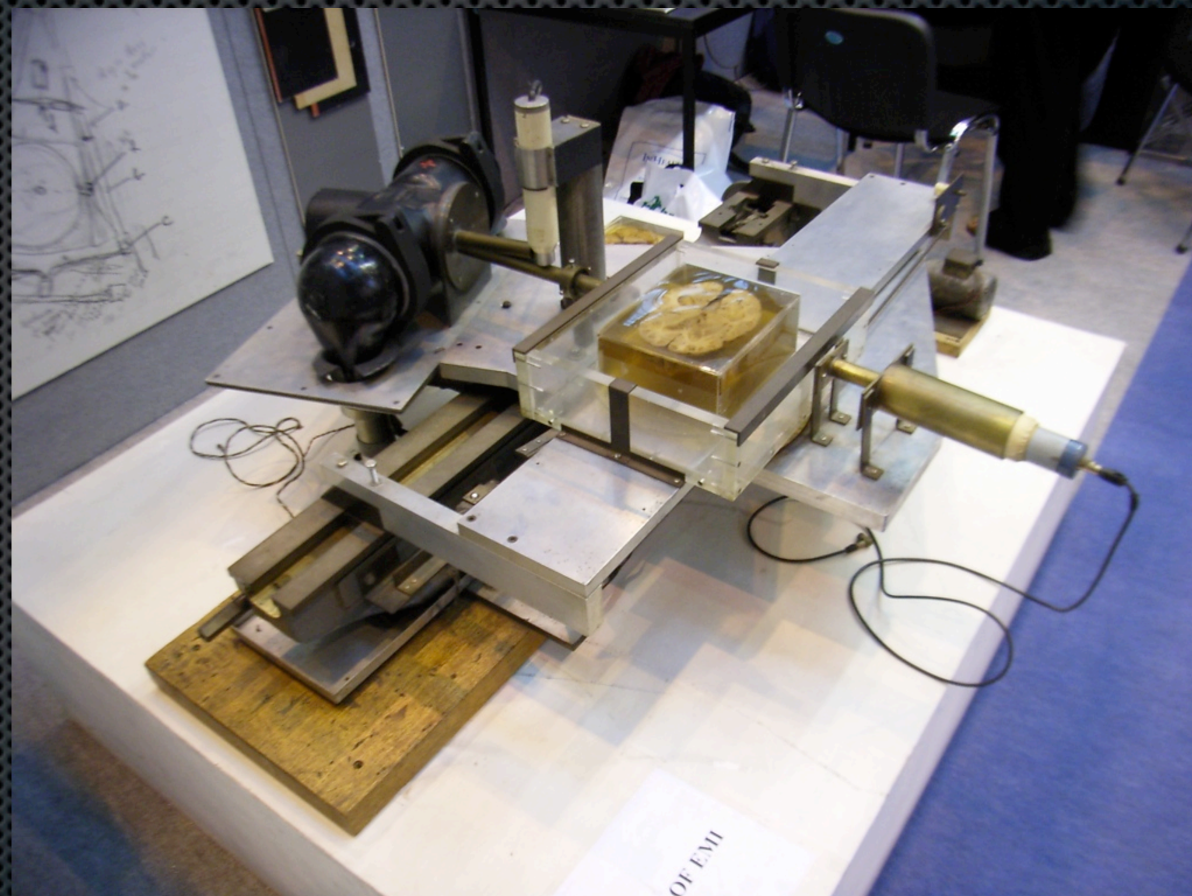
Introduction to Inverse Problems

Paolo Favaro

B39SI2 Digital Signal Processing II

How it all started...

- 1971: First clinical machine for detection of head tumors based on *X-ray Computed Tomography* (CT)
- 1979: G.L.Housfield and A.Cormack share the Nobel Prize for Physiology and Medicine



What is an inverse problem?

- **Mathematics**: J.B.Keller "We call two problems *inverse* of one another if the formulation of each involves all or part of the solution of the other. [...] one of the two problems has been studied extensively [...] while the other [...] is not so well understood. In such cases, the former is called the *direct problem*, while the latter is the *inverse problem."*
    - Data and unknowns can be exchanged arbitrarily.
- **Physics**: The direct problem is more fundamental and oriented along a cause-effect sequence based on well-established physical laws. E.g., compute the trajectories of particles from the knowledge of the forces. The inverse problem is to find the unknown causes of the known consequences.

What is an inverse problem?

- **Definition**: *A direct problem is a problem directed towards a loss of information*, i.e., its solution defines a transition from a physical quantity with a certain information content to another quantity with a smaller information content.

Example: Heat propagation

Direct problem: Compute the temperature distribution at a time t>0, given the temperature distribution at t=0.

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{D}\frac{\partial u}{\partial t}$$

Thermal conductivity *D* and boundary conditions

$$u(x,0) = f(x) \qquad u(0,t) = u(1,t) = 0$$

Example: Solution

Use Fourier series expansion

$$f(x) = \sum_{n=1}^{\infty} f_n \sin(\pi n x)$$

where

$$f_n = 2 \int_0^1 f(x) \sin(\pi n x) dx$$

then the solution is

$$u(x,t) = \sum_{n=1}^{\infty} f_n e^{-D(\pi n)^2 t} \sin(\pi n x)$$

Example: Loss of information

- Assume variables are known with precision ε (noise)
- then only coefficients f_n such that $|f_n| > \varepsilon$ are known
- Since f_n tend to zero as $n \mapsto \infty$, only N_ε coefficients are known
- this is the initial information content of the data of the direct problem
- Notice that at time *t*=*T* the number of Fourier coefficients of *u* is smaller than N_ε due to the decaying factor $e^{-D(\pi n)^2 T}$
- then there is a *loss of information* (entropy is increasing in time)

Example: The inverse problem

- Determine the temperature distribution at time $t=0$ ($f(x)$) given the temperature distribution at time $t=T$ ($u(x,T)$)

- Due to the loss of information exact recovery of $f(x)$ is not possible from $u(x,T)$: There exist many functions $f(x)$ that yield the same distribution at time $t=T$, $u(x,T)$

- Solving an inverse problem corresponds to a *gain of information*

What is an ill-posed problem?

- **Definition (Hadamard)**: A problem is *ill-posed* if any of the following is true

  (1) A solution does not always exist

  (2) The solution is not always unique

  (3) The solution does not depend continuously on the data

Example: Heat backpropagation

- Consider the following data at time *t=T*

$$u(x,T) = \frac{1}{n}\sin(\pi n x)$$

- then the solution of the inverse problem is

$$u(x,t) = \frac{1}{n}\sin(\pi n x)\, e^{D(\pi n)^2(T-t)}$$

- Notice that when $n \mapsto \infty$, the data function tends to *0* while the solutions tends to infinity
- This is due to the lack of continuity of the solution (ill-posedness)

Mathematical framework

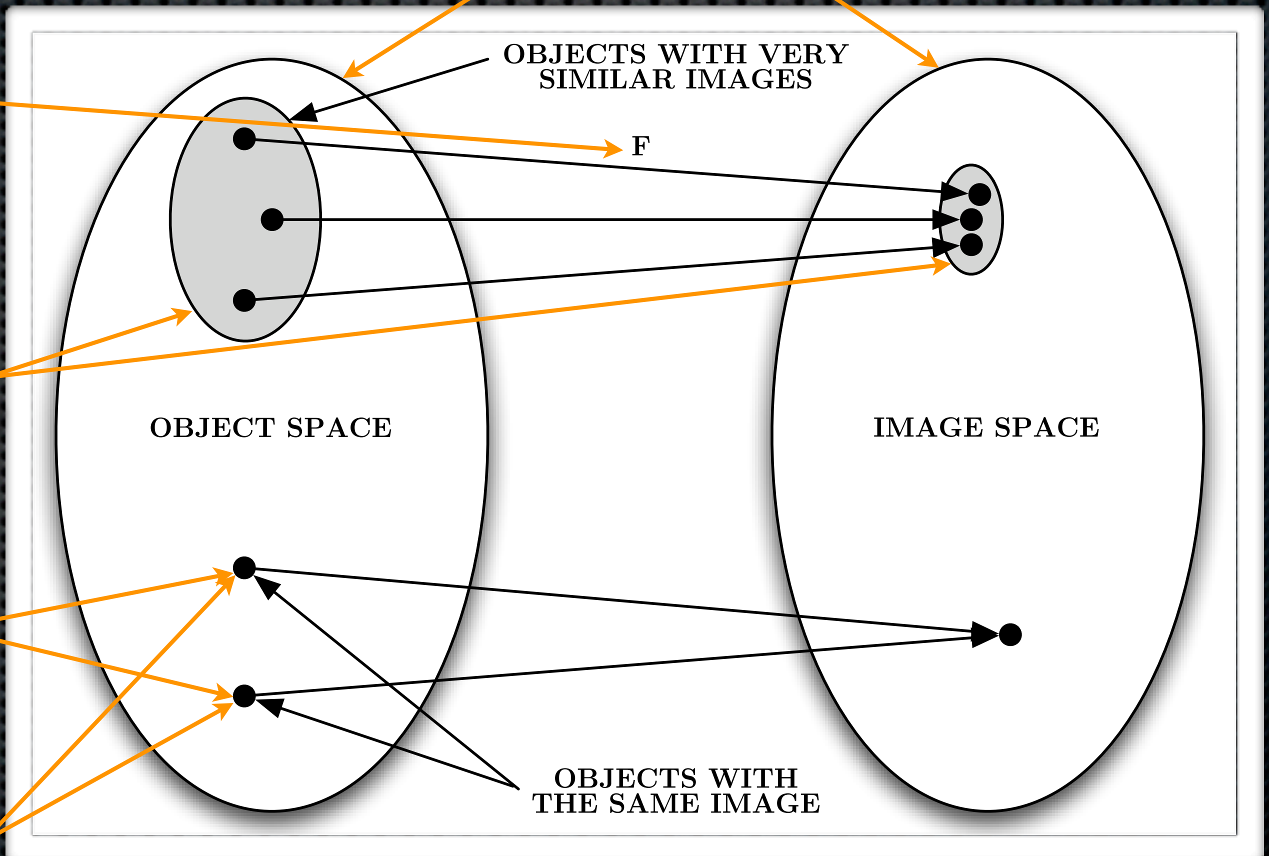
Equipped with a *metric* (to determine when objects/images are close or far)

The *range* of *F* is the set of *noise-free* images (may not coincide with the image space)

Continuity: images of two close objects are also close

Existence of *invisible objects*: Different objects map to the same image

Lack of continuity: Objects that are far apart map to images that are close



How to cure ill-posedness

- Remark: Continuity/discontinuity applies only to problems in infinite-dimensional spaces
- Discrete domain: Small oscillations in the data produce large oscillations in the solution (*ill-conditioning)*
- When multiple solutions exist, how do we pick the good one?
- **Golden rule**: Search for approximate solutions satisfying additional constraints coming from the physics of the problem (*a priori* information)
- Ex.: Finite energy, positiveness, upper bounds, family of functions, smoothness, statistical properties

A quick review of DSP

- Fourier transform (FT)

$$\hat{f}(\omega) \doteq \int e^{-jw\cdot x} f(x)dx$$

- Inverse Fourier Transform (IFT)

$$f(x) \doteq \frac{1}{(2\pi)^n} \int e^{j\omega\cdot x} \hat{f}(\omega)d\omega$$

- Generalized Parseval equality

$$\int f(x)h^*(x)dx = \frac{1}{(2\pi)^n} \int \hat{f}(\omega)\hat{h}^*(\omega)d\omega$$

Bandlimited functions and sampling theorems

- *Spacelimited*: Bounded support in space
- *Bandlimited*: Bounded support in frequency
- FT of a spacelimited function is never bandlimited and IFT of a bandlimited function is never spacelimited
- Bandlimited functions can be represented without loss of information by using samples $x_n = n\frac{\pi}{\Omega}$ (*Wittaker-Shannon Theorem*)

$$f(x) = \sum_{n=-\infty}^{+\infty} f\left(n\frac{\pi}{\Omega}\right) \operatorname{sinc}\left[\frac{\Omega}{\pi}\left(x - n\frac{\pi}{\Omega}\right)\right]$$

Convolution

- The convolution product of *f(x)* and *K(x)* is

$$g(x) = \int K(x-y) f(y) dy$$

$$g = K * f$$

- Convolution theorem

$$\hat{g}(\omega) = \hat{K}(\omega) \hat{f}(\omega)$$

- Inverse of the convolution

$$(K^{-1} g)(x) = \frac{1}{(2\pi)^n} \int \frac{\hat{g}(\omega)}{\hat{K}(\omega)} e^{j\omega \cdot x} d\omega$$

Inverse problems in imaging

$\delta(x-y)$

$K(x,y)$

point source

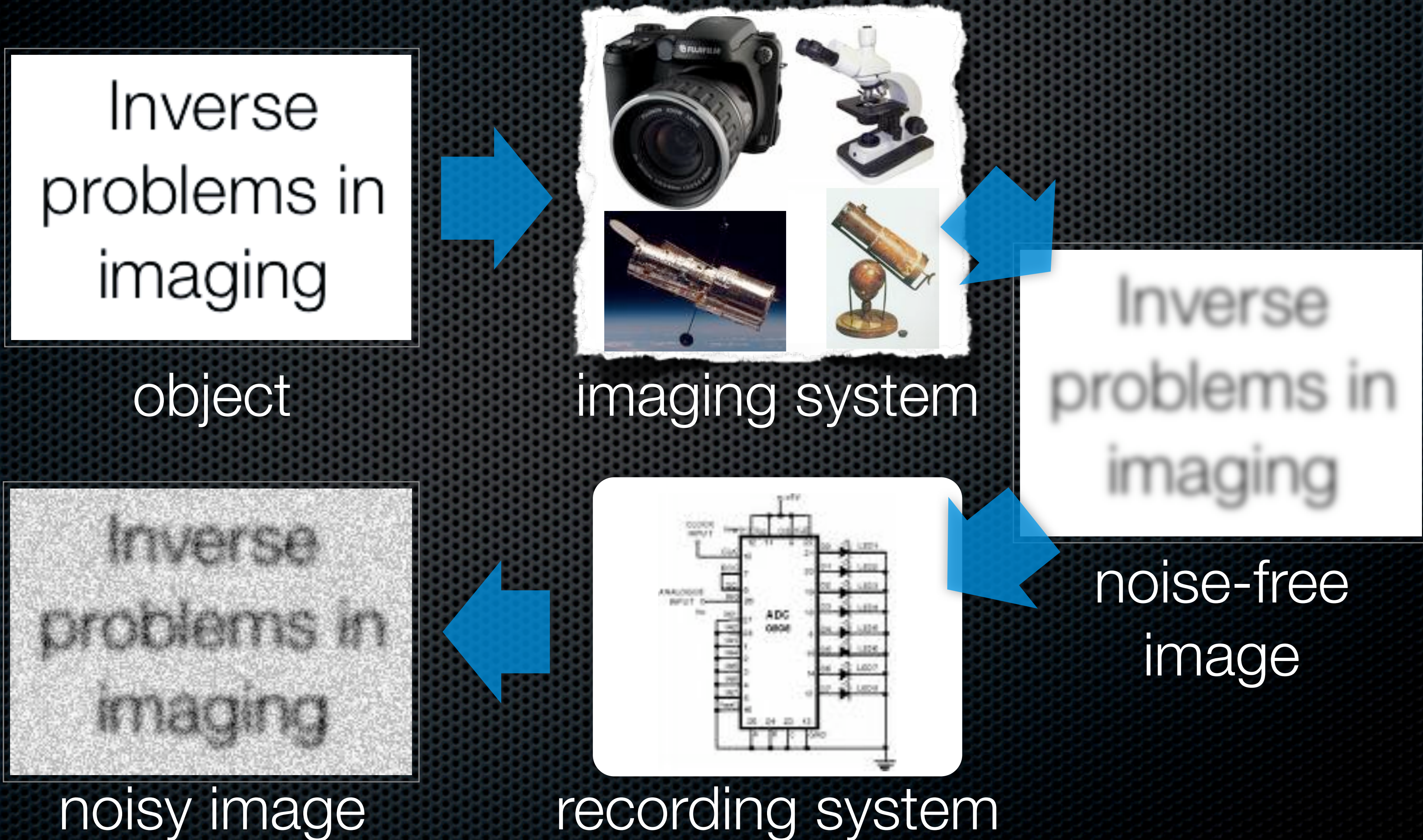
imaging system

image of the point source (Airy disc)

noise-free image $g_0(x) = \int K(x,y) f_0(y) dy$

point spread function (PSF)

Image formation



Convolutional model

- In many imaging systems the PSF is invariant to translations, i.e., *K(x,y) = K(x-y,0)*

$$g(x) = \int K(x-y) f_0(y) dy + w(x)$$
$$g = K * f_0 + w$$

- The *transfer function* $\hat{K}(\omega)$ is the FT of the PSF *K*

An example: The microscope

Every lens element alters the image in some way

The point spread function of the system is a convolution of all the point spread functions of all the lens elements and apertures in the optical train

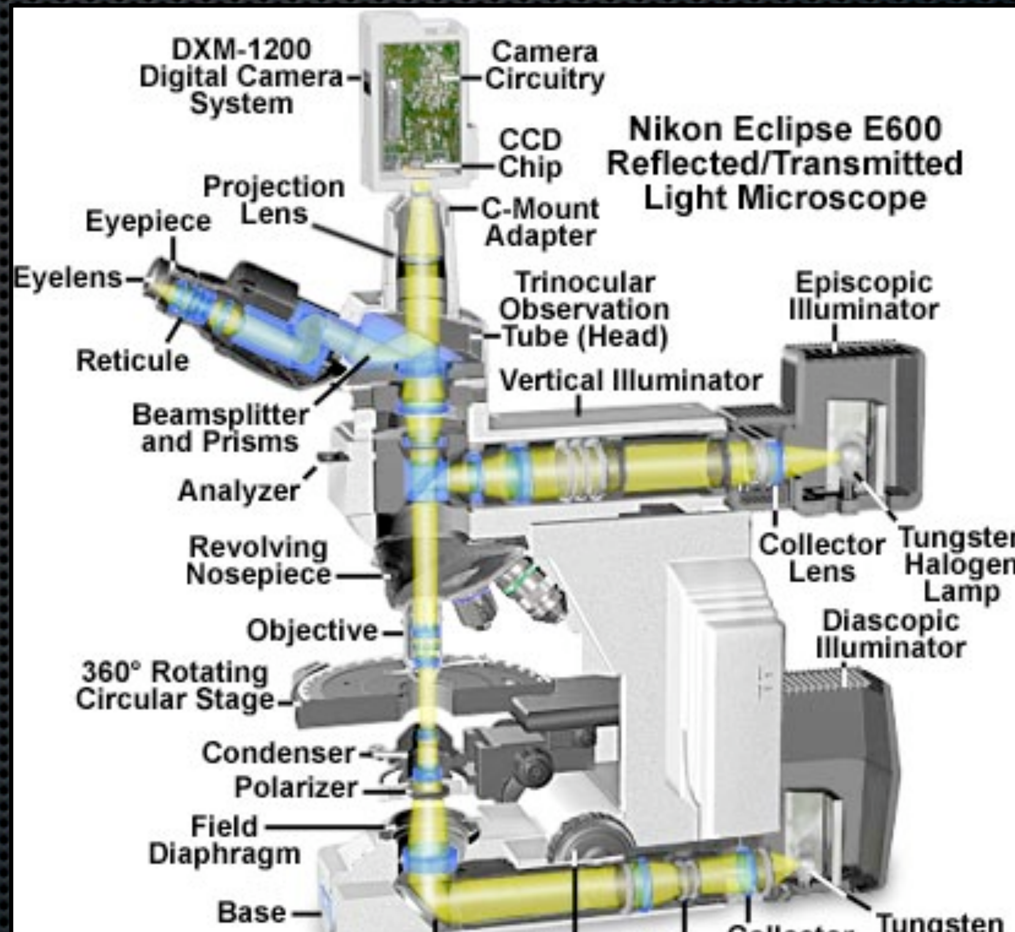


Figure 1

The ill-posedness of image deconvolution

- Recall that images can be written as

$$g(x) = \int K(x-y) f_0(y) dy + w(x)$$

- In general one may not know the PSF *K* or know only the parametric family of the PSFs (image identification, blur identification, image-blur identification, blind deconvolution)
- We suppose that the PSF is known sufficiently well

Discretization

- Consider the discretization of
$$g(x) = \int K(x-y) f_0(y) dy + w(x)$$
as
$$g = Af_0 + w$$
where A represents a matrix operator
- If there is no noise, this amounts to solving a linear algebraic system

Frequency domain

- Both in the continuous and discrete cases we can apply the FT or the DFT and obtain a linear set of equations in the form of

$$\hat{K}(\omega)\hat{f}(\omega) = \hat{g}(\omega)$$

Well-posedness

- The solution of the problem is unique
- The solution exists for any data
- The solution depends continuously on the data*

*Courant: “Data in nature cannot be considered as rigidly fixed; the mere process of measuring them involves small errors. Therefore a mathematical problem cannot be considered as realistically corresponding to physical phenomena unless a variation of the given data in a sufficiently small range leads to an arbitrary small change in the solution. This requirement for *stability* is not only essential for meaningful problems in mathematical physics, but also for approximation methods”.

Uniqueness

- If the solution is not unique, then there exist at least two distinct objects f_1 and f_2 such that $Af_1 = g$ and $Af_2 = g$
- Since *A* is linear then $A(f_1 - f_2) = 0$ and $f = f_1 - f_2 \neq 0$
- Conversely, if there exist an $f \neq 0$ that satisfies $Af = 0$ and f_1 is such that $Af_1 = g$ then $f_2 = f_1 + f$ satisfies $Af_2 = g$
- Uniqueness of $Af = g$ is guaranteed if and only if $Af = 0$ is satisfied only for $f = 0$

Uniqueness

- In frequency, the analysis of uniqueness corresponds to studying
$$\hat{K}(\omega)\hat{f}(\omega) = 0$$
- If the support of $\hat{K}(\omega)$ covers the entire frequency space, then the only possible solution is $\hat{f}(\omega) = 0$
- Bandlimited systems do not satisfy the uniqueness constraint

Existence

- Suppose that the solution is unique, let us now consider the *existence*
- If the system is not bandlimited, then the solution is given by

$$\hat{f}(\omega) = \frac{\hat{g}(\omega)}{\hat{K}(\omega)}$$

  whenever the ratio above defines the FT of a function
- In some isolated points $\hat{K}(\omega)$ might be 0 and the ratio will not be defined (the solution does not exist)
- Even if $\hat{K}(\omega)$ is always different from 0, $\hat{K}(\omega)$ and $\hat{g}(\omega)$ tend to 0 as ω goes to infinity, but their ratio may not

Existence

- The condition for existence can be written as

$$\int \left| \frac{\hat{g}(\omega)}{\hat{K}(\omega)} \right|^2 d\omega < \infty$$

- Notice that uniqueness does not imply existence
- However, in the discrete case when uniqueness holds, the problem is well-posed (*inverse filtering*)

Discrete case: Ill-conditioning

- We have seen that by discretizing an ill-posed problem (with uniqueness) we obtained a well-posed problem

- Unfortunately, such solution is typically completely corrupted by noise...

- Why?

Ill-conditioning

- Consider a small variation of the discrete image *g*, then the corresponding small variation of the object is

$$\delta f^\dagger = A^{-1}\delta g$$

- Then, we can obtain the following bound

$$||\delta f^\dagger|| \leq \frac{1}{\hat{K}_{\mathrm{min}}}||\delta g||$$

- and we also have that

$$||g|| \leq \hat{K}_{\mathrm{max}}||f^\dagger||$$

- so that we obtain

conditioning number

$$\frac{||\delta f^\dagger||}{||f^\dagger||} \leq \frac{\hat{K}_{\mathrm{max}}}{\hat{K}_{\mathrm{min}}}\frac{||\delta g||}{||g||}$$

Ill-conditioning

- Notice that the more we increase the accuracy of the discretized problem and the more the conditioning number grows
- Indeed, as ω goes to infinity, $\hat{K}(\omega)$ goes to *0*
- Hence, we will obtain smaller and smaller $\hat{K}_{\min}$

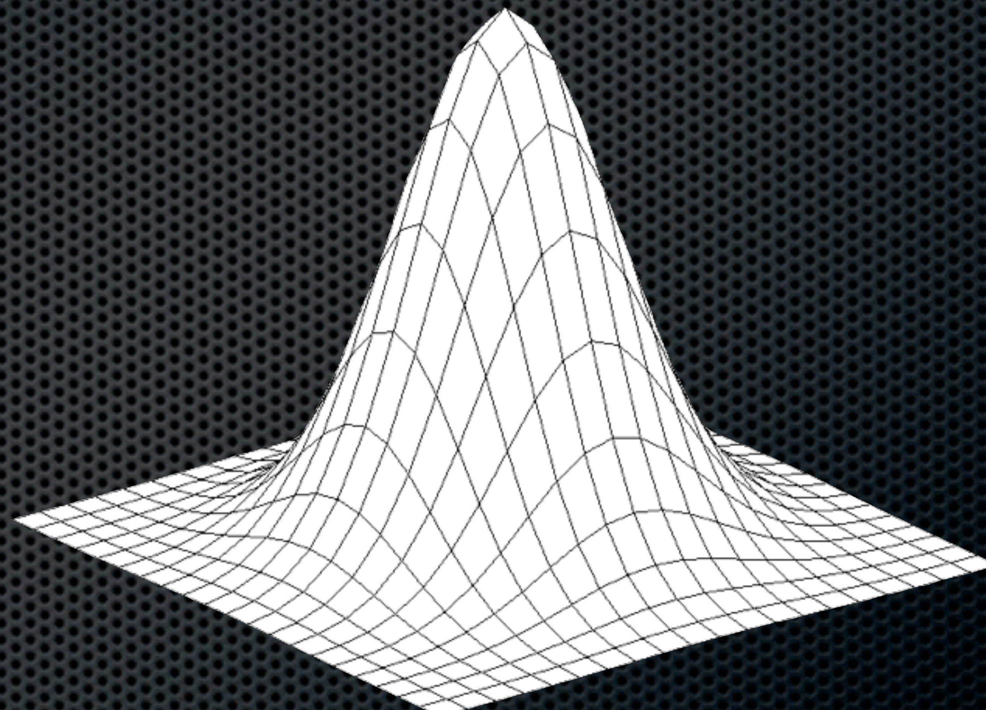
Example: Inverse filtering

object

image

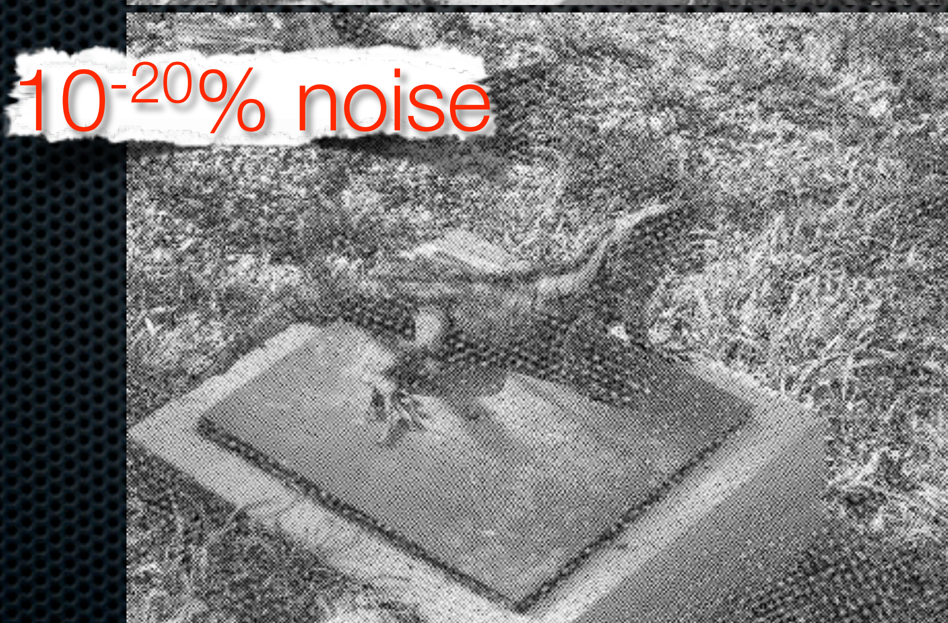
Example: Gaussian PSF

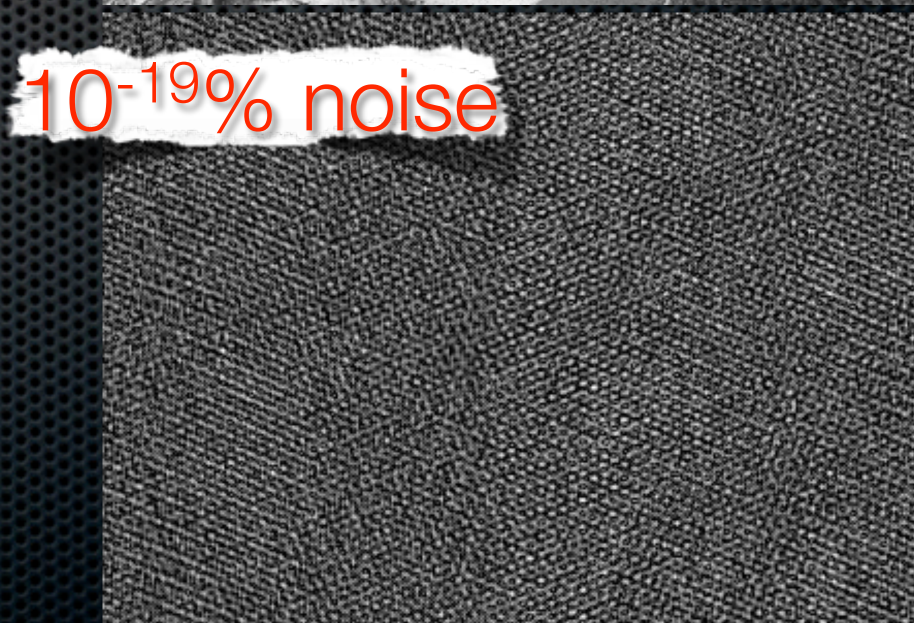
PSF intensities (grayscale)



PSF mesh plot

Example: Reconstructions





Out-of-band noise

- Recall that band-limited systems do not have a unique solution

$$g(x) = \int K(x-y) f_0(y) dy + w(x)$$

$$\hat{g}(\omega) = \hat{K}(\omega)\hat{f}(\omega) + w(\omega)$$

- If $\hat{K}(\omega)$ is nonzero within a band *B* then for any frequency outside B we have the equation

$$0 = w(\omega)$$

- The noise at these frequencies is called the out-of-band noise

Out-of-band noise

- If noise is generated only as out-of-band noise, then we can easily remove it
- We need to introduce a *projection* operator *P*

$$(P_B g)(x) \doteq \frac{1}{(2\pi)^n} \int_B \hat{g}(\omega) e^{j\omega \cdot x} d\omega$$

- and instead of $Af = g$ use the following image model

$$Af = P_B g$$

Example: Image filtering

bandlimited object

image + out-of-band noise

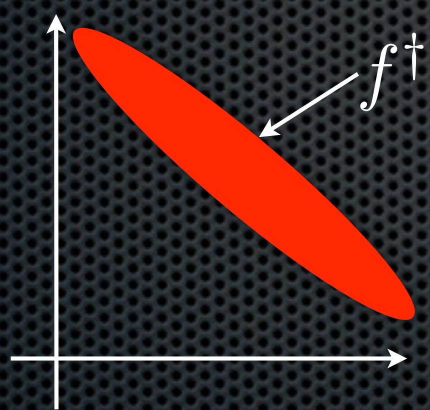
Example: Image filtering

bandlimited object

recovered object

The use of *a-priori* information

- Even when we can remedy existence, uniqueness may remain an issue
- Typically, the set of admissible solutions is too large



object space

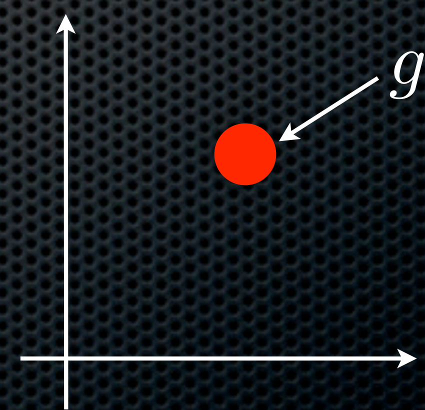
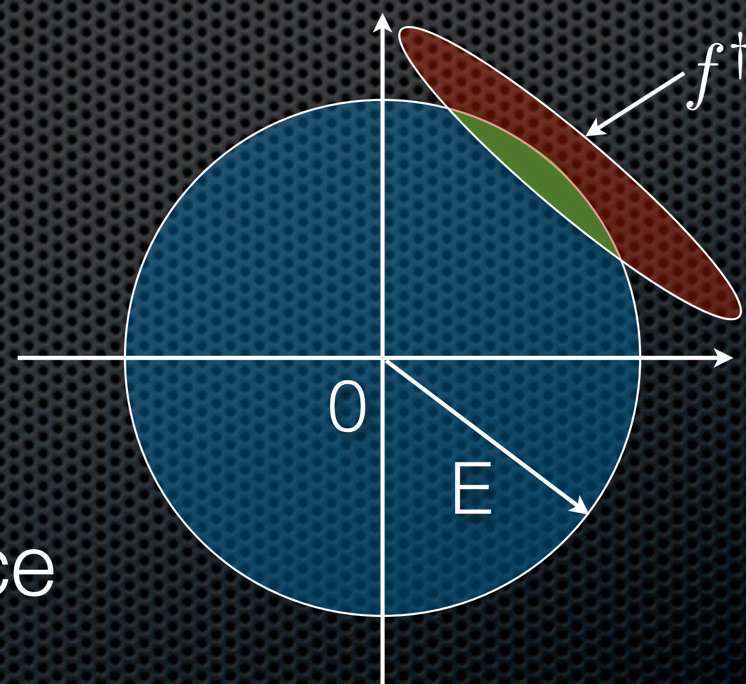


image space

The use of *a-priori* information

- The set of admissible solutions can be reduced by introducing additional constraints
- E.g.: bound on the norm (<E)



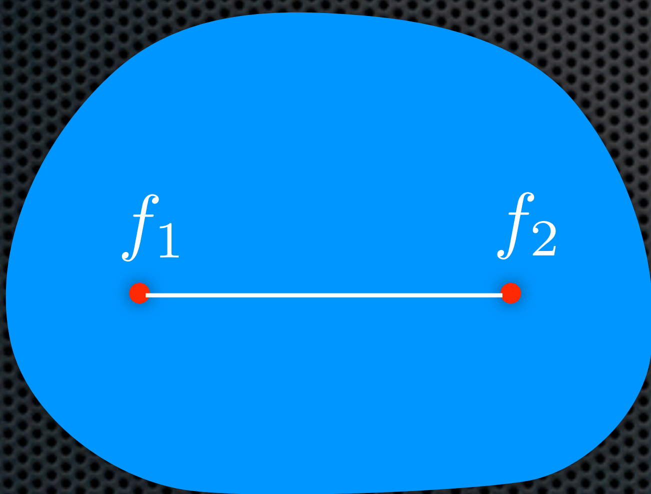
object space

The use of *a-priori* information

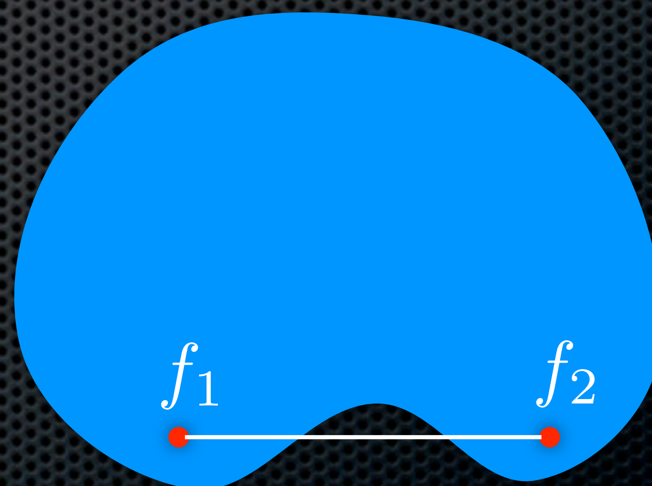
- Bound on the energy
- Bound on the magnitude of the derivatives
- Spacelimited functions
- Bandlimited functions
- Nonnegative functions
- Functions fixed on point sets
- Any intersection of the above

The use of *a-priori* information

- Previous constraints form closed and convex sets



convex



non convex

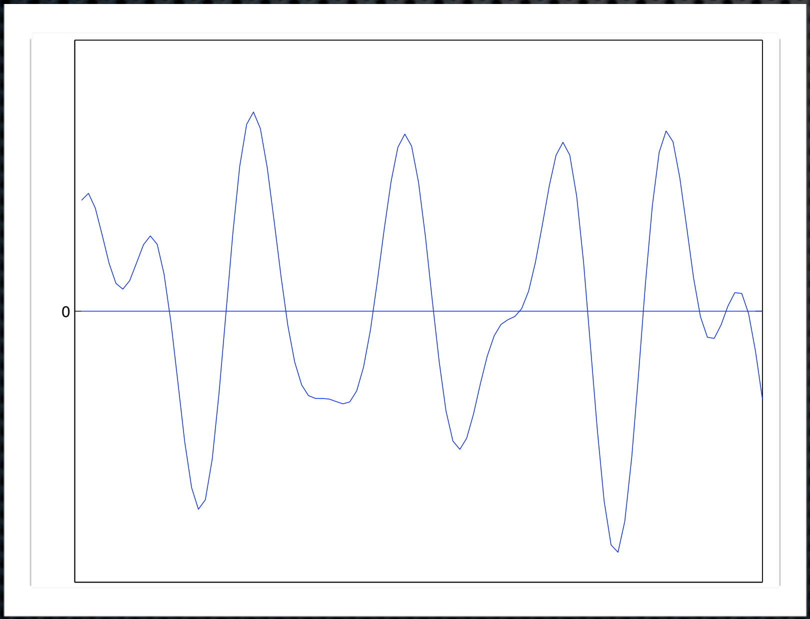
The use of *a-priori* information

- **Property (convex sets)**: If C is a convex set, then given an arbitrary element f of the object space, there exists a unique element f_C of C which has minimal distance from f.

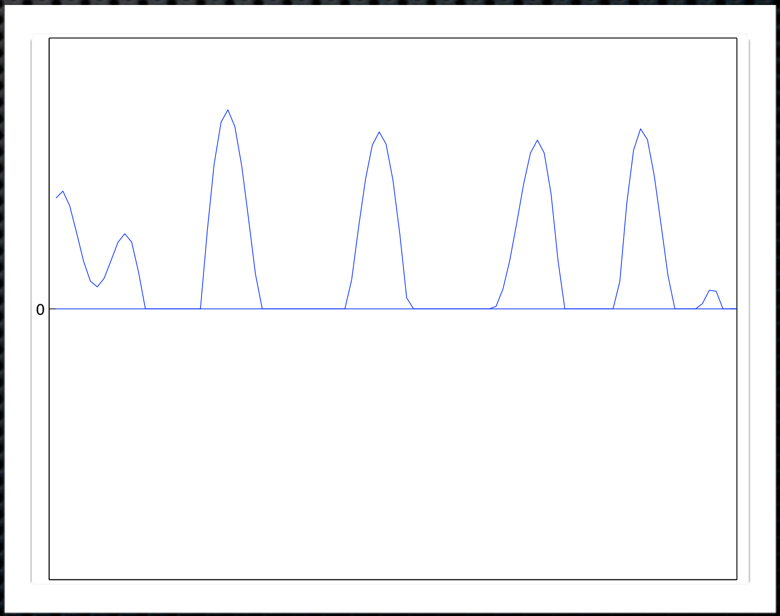
The use of *a-priori* information

- Example: Projection onto nonnegative functions

$$(P_C f)(x) = \begin{cases} f(x) & \text{if } f(x) > 0 \\ 0 & \text{if } f(x) \leq 0 \end{cases}$$



original function



projected function

Regularization methods

- **Definition (A.N.Tikhonov)**: Consider a family of approximate solutions depending on a positive parameter called the *regularization parameter.* Properties: When there is no noise, the family converges to the correct solution as the parameter goes to *0*; when there is noise, the approximate solution can be obtained for a non zero parameter.

Least-squares with prescribed energy

- Find the minima of

$$\epsilon^2(f;g) = \|Af - g\|^2$$

  subject to

$$E^2(f) = \|f\|^2 = E^2$$

- Use *Lagrange multipliers*

$$\Phi_\mu(f;g) = \epsilon^2(f;g) + \mu E^2(f) = \|Af - g\|^2 + \mu\|f\|^2$$

Least-squares with prescribed energy

- When the PSF is bandlimited, the solution is

$$\hat{f}_\mu(\omega) = \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu} \hat{g}(\omega)$$

- and its IFT satisfies the following properties:

1. It is square integrable for any *g* and any positive μ ;
2. It depends continuously on *g*;
3. It is orthogonal to the subspace of all invisible objects.

$$E^2(f_\mu) = \|f_\mu\|^2 \leq \frac{1}{\mu}\|g\|^2$$

Least-squares with minimal energy

- Find the minima of

$$E^2(f) = \|f\|^2$$

  subject to

$$\epsilon^2(f;g) = \|Af - g\|^2 \leq \epsilon^2$$

- If the null solution does not belong to the prescribed set, then equality holds
- The previous solution (via Lagrange multipliers) applies

Regularization (Tikhonov)

- Recall

$$\Phi_\mu(f;g) = \epsilon^2(f;g) + \mu E^2(f) = \|Af - g\|^2 + \mu\|f\|^2$$

- Then we have that the Euler-Lagrange equations are

$$A^*(Af_\mu - g) + \mu f_\mu = 0$$

- which can be rewritten as

$$(A^*A - \mu I_d)f_\mu = A^*g$$

- and the *regularized solutions* are

$$f_\mu = (A^*A + \mu I_d)^{-1}A^*g = R_\mu g$$

Regularization (Tikhonov)

- The operator R_μ must satisfy the following two properties

 1. For $\mu > 0$ it is a linear and continuous operator

 2. For any noise-free image g_0

$$\lim_{\mu \mapsto 0} R_\mu g_0 = P_B f_0$$

Error analysis

- Suppose we apply the regularization operator to noisy data

$$R_\mu g = R_\mu A f_0 + R_\mu w$$

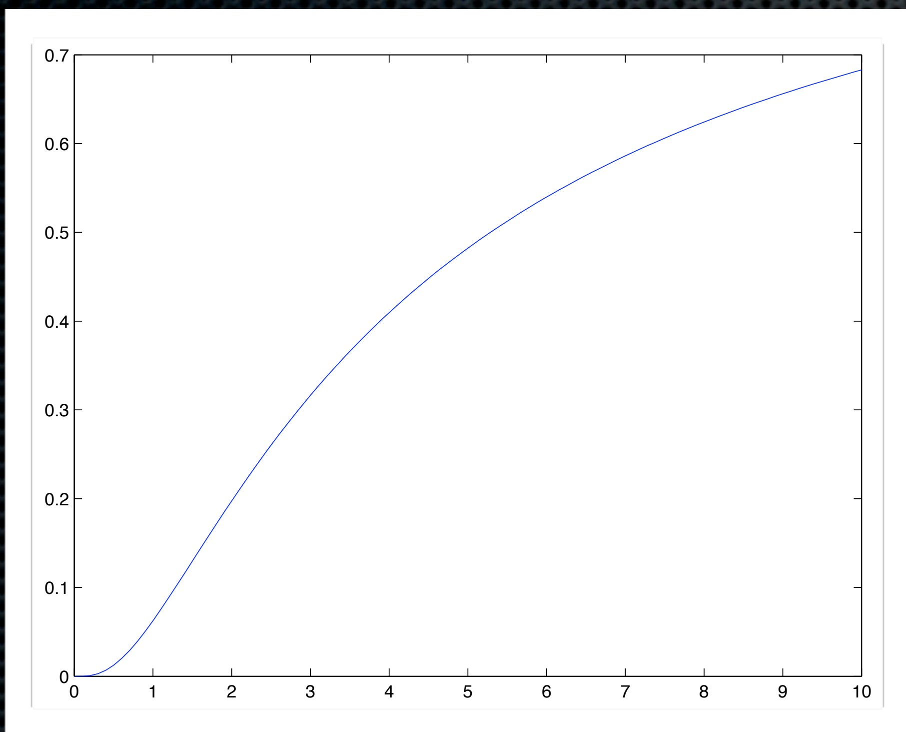
- Error can be decomposed into: Approximation error and noise-propagation error

$$R_\mu g - P_B f_0 = (R_\mu A f_0 - P_B f_0) + R_\mu w$$

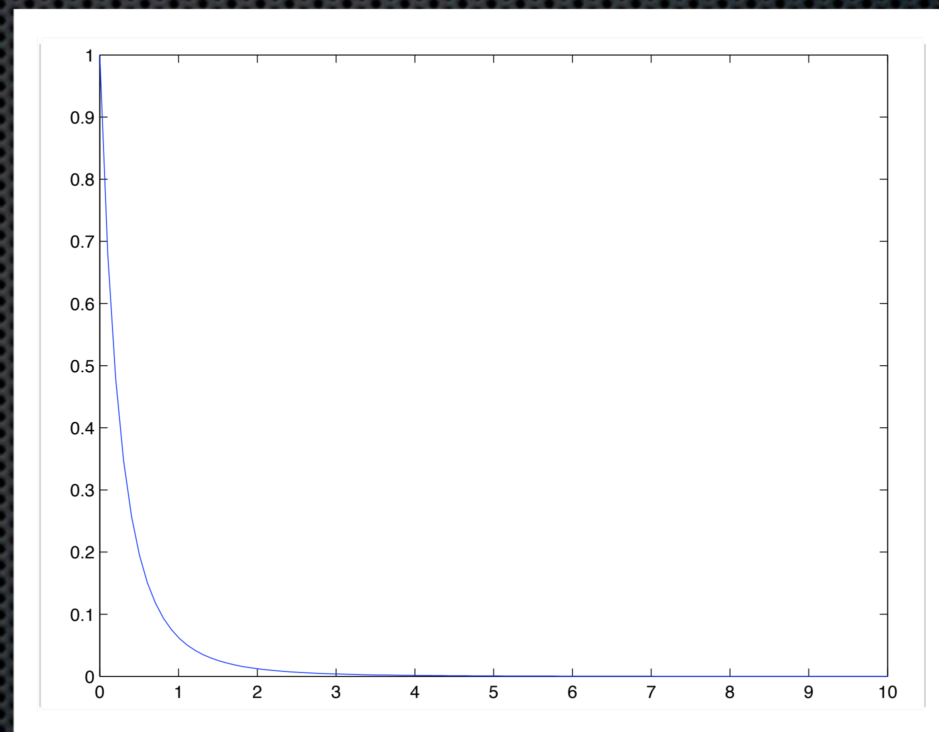
$$\|R_\mu g - P_B f_0\| \leq \|R_\mu A f_0 - P_B f_0\| + \|R_\mu w\|$$

approximation error — noise-propagation error

Error analysis

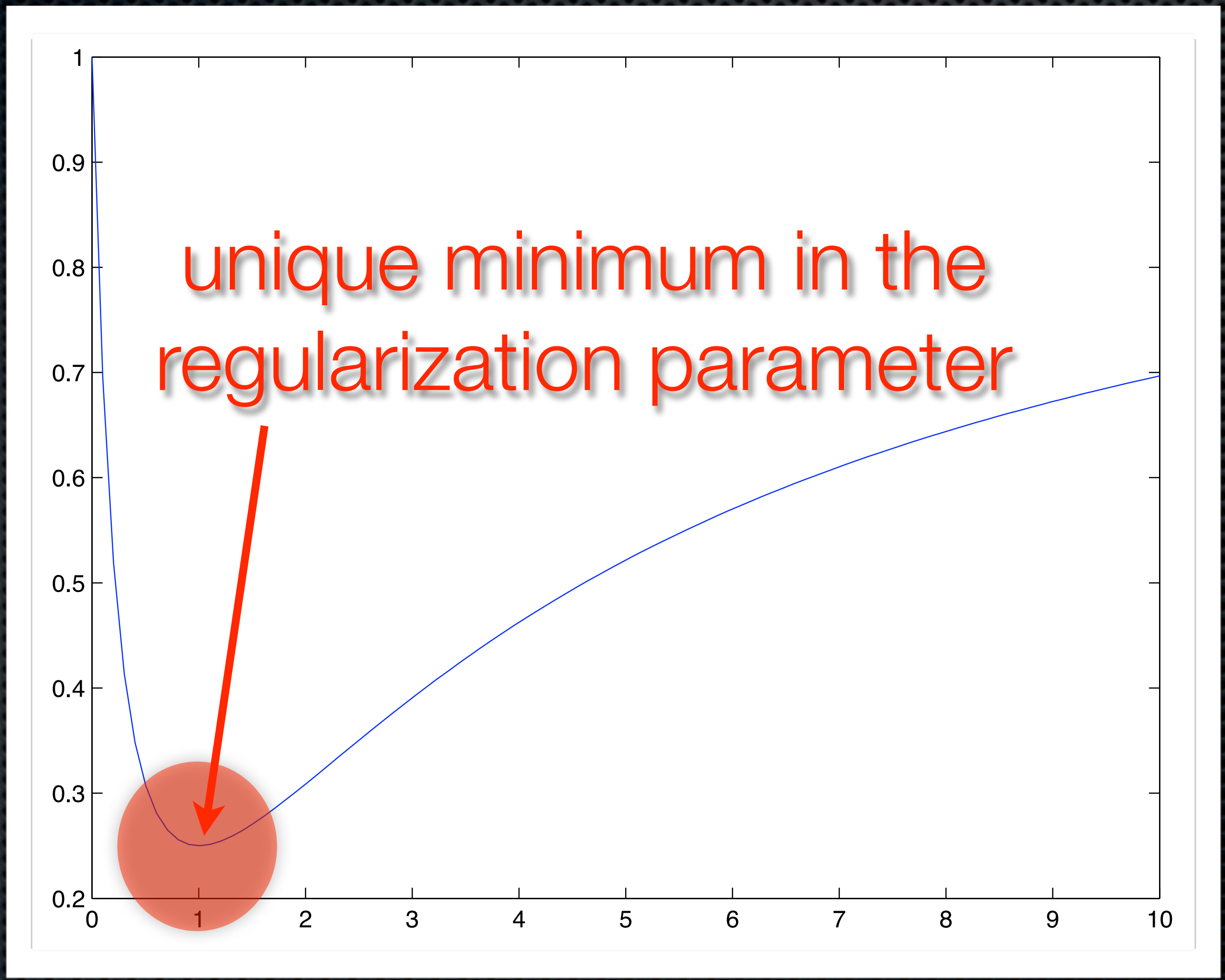


approximation error



noise-propagation error

Error analysis



combined error

Examples

blurred image

blurred and noisy image

Examples

$\mu = 1$ $\mu = 10^{-1}$ $\mu = 10^{-2}$ $\mu = 10^{-3}$

$\mu = 10$ $\mu = 100$ $\mu = 10^{3}$ $\mu = 10^{4}$

Tikhonov Regularization (General Form)

- General regularization term as a linear combination of the energy of *0*-th to *n*-th derivatives of the solution
- Yields a polynomial *P* of quadratic terms (w^2) in the frequency space
- Extensions from 2D to 3D: consider all combinations of partial derivatives

$$f_\mu = R_\mu g$$

$$R_\mu g = K_\mu^\dagger * g \qquad K_\mu^\dagger(x) = \frac{1}{(2\pi)^n} \int \frac{\hat{K}^*(\omega)}{|\hat{K}(\omega)|^2 + \mu \hat{P}(\omega)} e^{j\omega\cdot x} d\omega$$

Resolution vs Ringing

- For linear methods the band-limiting assumption implies that the solution is also band-limited
- If there are edges, the band-limiting assumption generates ringing
- To reduce ringing we can employ *windowing* and loose resolution
- Hence, there is a trade-off between resolution and ringing

Choice of the regularization parameter

- No general recipe (so far)
- In practice, use an estimate of the numerical accuracy of the sensor + an estimate of the noise level
- If noise energy is known accurately, one can use the *discrepancy principle*
- If bounds on noise energy and solution energy are known, then one can use the *Miller* method
- If no bounds are known, then one can use the *generalized cross-validation*

Iterative regularization methods

- Iterative methods have been introduced to solve linear algebraic systems
- Hence, such methods can be used to find a regularized solution
- The number of iterations will play the role of the regularization parameter: As the number of iterations increases, the solution will first approach the unknown object and then move away from it (*semiconvergence*)

The van Cittert and Landweber methods

- Start from the least-squares problem
$$||Af - g||^2$$
- The least-squares solution satisfies (Landweber)
$$A^*Af = A^*g$$
- Alternatively one can aim at solving (van Cittert)
$$Af = g$$
- To treat both methods simultaneously, we use
$$\bar{A}f = \bar{g}$$

Successive approximations

- Define the operator

$$T(f) = f + \tau(\bar{g} - \bar{A}f)$$

where τ is the *relaxation parameter*. The solution of the original problem is a *fixed point* of this operator.

- Method of *successive approximations*

$$f_{k+1} = T(f_k)$$

Convergence properties

- The method of successive approximations does not converge when there is out-of-band noise
- van Cittert: $Real\{\hat{K}(\omega)\} > 0$ $\quad 0 < \tau < \frac{2}{\hat{K}_{max}}$
- Landweber: $\quad 0 < \tau < \frac{2}{\hat{K}^2_{max}}$

Landweber vs Tikhonov

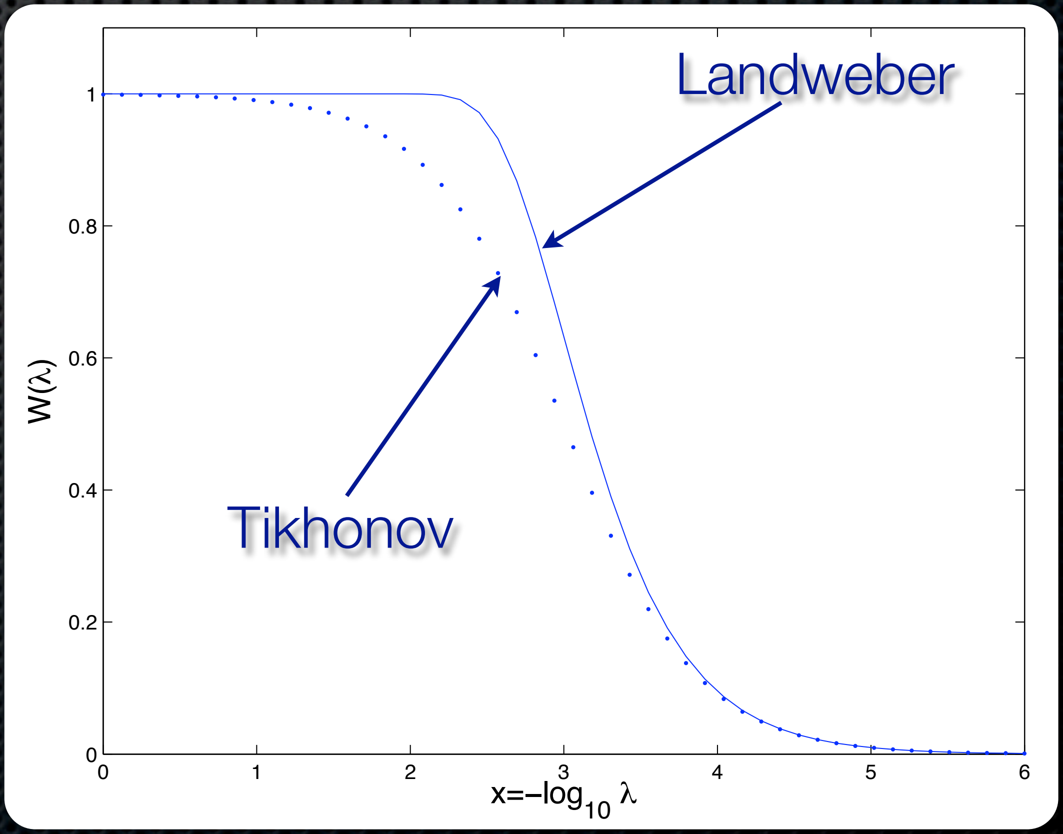
- They have similar convergence properties
- Both can be shown to be equivalent to windowing of the inverse convolution

$$\hat{W}_\mu(\lambda) = \frac{\lambda}{\lambda+\mu}$$

Tikhonov regularization

$$\hat{W}_k(\lambda) = 1 - (1-\lambda)^k$$

Landweber regularization

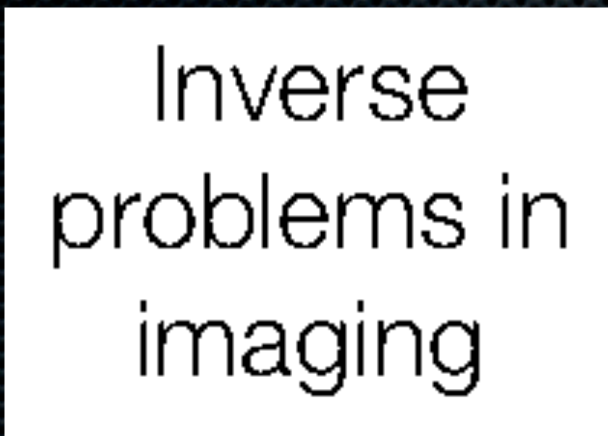


Projected Landweber

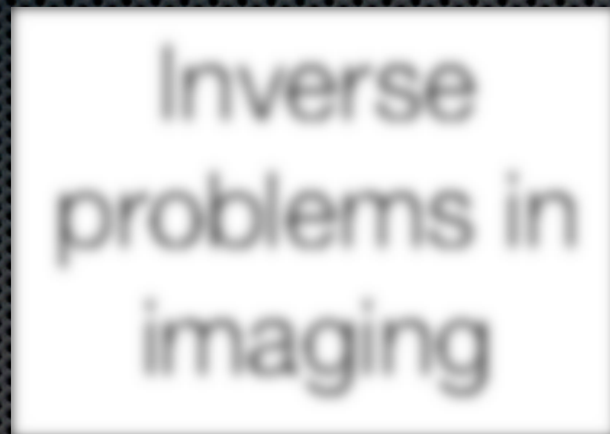
- The *Landweber* method can be used to solve constrained least-squares problems
- The *projected Landweber* method yields approximate solutions that converge to the true object without noise, and has *semiconvergence* with noise

$$f_{k+1} = P_C[f_k + \tau(\bar{g} - \bar{A} f_k)]$$

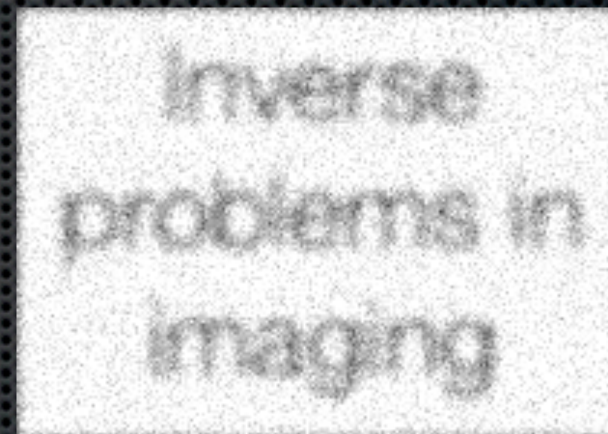
Example



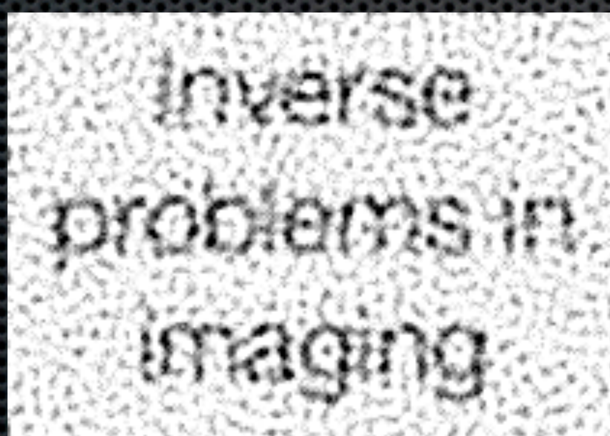
original



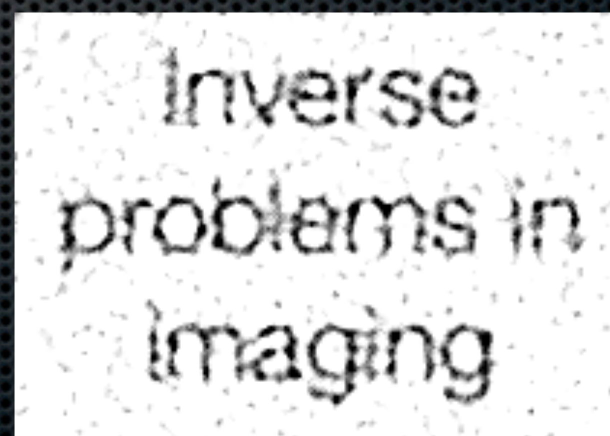
blurred



blurred+noisy



Landweber



projected Landweber

Steepest descent

- The Landweber method is an example of *gradient methods*, i.e., methods where at each step the approximation is modified in the direction of the gradient of the discrepancy functional

energy $$||Af - g||^2$$

explicit form $$(Af - g)^*(Af - g)$$

energy gradient $$\nabla_f E = 2(A^*Af - A^*g)$$

Steepest descent

- Update the approximation in the direction opposite to the gradient

$$f_{k+1} = f_k - \tau \nabla_f E$$

- The optimal step is the one that minimizes the energy E given the direction (the gradient)

$$E_{k+1} = E_k + \tau^2 \|A \nabla_f E\|^2 - 2\tau \|\nabla_f E\|^2$$

- which gives

$$\tau = \frac{\|\nabla_f E\|^2}{\|A \nabla_f E\|^2}$$

Steepest descent

- Define the residuals as

$$r_k \doteq \nabla_f E_k$$

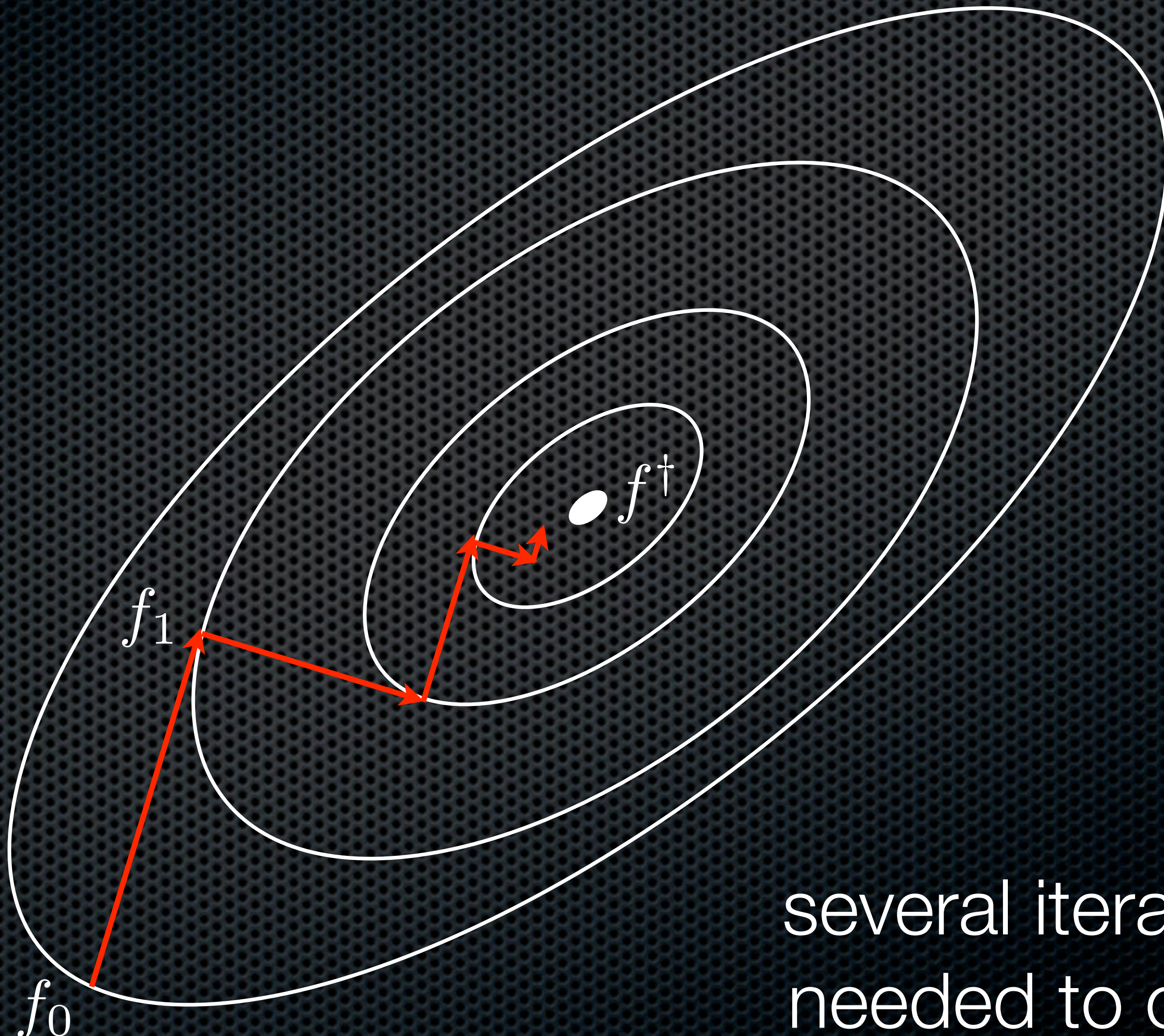
- Residuals iterate according to

$$r_{k+1} = r_k - \tau A^* A r_k$$

- and they are orthogonal by construction

$$r_{k+1} \perp r_k$$

Steepest descent:Illustration



several iterations are
needed to converge

Conjugate gradient descent

- Definition: Solutions *f* and *h* are *conjugate* with respect to A^*A if
$$f^*A^*Ah = 0$$
- Steepest descent is a linear combination of functions
$$A^*g, A^*AA^*g, (A^*A)^2A^*g, \ldots, (A^*A)^kA^*g$$
which span the so-called *Krylov subspace*

Conjugate gradient descent

- The CGD does not project the generalized solution on the *Krylov* subspace, but instead computes a solution in the *Krylov* subspace that minimizes the energy *E*
- Denote with P_k the *projection* onto the k-th *Krylov* subspace
- The CGD is a *projection method*: At each step *k* it computes the projection of the least-squares solution onto the *Krylov* subspace

$$||AP_k f - g||^2 \qquad P_k f = f$$

then the solution satisfies

$$P_k A^* A P_k f = P_k A^* g$$

Conjugate gradient descent

- Algorithm

1. initialize $r_0 = p_0 = A^* g \qquad f_0 = 0$
2. compute

$$\alpha_k = \frac{\|r_k\|^2}{r_k^* A^* A p_k}$$

$$f_{k+1} = f_k + \alpha_k p_k$$

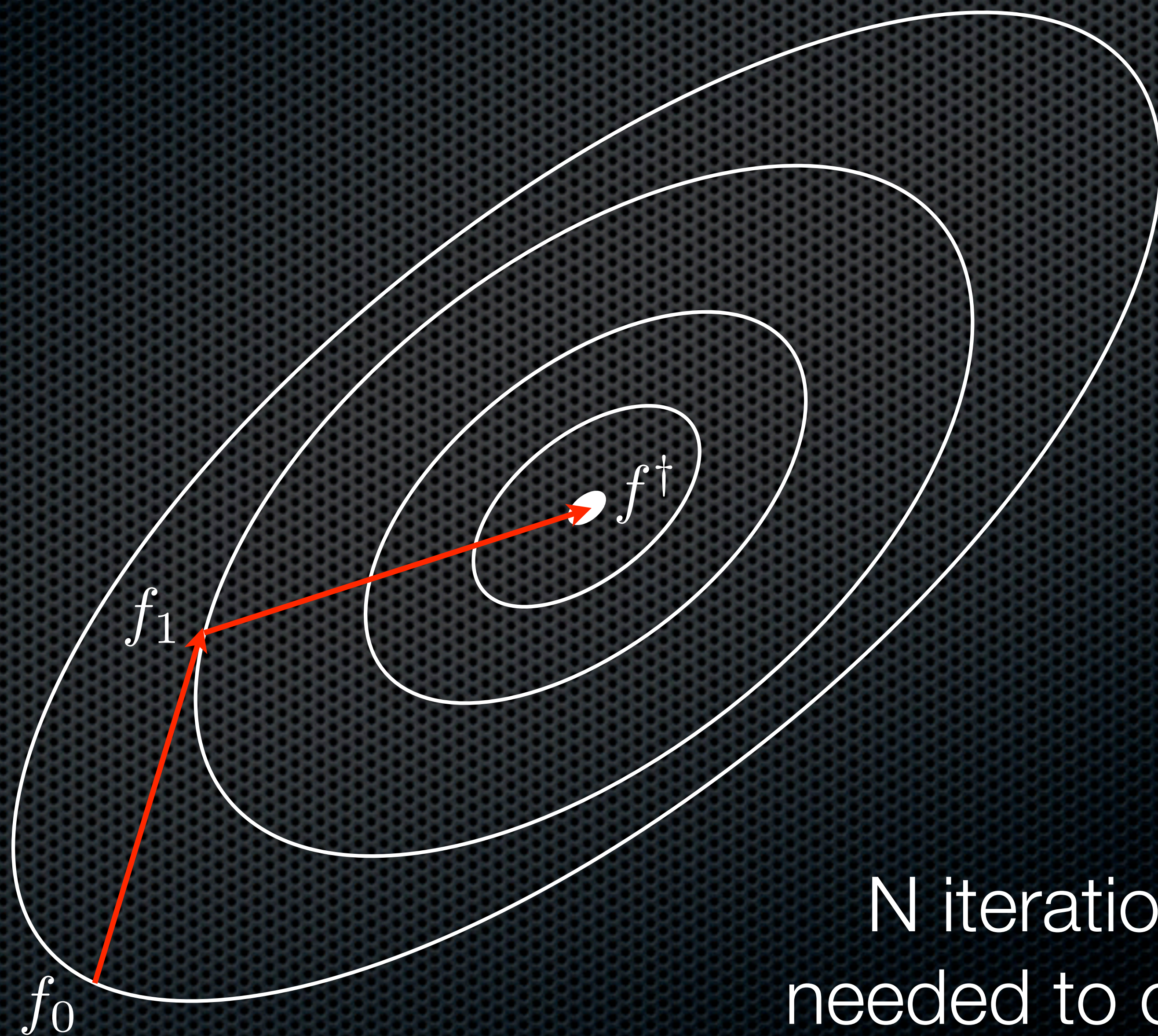
$$r_{k+1} = r_k - \alpha_k A^* A p_k$$

$$\beta_k = -\frac{r_{k+1}^* A^* A p_k}{p_k^* A^* A p_k}$$

$$p_{k+1} = r_{k+1} + \beta_k p_k$$

- Notice that $r_{k+1} \perp r_k \qquad A p_{k+1} \perp A p_k$

CGD: Illustration



N iterations are needed to converge