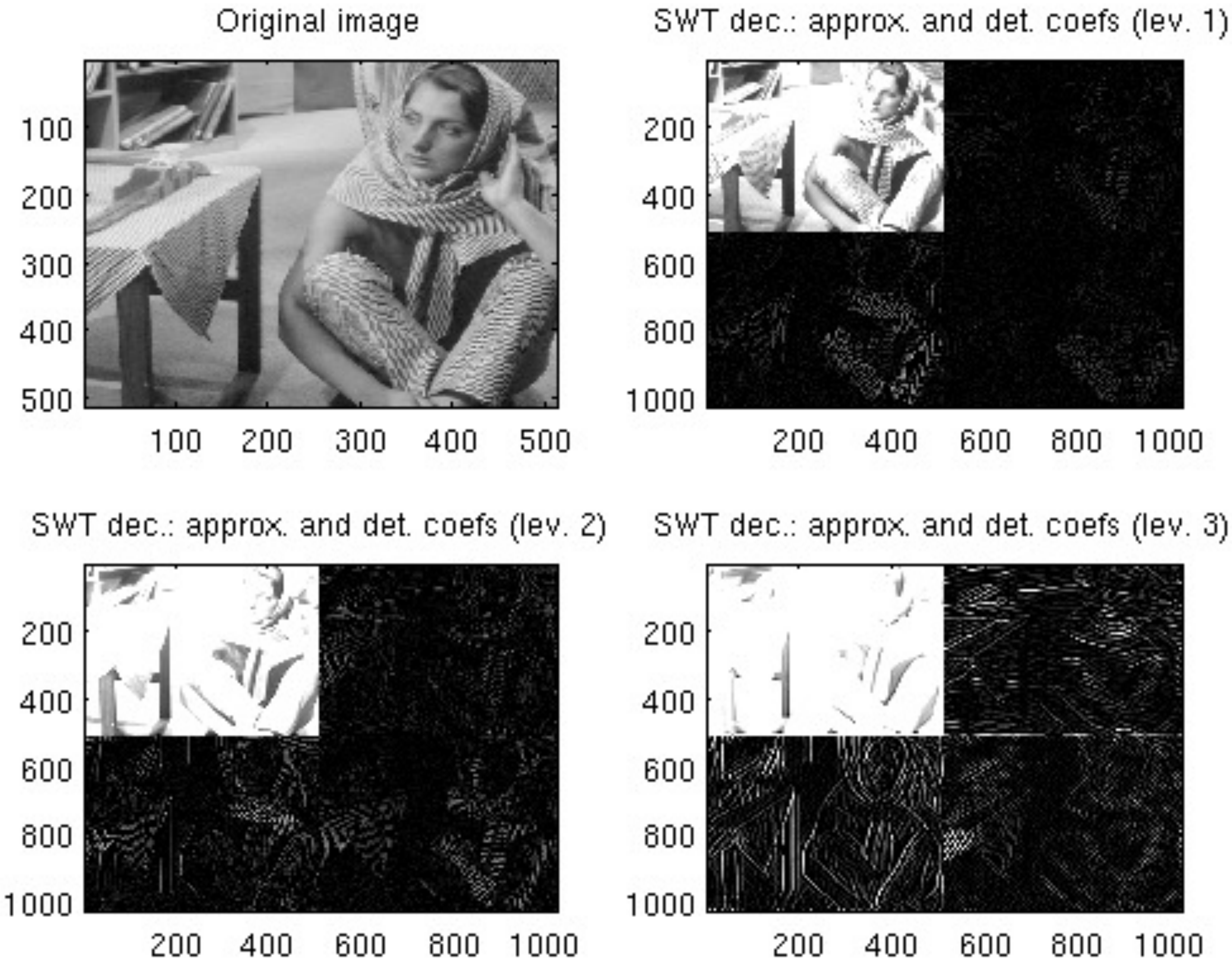
$$\|g - Hu\|_2^2 + \lambda \sum_{i,j=1}^{M} \left|\langle u, \psi_{i,j} \rangle\right|$$

  - Regularization in a dictionary of waveforms

A 2D decomposition (Haar)



A 2D decomposition (Symlet)



Why WT for denoising?

- WT gives a sparse representation of images, that is
  - few coefficients c_j with high values,
  - Lot of coefficients around 0.

- Then, when dealing with a noisy image, the few coefficients with high values will be above the noise.
- The rest of the coefficients with small values will be set to zero.

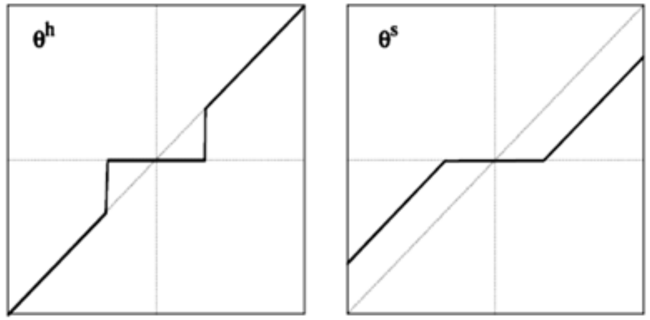
Connection with the variational approach

- Wavelet domain: compact representation of the signal

$\rightarrow$ Find u minimising

$$\|g-u\|^2+\lambda\,\#_{\{TO(u)\neq 0\}} \quad \Leftrightarrow \quad \|TO(u)-TO(g)\|^2+\lambda\,\#_{\{TO(u)\neq 0\}}$$

Solution : $u_{opt} = \theta_\lambda(g)$ hard thresholding.



$\rightarrow$ Find u minimising

$$\|u-g\|^2+\lambda\,|TO(u)| \quad \Leftrightarrow \quad \|TO(u)-TO(g)\|^2+\lambda\,|TO(u)|$$

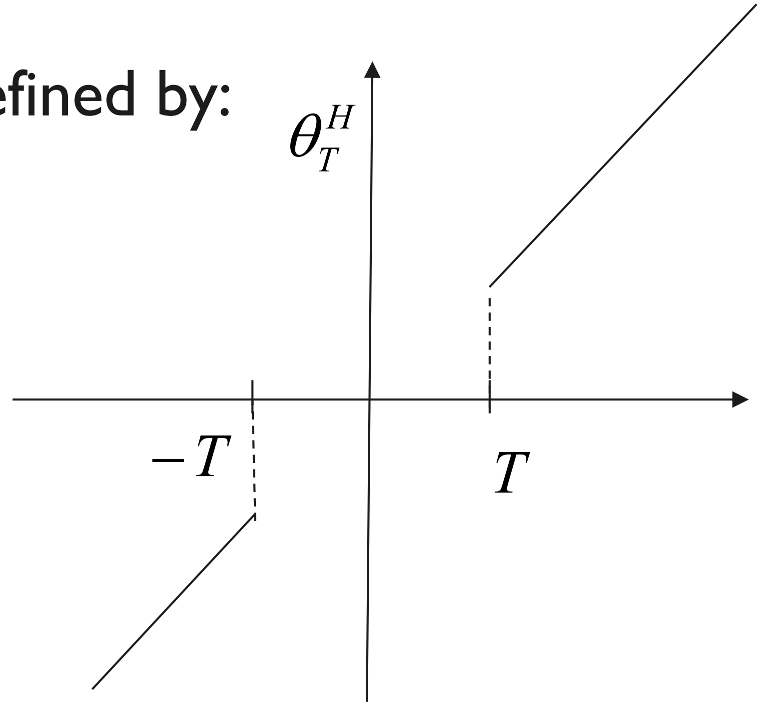
Solution : $u_{opt} = \theta_\lambda(g)$ seuillage soft thresholding.

- Related problem : image decomposition in a dictionary of atoms. Then it is no more a decomposition on a basis but on a dictionary. This dictionay can be composed of a union of bases.

Hard Thresholding

- The hard thresholding function is defined by:

$$\theta_T^H(t) = \begin{cases} t & \text{if } |t| > T \\ 0 & \text{if } |t| \le T \end{cases}$$

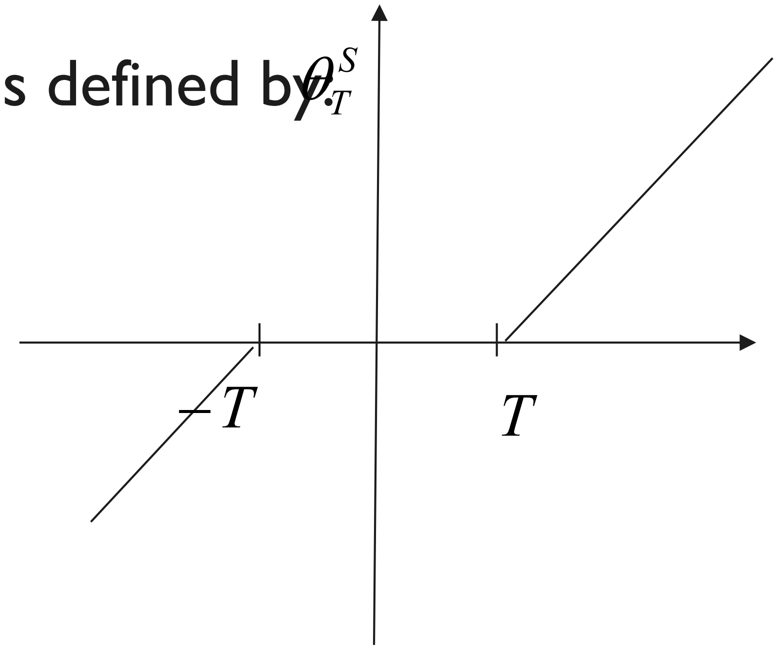


- When B is a wavelet basis, the hard thresholding preserves high coefficients which are above the noise level. These coefficients are on contours of the image Then there is no smooting of the edges.
- Small coefficients (smaller than T) are set to 0: in areas where varirations of the image are small, by setting the wavelet coefficients to 0, a local mean is done.

Soft thresholding

- The soft thresholding function is defined by:

$$\theta_T^S(t)=\begin{cases} t-T & \text{if } t>T \\ t+T & \text{if } t<-T \\ 0 & \text{if } |t|\le T \end{cases}$$



- This function is more regular than the hard thresholding function
- Reduction by T (pour T choisi selon le critère de Donoho Johnstone, voir page suivante) of coefficient magnitude allows to ensure that the magnitude of estimated coefficient are smaller than the ones of the noisy signal. Then the thresholded signal is as least as regulare than the original signal. This is not true for the hard thresholding.

Deconvolution and l^1 regularization

$\psi : R^n \rightarrow R^N$ is a wavelet basis (N=n) or a redundant wavelet transform (n<N) or a serie of wavelet basis (dictionary of waveform)

Analysis Problem (AP)

$$\min_{u \in R^n} \left\{ \left\| g - Hu \right\|_2^2 + \lambda \left\| \psi\, u \right\|_1 \right\}$$

Synthesis Problem (SP)

$$\min_{c_u \in R^N} \left\{ \left\| g - H\psi^* c_u \right\|_2^2 + \lambda \left\| c_u \right\|_1 \right\}$$

- Not equivalent problems, except if ψ is a basis. Larger dimension of the unknown space in the synthesis problem when ψ is redundant wavelet transform or a dictionary of shapes.
- See algorithms for the SP problem in the course dictionary optimisation L1 L0
- Algorithms for AP problem: primal/dual; ADMM among others

Deconvolution results

- Lena image

blurred image

blurred+noisy

regul. par ondelettes (Base d’ondelettes : pas de redondance: non invariance par translation)

Regularization using wavelets

$$\min_{u} \|g - Hu\|_2^2 + \lambda \|\Psi u\|_1$$

Ψ is a redundant basis in order to have translation invariance (frame rather than a basis)

Blurred and noisy image

Restored Image using dual tree complex wavelets (N. Kingsbury)

Sparse model

$$\min_{c_u} \left\| g - H\Psi^* c_u \right\|_2^2 + \lambda \left\| c_u \right\|_1$$

Ψ is a dictionary. Minimization can be done in the wavelet domain (SP).

Blurred and noisy image

Restored Image using dual tree complex wavelets (N. Kingsbury)

Régularisation par TO ou par TV?

- Pour la restauration de $g=Hu+n$ on considère la fonctionnelle à minimiser

$$\|g - Hu\|^2_{L^2(\Omega)} + \lambda \|u\|_Y$$

$\|.\|_Y$ est la norme de l'approximation u de g dans un espace régularisant Y : BV, $B^1_{1,1}$ …

- Pas d'espace optimal, dépend du contenu de l'image
  - TV lisse les oscillations reconstruit bien les contours et permet interpolation et extrapolation spectrale.
  - TO reconstruit bien les détails et les textures
- Modèles ultérieurs pour le débruitage
  - décomposition sur des trames, X-lets [,…]
  - décomposition sur un dictionnaire [Stark-Elad-Donoho05,…]
  - Régularisation mixte l^1 , par exemple TV + TO(s)

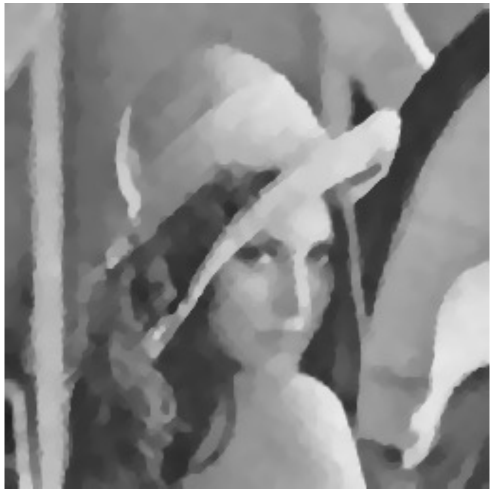
Deconvolution results

- Lena image

blurred image

blurred+noisy

Wavelet regul.



TV regul.

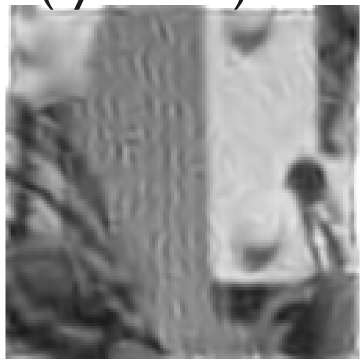
Wavelet + TV regul.

Comparaison TV, wavelet and sparsity

Wavelets are complex wavelets on N. Kingsbury's dual shaft, it is a frame and has a base, there is redundancy. This transform is practically invariant by translation and rotation

Image

Sparse model (synthesis)



Wavelet model (analysis)

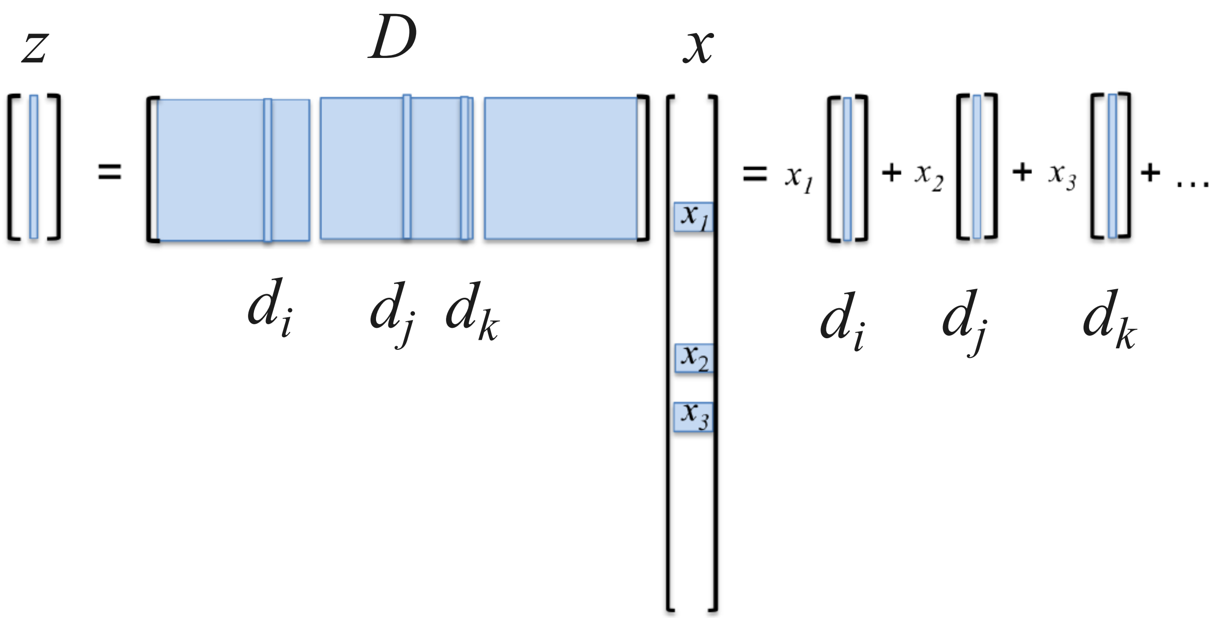
TV regularization

Dictionary Learning

- Dictionary representation of an image z in R^{M} :

$$\underset{x \in R^N}{Arg\,Min} \left\{ \|z - Dx\|_2^2 + \lambda \|x\|_0 \right\}$$

- $D=[d_1, d_2,\ldots, d_N] \in \mathrm{R}^{\mathrm{MxN}}$, is a set of basis vectors (could be one basis or a set of bases), or a set of known waveforms



Dictionary Learning

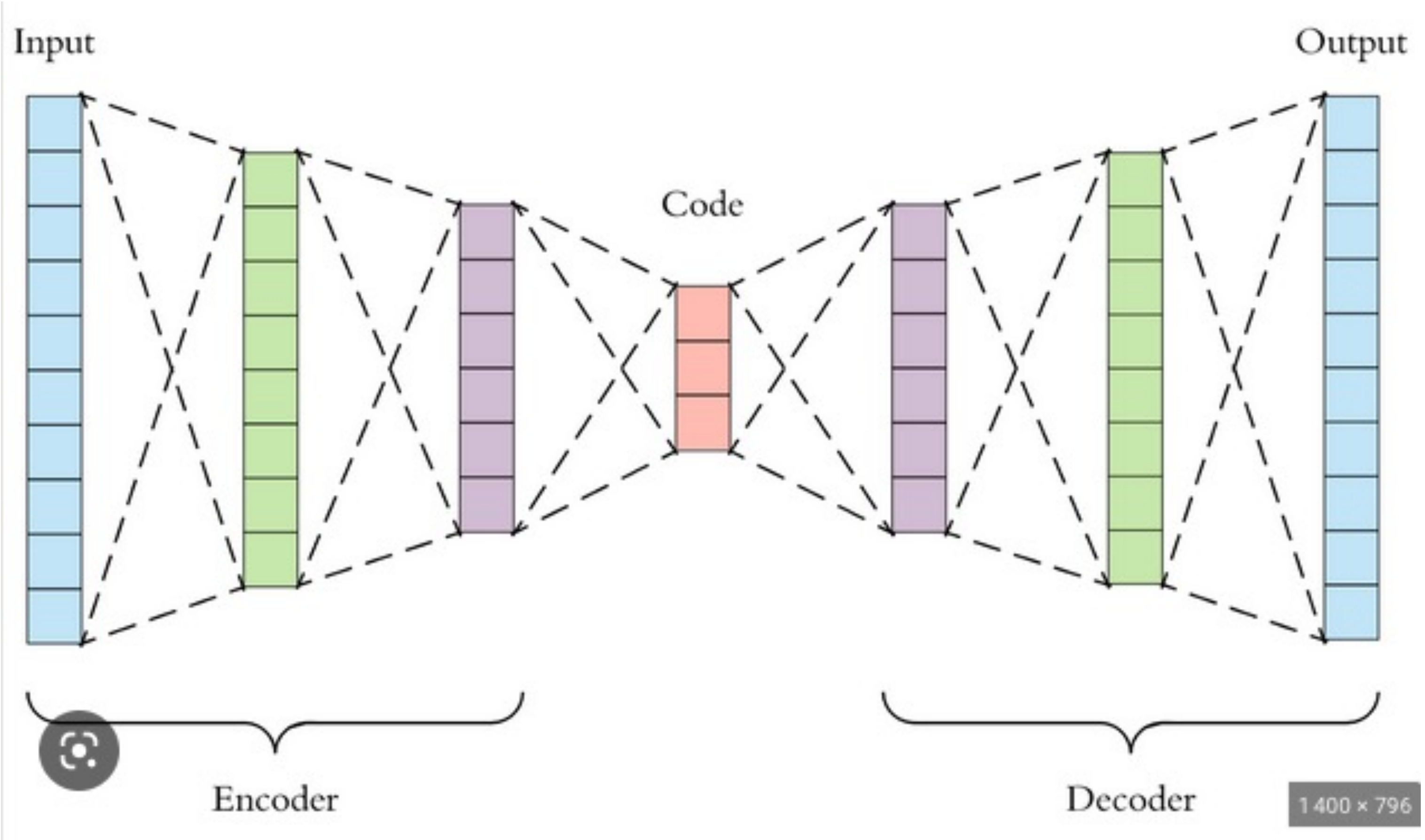
- D could be learned from a set of representative images, i.e. the training images $\left\{z^i\right\}_{i=1}^{P}$

- The goal is to design the dictionary D that leads to the best possible representations for each member in the training set with strict sparsity constraint

$$\forall i, \underset{D,x^i}{Min}\left\{\left\|z^i - Dx^i\right\|_2^2\right\} \quad \text{subject to} \quad \left\|x^i\right\|_0 \leq T$$

Pour tout $i=1,..P$

Same idea as an auto-encoder



©wikipedia

Dictionnary learning

- The ith image z^i is in a vector Mx1, we denote

$$Z = \left[z^1, z^2, \ldots, z^P \right] \in R^{MxP}$$

- We do the same for $X=[x^1, x^2, \ldots, x^P] \in R^{NxP}$, is a the set of sparse representation of the images z_i in Z.
- Dictionary learning from images z:

$$\underset{D,X}{Min} \left\{ \left\| Z - DX \right\|_F^2 \right\} \quad \text{subject to} \quad \forall i, \left\| x^i \right\|_0 \leq T$$

- $\left\| A \right\|_F^2 = \sum_{i,j} A_{i,j}^2$ is the Frobenius norm on matrix.

- Iterative minimization wrt D and X.

Dictionnary learning

$$\underset{D,X}{Min}\left\{\left\|Z-DX\right\|_F^2\right\} \quad \text{subject to} \quad \forall i, \left\|x_i\right\|_0 \le T$$

- To prevent D from being arbitrarily large, we impose D to be in the convex constraint set
$$C=\left\{D\in R^{M\times N} \,/\, \forall j=1,..,N,\; d_j^t d_j \le 1\right\}$$
- Fix D, estimate for all i the sparse approximation x^i, for example by using the (Orthogonal) Matching Pursuit algorithm.
- Fix X, estimate the dictionary D by cancelling the gradient of the criterion to minimize wrt D:
$$\hat{D}=Z.X^*\left(X.X^*\right)^{-1}$$

and renormalize the column of D.

- Remark: if XX^* is non invertible, use the pseudo-inverse (or Moore-Penrose inverse):
$$\hat{D}=Z.X^+ \quad \text{where} \quad X^+=\lim_{\delta\to 0^+} X^*\left(XX^*+\delta I\right)^{-1}$$

Dictionary learning

- Context of large scale datasets (e.g. millions of training samples in Z)
- Accessing the whole training set at each iteration is not possible
- Several approaches

  - K-SVD, [Aharon, Elad, Bruckstein, IEEE Trans on Signal processing, 54(11) 2006]
  - Stochastic gradient [Aharon, Elad, SIAM Journal on Imaging Sciences. 1 (3), 2008]
  - OnLine estimation [Mairal, Bach, Ponce, Sapiro, J. Mach. Learn. Res. 11, 2010]
  - …

K-SVD algorithm

- [Aharon, Elad, Bruckstein, IEEE Trans on Signal processing, 54(11) 2006]
- 1. Updating X: Fix dictionary D.

As $\left\| Z - DX \right\|_F^2 = \sum_{i=1}^{P} \left\| z^i - Dx^i \right\|_2^2$

the minimization wrt to X reduces to

$$\underset{x^i}{Min} \left\{ \left\| z^i - Dx^i \right\|_2^2 \right\} \quad \text{subject to} \quad \left\| x^i \right\|_0 \leq T, \quad \text{for } i = 1, \ldots P$$

- This can be done by algorithms as greedy algorithms, or approximations by basis pursuit algorithms (replacing l_0 by l_1 norm)

K-SVD algorithm

- II. Updating dictionary D: Fix X.

We have: $$\|Z - DX\|_F^2 = \left\| Z - \sum_{j=1}^{N} d_j . X_{j,.} \right\|_2^2 \quad d_j \in R^{Mx1}, \quad X_{j,.} \in R^{1xP}$$

d_j is the jth column of D and $X_{j,.}$ is the jth line of X.

- $d_j . X_{j,.} \in R^{MxP}$ is a rank one matrix
- The multiplication DX has been decomposed in a sum of N rank-1 matrices.
- The SVD algorithm iterates on column d_k

K-SVD algorithm

- We isolate only one such rank-1 matrix of index k:

$$\left\|Z - DX\right\|_F^2 = \left\|\left(Z - \sum_{j \neq k} d_j . X_{j,.}\right) - d_k . X_{k,.}\right\|_F^2 = \left\|E_k - d_k . X_{k,.}\right\|_F^2$$

Matrix E_k stands for the error for all the P examples when the kth atom d_k is removed. We assume they are all fixed, only d_k is optimized.

- By a SVD on $E_k \in R^{MxP}$, we could find the closest rank-1 matrix approximating E_k in the Frobenius norm sense. But we want $X_{k,.}$ to be sparse.
- Define $\omega_k = \left\{ i / 1 \leq i \leq P, \quad X_{k,i} \neq 0 \right\}$
- Restrict E_k by choosing only the columns corresponding to ω_k: E_k^{ω}
- Apply the restriction to the line $X_{k,.}$: $X_{k,.}^{\omega}$

K-SVD algorithm

- Apply SVD decomposition on $E_k^\omega = U \Delta V^t$
- Choose the update dictionary column d_k equal to the first column of U: $\hat{d}_k = U_{.,1} \; (= u_1)$
- Update the restricted coefficient vector $\widehat{X}_{k,.}^\omega = v_1 . \Delta(1,1)$

And re-expand it (with 0 coordinates) in $\mathrm{R}^{1\mathrm{xP}}$: $\hat{X}_{k,.} \in R^{1 \times P}$

- Iterate on column d_k

- Globally the algorithm iterates minimization wrt X and D.
  - For D the algorithm iterates on each column d_k (each waveform)