Regularization Theory

Nicolas Rougon

Institut Mines-Télécom / Télécom SudParis
ARTEMIS Department; CNRS UMR 8145

nicolas.rougon@telecom-sudparis.eu

May 11, 2015

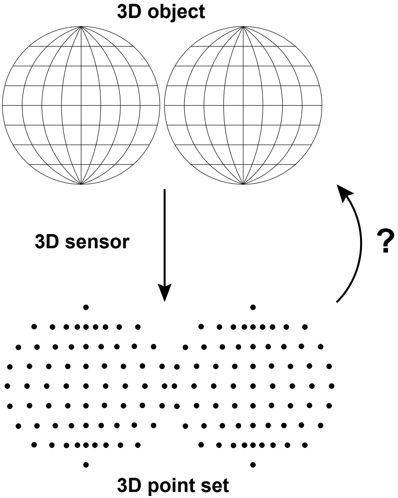
Course overview

- Basic problems in low-level image analysis are hard to solve
- The encountered difficulties have common nature and origins, linked to the key notion of ill-posed inverse problem
- Generic mathematical approaches for fixing these difficulties have been developed in the deterministic and stochastic frameworks. They are known as regularization techniques
- In the next courses, deterministic regularization will be applied to 3 basic image analysis problems

Contour-based segmentation	Region-based segmentation	Motion estimation
Active contours	Mumford-Shah	Optical flow

Outline

An example - Surface reconstruction



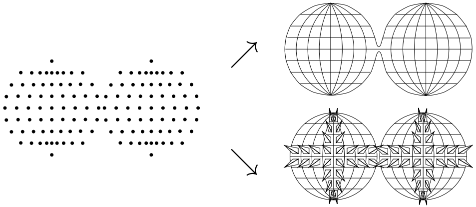
Problem statement

Given a 3D point set, estimate the geometry of the surface (*i.e.* the shape) of the underlying 3D object

- Solving for an interpolation problem over $\mathbb{R}^3$

An example - Surface reconstruction

- A hard problem
  - Topology is lost
    → connectivity? homotopy?
  - Geometry is lost
    → metrics? orientation?

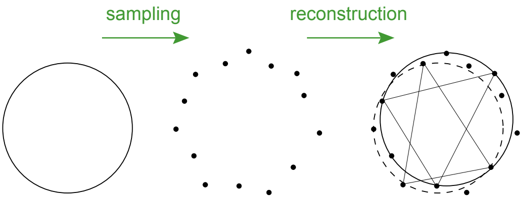


under-constrained problem
⇒ multiple solutions

- **Solution:** enforcing prior constraints
  - Topological constraints
    - connectivity
    - homotopy
    - boundary (open/close)
  - Geometric constraints
    - continuity (smoothness)
    - support (global/patch)
    - subspace

An example - Surface reconstruction

- A hard problem (cont'd)
  - Discretization and noise generate ambiguities



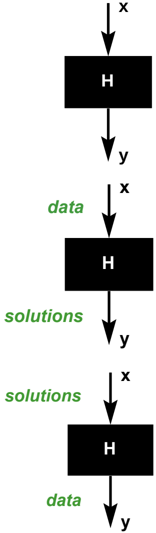
Assume subspace constraints (*e.g.* circle shape space)

over-constrained problem
$\Rightarrow$ solutions may not exist

noise-sensitive problem
$\Rightarrow$ solutions may vary radically with noise

- **Solution:** defining an approximated problem

Direct *vs.* inverse problems



- Imaging system model $H : X \to Y$

  H continuous (linear/nonlinear) operator

  X, Y Hilbert spaces

- Direct problem

  Given an input $x \in X$, generate $y \in Y$ such that $y = Hx$

  $\to$ Image synthesis

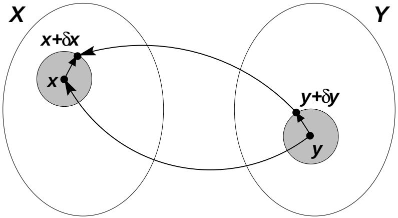
- Inverse problem

  Given the data $y \in Y$, estimate $x \in X$ such that $y = Hx$

  $\to$ Image analysis

Well-posed *vs.* ill-posed problems

- A (direct/inverse) problem is well-posed *iff* the following so-called Hadamard conditions hold
  - for any data $y \in Y$, there exists a solution $x \in X$
  - the solution $x \in X$ is unique
  - the solution $x \in X$ depends continuously on data $y \in Y$



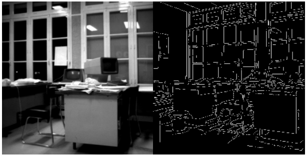
- A problem is ill-posed *iff.* at least one Hadamard condition fails

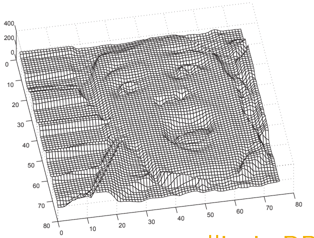
Direct problems

- Since H is continuous, classical direct problems in mathematics, physics and engineering are well-posed
  - Physical simulation problems
    - propagation $\rightarrow$ elliptic PDE + Dirichlet conditions
      e.g. wave equation
    - diffusion $\rightarrow$ parabolic PDE + Neumann conditions
      e.g. heat equation
    - transport $\rightarrow$ hyperbolic PDE + Cauchy conditions
      e.g. eikonal (Burger) equation
  - Image synthesis problems

Inverse problems in image analysis

- Inverse problems in image science are usually ill-posed
  - Segmentation / surface reconstruction / shape from X





- Edge detection

  H is an integral operator

$$(Hx)(\eta) = \int_{-\infty}^{\eta} x(\xi)\, d\xi$$

  η : pixel location

- Surface reconstruction

  H is a projector onto a local basis

$$(Hx)(\eta_i) = < x, \varphi_i >_X$$

  η_i : surface local coordinates

$\rightarrow$ elliptic PDE + Cauchy conditions

Inverse problems in image analysis

- Inverse problems in image science are usually ill-posed
  - Matching

H is a warping operator

$$(Hx)(\eta) = (x \circ \phi)(\eta)$$

η : pixel location

- Motion estimation

  $\phi = \mathbb{I}\mathrm{d} + v$ where v : optical flow

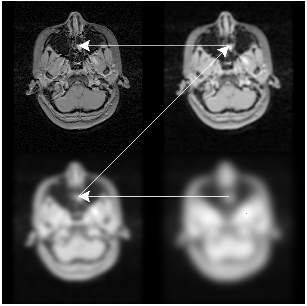
- Stereovision

  $\phi = \mathbb{I}\mathrm{d} + d$ where d : disparity

$\longrightarrow$ Optimal isomorphism

Inverse problems in image analysis

- Inverse problems in image science are usually ill-posed
  - Deconvolution



H is a convolution operator

$$(Hx)(\eta) = (K \star x)(\eta)$$

η : pixel location

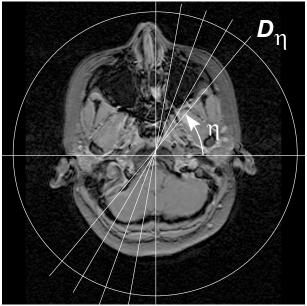
- Deblurring
- Linear scale-space

  K : Gaussian kernel

$\longrightarrow$ Inverse diffusion

Inverse problems in image analysis

- Inverse problems in image science are usually ill-posed
  - Transmission tomography reconstruction



H is a Radon transform

$$(Hx)(\eta) = \int_{D_\eta} x(l)\, dl$$

η : projection angle

- Computerized Tomography (CT)

$\longrightarrow$ Inverse Radon transform

Inverse problems in image analysis

- Inverse problems in image science are usually ill-posed
  - Emission tomography reconstruction
    - Positon Emission Tomography (PET)
    - Single Photon Emission Computerized Tomography (SPECT)

    $\rightarrow$ Inverse Fredholm integral
  - Super-resolution
    - Satellite / Aerial / Astromonical imaging

    $\rightarrow$ Interpolation from multiple data
  - Phase unwarping
    - Synthetic Aperture Radar (SAR)
    - Digital Holographic Microscopy (DHM)

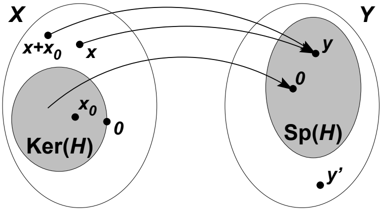
    $\rightarrow$ Analytic continuation
  - ...

Outline

Problem statement

Inverse problem (P)

Given the data $y \in Y$, estimate $x \in X$ such that $y = Hx$



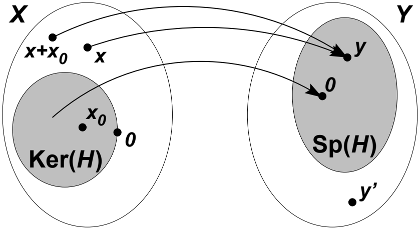
The invertibility of H depends on

- its kernel $\text{Ker}(H) \subseteq X$ (closed set)
- its span $\text{Sp}(H) \subseteq Y$

▶ The problem (P) is well-posed *iff.* H is injective ($\text{Ker}(H) = \{0\}$) and onto ($\text{Sp}(H) = Y$). In this case

- H^{-1} exists and is continuous
- The solution of (P) is $x = H^{-1}y$

Restoring well-posedness



Assumptions

- H is not injective
  $\mathrm{Ker}(H) \neq \{0\}$
- H is not onto
  $\mathrm{Sp}(H) \neq Y$

▶ How to restore the existence and unicity properties?

Key idea

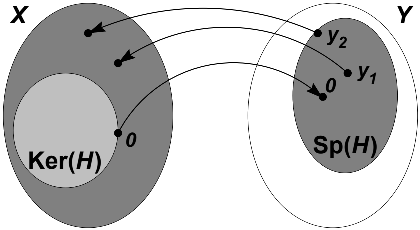
Redefine the problem (P) so that

- $\mathrm{Ker}(H) \cap X = \{0\}$
- $\mathrm{Sp}(H) = Y$

Restoring well-posedness

Track #1

Redefine the data space Y and the solution space X



Assumptions

- H is not injective
  $\mathrm{Ker}(H) \neq \{0\}$
- $\mathrm{Sp}(H)$ is closed
  $Y = \mathrm{Sp}(H) \oplus \mathrm{Sp}^{\perp}(H)$

- The problem (P) is well-posed over
  - $X = \mathrm{Ker}^{\perp}(H)$
  - $Y = \mathrm{Sp}(H)$

Restoring well-posedness

Track #1

Redefine the data space Y and the solution space X

- Bridging the gap from theory to practice is rarely possible since there is no generic algorithms to
  - compute $\text{Ker}^{\perp}(H)$
  - test for $y \in \text{Sp}(H)$

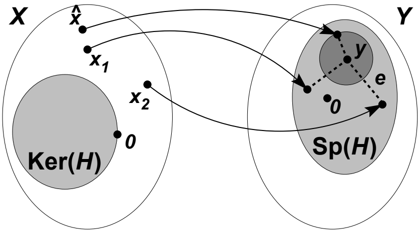
Track #2

Redefine the problem operator H to end up with a computable inverse operator

Problem statement

Idea

Find $x \in X$ such that Hx is as close as possible to data $y \in Y$



Assumptions

- $\text{Sp}(H)$ is closed

Pseudo-solutions

Pseudo-inverse problem $(\hat{\mathrm{P}})$

Given data $y \in Y$, find a minimizer $\hat{x} \in X$ of the error criterion $e(x) = \|Hx - y\|_Y$

$$\hat{x} = \arg\min_{x \in X} \|Hx - y\|_Y$$

- $\hat{x}$ is refered to as a pseudo-solution of (P)
- When $\|\cdot\|_Y$ is the L^2-norm
  - $e(x)$: mean square error
  - $\hat{x}$: least square solution

Pseudo-inverse

- Let H^* be the adjunct operator of H, defined by

$$< Hx, y >_Y = < x, H^* y >_X$$

Normal equations

Pseudo-solutions $\hat{x}$ verify the normal equations

$$H^* Hx \;=\; H^* y$$

- Special case of Euler equations $\partial_x e(x) \;=\; 0$

Pseudo-inverse

The pseudo-inverse of H is the operator $\hat{H} : Y \to X$ defined by

$$\hat{H} = (H^* H)^{-1} H^*$$

- $\hat{H}$ is continuous

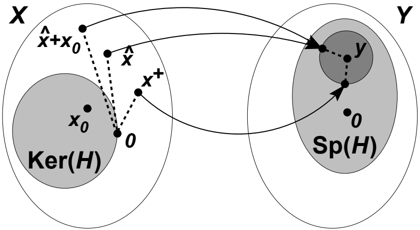
Well-posedness

- Since $\hat{H}$ is always defined, pseudo-solutions always exist
  - Let P_H be a projector on $\mathrm{Sp}(H)$
    The pseudo-solutions of $(\hat{\mathrm{P}})$ for data y and $P_H y$ are the same
- If H is injective
  - The problem $(\hat{\mathrm{P}})$ is well-posed
  - Its solution is $\hat{x} = \hat{H}y$
- If H is not injective
  - The problem $(\hat{\mathrm{P}})$ is under-constrained (multiple solutions)

Problem statement

Idea

Among all pseudo-solutions, retain the one with minimal norm



Assumptions

- $\mathrm{Sp}(H)$ is closed
- H is not injective

Generalized solutions

Generalized inverse problem ($P^\dagger$)

Given $y \in Y$, find a minimizer $x^\dagger \in X$ with minimal norm of the error criterion $e(x) = \|Hx - y\|_Y$

$$x^\dagger = \begin{cases} \arg\min_{x \in X} \|Hx - y\|_Y \\ \arg\min_{x \in X} \|x\|_X \end{cases}$$

- $x^\dagger$ is refered to as a generalized solution of (P)

Generalized inverse

- $x^\dagger$ is unique
  An inverse operator can therefore be defined

Generalized inverse

The generalized inverse of H is the operator $H^\dagger : Y \to X$ such that

$$H^\dagger y = x^\dagger$$

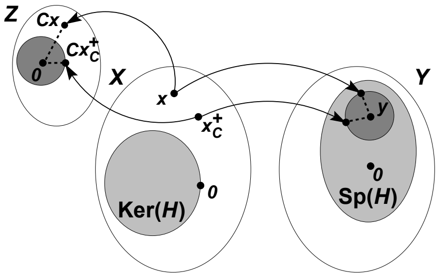
- $H^\dagger$ is continuous
- $x^\dagger \in \mathrm{Ker}^\perp(H)$

- The problem $(\mathrm{P}^\dagger)$ is well-posed *iff.* $\mathrm{Sp}(H)$ is closed
- Generalized inverses may be trivial or unrealistic w.r.t. to (P)

Problem statement

Idea

Introduce prior knowledge by enforcing constraints on solutions



Assumptions

- $\mathrm{Sp}(H)$ is closed
- H is not injective

- constraint space Z (Hilbert)
- linear constraint operator $C : X \to Z$
  - inducing a semi-norm $\|Cx\|_Z$ over X

C-generalized solutions

C-generalized inverse problem ($\mathrm{P}_C^\dagger$)

Given $y \in Y$, find a minimizer $x_C^\dagger$ with minimal constraint norm of the error criterion $e(x) = \|Hx - y\|_Y$

$$x_C^\dagger = \begin{cases} \arg\min_{x \in X} \|Hx - y\|_Y \\ \arg\min_{x \in X} \|Cx\|_Z \end{cases}$$

- $x_C^\dagger$ is refered to as a C-generalized solution of (P)

C-generalized inverse

- $x_C^\dagger$ does not necessarily exist

Theorem

The C-generalized solution $x_C^\dagger$ exists when C verifies

1. $\mathrm{Ker}(C) \cap \mathrm{Ker}(H) = \{0\}$
2. the domain of C is dense in X
3. C is closed
4. the pseudo-solutions in $\mathrm{Ker}(C)$ are mapped by H into a closed set in Y
5. C is bounded

C-generalized inverse

- Under the previous conditions, an inverse operator can be defined

C-generalized inverse

The C-generalized inverse of H is the operator $H_C^\dagger : Y \to X$ such that

$$H_C^\dagger y = x_C^\dagger$$

- $H_C^\dagger$ is bounded
- If C has a bounded inverse

$$H_C^\dagger = C^{-1} \left(H C^{-1} \right)^\dagger$$

Well-posedness

- The problem $(P_C^\dagger)$ can be ill-posed
- Differential operators are not bounded

  Consequently, the C-generalized inverse framework does not allow for integrating

  - smoothness constraints
  - local geometric constraints

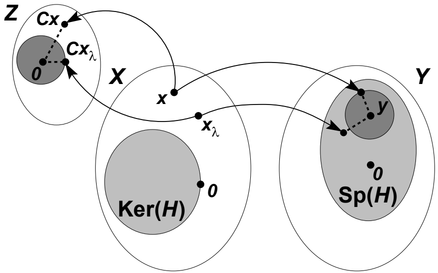
  which play a pivotal role in image analysis

Outline

Problem statement

Idea

Find $x \in X$ ensuring the best trade-off between prior constraints and closeness of Hx to data $y \in Y$



Assumptions

- $\mathrm{Sp}(H)$ is not closed
- H is not injective

- This is the case for kernel operators $Hx = K \star x$

Regularized solutions

Regularized inverse problem (P_λ)

Given $y \in Y$, find a minimizer x_λ of the cost function $E_\lambda(x)$ such that

$$x_\lambda = \arg\min_{x \in X} \left(\|Hx - y\|_Y^2 + \lambda \|Cx\|_Z^2 \right)$$

- x_λ is refered to as a regularized solution of (P)
- The regularization functional (energy) $E_\lambda(x)$ comprises
  - a data consistency term $\|Hx - y\|_Y^2$
  - a model term $\|Cx\|_Z^2$ defining prior constraints on solutions
- The regularization parameter $\lambda > 0$ controls the trade-off between data and prior knowledge
  - when data noise $\nearrow$ (*i.e.* data reliability $\searrow$), the model influence can be strengthened by letting $\lambda \nearrow$

Tikhonov stabilizers

- Admissible constraints

Tikhonov stabilizers

$$\| Cx \|_Z^2 = \sum_{i=1}^{n} \int c_i(\xi) \left| D^i x(\xi) \right|^2 d\xi$$

- The potential function $c_i(\xi) \left| D^i x(\xi) \right|^2$ defines a quadratic smoothness constraint of order i, enforcing C^i-continuity
- The weighting functions $c_i(\xi) > 0$ control locally the smoothness of solutions. They are usually chosen constant

- Oversmoothing artefacts
  - Quadratic smoothness contraints do not allow for preserving salient data discontinuities

Regularized inverse

Euler equations

Regularized solutions x_λ verify the following Euler equations

$$\left[H^*H + \lambda C^*C\right]x = H^*y$$

Regularized inverse

The regularized inverse of H is the operator $R_\lambda : Y \rightarrow X$ defined by

$$R_\lambda = \left[H^*H + \lambda C^*C\right]^{-1} H^*$$

- R_λ is continuous
- $\lim_{\lambda \rightarrow 0} R_\lambda y = H^\dagger y$

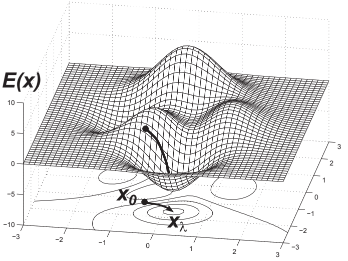
Well-posedness

Theorem (Tikhonov-Arsenin)

The regularized problem (P_λ) is well-posed

- Its solution is $x_\lambda = R_\lambda y$

▶ In many cases, H^* is not defined under closed form
Hence, x_λ cannot be computed analytically



- Minimizing $E_\lambda(x)$ is then performed iteratively using gradient descent techniques

Regularization filtering

- Singular Value Decomposition (SVD) of H

$$H\,u_k \;=\; \alpha_k\,v_k \qquad\qquad H^*\,v_k \;=\; \alpha_k\,u_k$$

  - u_k (v_k): singular functions of H (H^*)
  - α_k: singular values of H $\left(\lim_{k\to\infty} \alpha_k \;=\; 0\right)$

- Let c_k^2 be the coordinates of C^*C over the basis (u_k)
  The pseudo-solution $\hat{x}$ and regularized solution x_λ of (P) verify

$$\hat{x} \;=\; \sum_k \frac{1}{\alpha_k}\,\langle y, v_k\rangle_Y\;u_k$$

$$x_\lambda \;=\; \sum_k \frac{\alpha_k^2}{\alpha_k^2 + \lambda\,c_k^2}\,\frac{1}{\alpha_k}\,\langle y, v_k\rangle_Y\;u_k$$

  - x_λ is a filtered version of $\hat{x}$
  - x_λ coordinates are bounded due to the constraint terms c_k^2

Nonlinear regularization

- Tikhonov theory extends to nonlinear regularization based on non-quadratic discontinuity-preserving stabilizers

1^{st}-order discontinuity-preserving stabilizers

$$\|Cx\|_Z^2 = \int \varphi(|Dx(\xi)|)\, d\xi$$

1. $\varphi(\sqrt{u})$ positive, strictly concave
2. $\frac{\varphi'(u)}{u}$ bounded, $\varphi''(u) > 0$
3. $\lim_{u \to 0} \frac{\varphi'(u)}{u} = \lim_{u \to 0} \varphi''(u) = 1$
4. $\lim_{u \to +\infty} \frac{\varphi'(u)}{u} = \lim_{u \to +\infty} \varphi''(u) = 0$
5. $\lim_{u \to +\infty} \frac{\varphi''(u)}{\left(\frac{\varphi'(u)}{u}\right)} = 0$

Nonlinear regularization

- Some classical discontinuity-preserving norms

name	$\varphi(u)$	$\frac{\varphi'(u)}{u}$
Welsch	$1 - e^{-u^2}$	e^{-u^2}
Cauchy	$\ln\left(1 + u^2\right)$	$\frac{1}{1+u^2}$
Geman-McClure	$\frac{u^2}{1+u^2}$	$\frac{1}{(1+u^2)^2}$
Green	$\ln \cosh(u)$	$\frac{\tanh(u)}{u}$
L_1 - L_2	$\sqrt{1 + u^2} - 1$	$\frac{1}{\sqrt{1+u^2}}$
Total Variation (TV)	u	$\frac{1}{u}$

Hyperparameter estimation

- Optimal values for the regularization parameter λ can be derived when upper-bounds ε_s on $\|Cx\|_Z$ or ε_d on $\|Hx - y\|_Y$ (data noise) are known
  - ε_s known: λ is estimated as a Lagrange multiplier

$$\begin{array}{ll} & \arg\min_{x \in X} \|Hx - y\|_Y \\ \text{u.c.} & \|Cx\|_Z \leq \varepsilon_s \end{array} \quad \Rightarrow \quad \exists \lambda \quad \begin{array}{l} x_\lambda = \arg\min_{x \in X} E_\lambda(x) \\ \|Cx_\lambda\|_Z = \varepsilon_s \end{array}$$

  - ε_d known: λ is chosen using Morozov's discrepancy principle

$$\begin{array}{ll} & \arg\min_{x \in X} \|Cx\|_Z \\ \text{u.c.} & \|Hx - y\|_Y \leq \varepsilon_d \end{array} \quad \Rightarrow \quad \exists \lambda \quad \|Hx_\lambda - y\|_Y = \varepsilon_d$$

  - $\varepsilon_s, \varepsilon_d$ known: $\lambda = \left(\frac{\varepsilon_d}{\varepsilon_s}\right)^2$

Hyperparameter estimation

- When these upper-bounds are not available, a sub-optimal value for the hyperparameter λ can be derived using cross-validation techniques
- There is no reference method
  - Most often, hyperparameters are tuned empirically from systematic experimentation over a reference database

Stochastic regularization

- Data y and unknowns x are modeled as random variables

Stochastic inverse problem (P)

Given noisy data $y \in Y$, estimate $x \in X$ such that $y = Hx + n$

- Gaussian case
  - Data: n is a white Gaussian noise with covariance $\Sigma_n = \sigma_n^2 \mathbb{I}$

$$p(y|x) \propto \exp\left(-\frac{1}{2\sigma_n^2}\|Hx - y\|_Y^2\right)$$

  - Model: x is a centered Gaussian process with covariance Σ_x

$$p(x) \propto \exp\left(-\frac{1}{2}\langle x, \Sigma_x^{-1} x\rangle_X\right)$$

Stochastic regularization

- Maximum A Posteriori (MAP) estimate of x

$$x^* = \arg\max_{x \in X} p(x|y)$$

  - Using Bayes' formula: $p(x|y) \propto p(y|x)\, p(x)$

$$x^* = \arg\min_{x \in X} \left[\frac{1}{\sigma_n^2} \left(\|Hx - y\|_Y^2 + \sigma_n^2 \langle x, \Sigma_x^{-1} x \rangle \right) \right]$$

  - The MAP criterion is similar to a regularization functional
    - $\|Hx - y\|_Y^2$ is a data consistency term
    - $\langle x, \Sigma_x^{-1} x \rangle$ is a prior model term
    - σ_n^2 is a regularization parameter

- These conclusions hold if x obeys an arbitrary exponential law

Wiener filtering

Euler equations

The MAP estimate x^* verify the following Euler equations

$$\left[H^*\Sigma_n H + \Sigma_x\right]x = H^*\Sigma_n y$$

Wiener-Kolmogorov filter

The Wiener-Kolmogorov filter $\mathcal{F}_W : Y \to X$ is defined by

$$\mathcal{F}_W = \left[H^*\Sigma_n H + \Sigma_x\right]^{-1} H^*\Sigma_n$$

- The MAP estimate is $x^* = \mathcal{F}_W y$

Markov Random Fields

- Let x be a random variable indexed by a discrete spatial variable η

Markov Random Field (MRF)

x is a MRF *iff.* the probability of observing $x(\eta)$ at η depends only on the realizations of x in some neighborhood $\mathcal{N}(\eta)$ of η

Theorem (Hammersley-Clifford)

x is a MRF *iff.*

$$p(x(\eta)) \propto \exp\left(-V(\mathcal{N}(\eta))\right)$$

for some neighborhood operator $V(\mathcal{N}(\eta))$ refered to as a potential

- Differential kernels are neighborhood operators
- Consequently, MRFs allow for enforcing discrete smoothness constraints within the MAP estimation framework

Markov Random Fields

- Markov Random Fields
  - Model: x is a 1^{st}-order MRF *i.e.*

$$p(x(\eta)) \propto \exp\Big(-\sum_{c \in \mathcal{C}^1(\eta)} V(c)\Big)$$

for some potential V operating on 1^{st}-order maximal cliques

$$\mathcal{C}^1(\eta) = \{ (\eta, \xi) \mid \xi \in \mathcal{N}(\eta) \}$$

  - $\sum_c V(c)$ can be chosen as a gradient filter norm

Deterministic *vs.* stochastic regularization

- Deterministic and stochastic regularization optimize criteria with similar structure but context-specific interpretations, *i.e.* a cost function $E_\lambda(x)$ or a posterior conditional probability $p(x|y)$

- They differ on the optimization approach
  - Deterministic regularization uses either variational methods to zeroe the derivative $\partial_x E_\lambda(x)$, or graph-cut methods to directly minimize $E_\lambda(x)$
  - Stochastic regularization uses Monte-Carlo or graph-cut methods to directly maximize the posterior $p(x|y)$

Deterministic *vs.* stochastic regularization

modeling	optimization approach	criterion differentiability	detected extrema
deterministic	variational	yes	local*
	graph-cut	no	global
stochastic	Monte-Carlo	no	global
	graph-cut	no	global

* in the vicinity of the initialization

- Direct optimization approaches apply to a broader class of criteria and provide increased robustness w.r.t. spurious local extrema induced by noise