METHODS FOR IMAGE RESTORATION

Michele Piana
Dipartimento di Informatica
Universita' di Verona

Deterministic approach

DETERMINISTIC APPROACH

Functional analysis framework:

Solution space X: $\int |f(x)|^2 dx < \infty$

Data space Y: $\int |g(y)|^2 dy < \infty$

Linear continuous operator A:X→Y

The (linear) inverse problem consists in determining f in X from the knowledge of g in Y when f and g are related by

$$g = Af$$

DETERMINISTIC APPROACH

Remarks:

- X and Y are broad, since they must contain all possible solutions and all possible (noisy) data

- The L2 condition for X and Y means that all the signals in the game must have finite energy

- In a finite-dimensional framework, X and Y are Euclidean spaces and A is a matrix

ILL-POSEDNESS

Kernel: $N(A) = \{ f \in X \quad Af = 0 \}$

Range: $R(A) = \{ g \in Y \quad \exists f \in X, Af = g \}$

Inverse: $A^{-1} \quad A^{-1}A = AA^{-1} = I$

The solution is not unique $\Leftrightarrow N(A) \neq 0$

The solution does not exist $\Leftrightarrow R(A) \neq Y$

The solution does not depend continuously on the data

$\Leftrightarrow A^{-1}$ is not continuous

ILL-POSEDNESS

Well-posedness does not imply stability:

$$\frac{\|\delta f\|}{\|f\|} \leq \|A\|\|A^{-1}\| \frac{\|\delta g\|}{\|g\|}$$

Ill-conditioning:

If $C(A) = \|A\|\|A^{-1}\|$ is large the problem is numerically unstable

Remark: finite-dimensional (well-posed) problems coming from the discretization of ill-posed problems are ill-conditioned

PSEUDOSOLUTIONS

$$\|Af - g\| = \min \Leftrightarrow A^* A f = A^* g$$

The set of least-squares solution is closed and convex in X

$$\Downarrow$$

there exists a unique least-squares solution with minimum norm, the generalized solution f^+

Generalized inverse: $A^+ : Y \rightarrow X \quad A^+ g = f^+$

PSEUDOSOLUTIONS

Remark 1: ill-conditioning:

$$\frac{\|\delta f^+\|}{\|f^+\|} \leq \|A\| \|A^+\| \frac{\|\delta g\|}{\|g\|}$$

Remark 2: compact operators:

$$(Af)(y) = \int K(y,x) f(x) dx \qquad \iint |K(y,x)|^2 dy dx < \infty$$

A^+ is not continuous

REGULARIZATION

Given the linear inverse problem $g = Af$
a regularization algorithm is a one-parameter family of operator $\{R_\lambda\}_{\lambda>0}$ such that:

1. $R_\lambda : Y \to X$ is bounded

2. $\lim_{\lambda \to 0} \| R_\lambda A f^+ - f^+ \| = 0$

REGULARIZATION

A simple model for image formation:

$$g_\delta = Af^+ + w_\delta \qquad \|w_\delta\| \le \delta$$

$$\Downarrow$$

$$\|R_\lambda g_\delta - f^+\| \le \|R_\lambda A f^+ - f^+\| + \delta \|R_\lambda\|$$

Remark: a regularization algorithm works fine if one can find an optimal $\lambda_{opt} = \lambda_{opt}(\delta)$ such that

3. $\lim_{\delta \to 0} \lambda_{opt}(\delta) = 0$

4. $\lim_{\delta \to 0} \|R_\lambda g_\delta - f^+\| = 0$

Remark: finding $\lambda_{opt} = \lambda_{opt}(\delta)$ is difficult!!

TIKHONOV METHOD

One-parameter family of minimum problems:

$$\|Af - g\|_Y^2 + \lambda\|f\|_X^2 = \min \quad \Leftrightarrow \quad (A^*A + \lambda I)f = A^*g$$

Main result: $R_\lambda = (A^*A + \lambda I)^{-1} A^*$ is a regularization algorithm

Remark: $\| R_\lambda \| \leq \dfrac{1}{\sqrt{\lambda}}$

An optimal choice of the regularization parameter is such that:

$$\lim_{\delta \to 0} \lambda_{opt}(\delta) = 0$$

$$\lim_{\delta \to 0} \frac{\delta}{\sqrt{\lambda_{opt}(\delta)}} = 0$$

TIKHONOV: COMPUTATION

Matrices or (compact) operators:

Singular system:

$$Au_k = \sigma_k v_k \qquad A^* v_k = \sigma_k u_k \qquad k = 1,\dots, p$$

Tikhonov regularized solution:

$$f_\lambda = \sum_{k=1}^{p} \frac{\sigma_k}{\sigma_k^2 + \lambda} (g, v_k) u_k$$

TIKHONOV: COMPUTATION

Proof:

$$f_\lambda = \sum_{k=1}^{p} (f_\lambda, u_k) u_k + u \qquad Au = 0$$

$$u = 0 \Rightarrow (A^* A + \lambda I) f_\lambda = \sum_{k=1}^{p} (f_\lambda, u_k)(\sigma_k^2 + \lambda) u_k$$

$$g = \sum_{k=1}^{p} (g, v_k) v_k + v \qquad A^* v = 0$$

$$A^* g = \sum_{k=1}^{p} (g, v_k) \sigma_k u_k$$

$$(f_\lambda, u_k) = \frac{\sigma_k}{\sigma_k^2 + \lambda} (g, v_k)$$

TIKHONOV: COMPUTATION

Convolution: $(Af)(x) = (K * f)(x)$

Tikhonov regularized solution:

$$\hat{f}_\lambda(\omega) = \frac{\overline{\hat{K}}(\omega)\hat{g}(\omega)}{\left|\hat{K}(\omega)\right|^2 + \lambda}$$

Remark: Tikhonov method is nothing but a linear filter!!

TIKHONOV: COMPUTATION

Proof:

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

$$(A^*Af\hat{)}(\omega) = |\hat{K}(\omega)|^2 \ \hat{f}(\omega)$$

$$(A^*g\hat{)}(\omega) = \overline{\hat{K}}(\omega)\hat{g}(\omega)$$

The Euler equation becomes:

$$(|\hat{K}(\omega)|^2 + \lambda)\hat{f}(\omega) = \overline{\hat{K}}(\omega)\hat{g}(\omega)$$

TIKHONOV: COMMENTS

- Generalization: $\left\| Af - g \right\|_Y^2 + \lambda \left\| Cf \right\|_X^2 = \min$

  Example: C is a differential operator

- Computational heaviness for general operators

- Difficulties in accounting for sophisticated a priori constraints

- General reconstruction ‘behaviours’: effective in smoothing, less effective in enhancing edges or resolving small features

DIGRESSION: LEARNING

- X compact in R^d ; Y compact in R ; $Z = X \times Y$
- $\rho : B(Z) \to [0,1]$ probability distribution such that:

ν, $\nu(E) = \rho(E \times Y)$ $E \in B(X)$ marginal distribution

$\rho(y \mid x)$, $d\rho(x, y) = d\rho(y \mid x) d\nu(x)$ conditional distribution

- $V : R \times Y \to [0, \infty]$ is the loss function such that

$I[f] = \int_Z V(f(x), y) d\rho(x, y)$ is the expected risk

- $D = \{(x_i, y_i)\ i = 1, \ldots l\}$ is the training set, where all the pairs are selected according to $\rho(x, y)$

THE LEARNING PROBLEM

Let $\mathcal{F}$ be the target space; $f_0 : X \to Y$ the target function such that

$$I[f_0] = \min_{f \in \mathcal{F}} I[f]$$

Then $\forall D \in Z^l$ determine $f_D \in H$ such that:

$\forall \eta,\ 0 < \eta < 1 \quad \exists \varepsilon = \varepsilon_l(\eta),$

$$\Pr\{I[f_D] - I[f_0] \leq \varepsilon\} \geq 1 - \eta$$

H is the hypothesis space and it is a RKHS

The map $D \to f_D$ is a learning algorithm

THE LEARNING PROBLEM

Regularization networks:

$$\frac{1}{l}\sum_{i=1}^{l} V(f(x_i), y_i) + \lambda \|f\|_H = \min$$

Quadratic (Euclidean) loss function → Tikhonov problem

More general loss functions: the generalization capability increases but the computational effectiveness descreases

Remark: there is still the problem of an optimal choice for λ

LANDWEBER METHOD

Successive approximations:

$$g = Af$$

$$f = f + g - Af$$

$$f_{n+1} = f_n + g - Af_n$$

$$f_{n+1} = f_n + \tau(g - Af_n)$$

Landweber method=successive approximations applied to the least-squares problem (i.e., to the Euler equation):

$$f_{n+1} = f_n + \tau A^*(g - Af_n) \quad f_0 = 0 \quad 0 < \tau < \frac{2}{\|A\|^2}$$

LANDWEBER METHOD: COMPUTATION

Matrices and compact operators:

$$f_n = \sum_{k=1}^{p} \left\{ 1 - \left[1 - \tau \sigma_k^2 \right]^n \right\} \frac{(g, v_k)}{\sigma_k} u_k \qquad 0 < \tau < \frac{2}{\sigma_1^2}$$

Convolution:

$$\hat{f}_n(\omega) = \left\{ 1 - \left[1 - \tau \left| \hat{K}(\omega) \right|^2 \right]^n \right\} \frac{\hat{g}(\omega)}{\hat{K}(\omega)} \qquad 0 < \tau < \frac{2}{\left| \hat{K}(\omega) \right|_{\max}^2}$$

Remark: even the Landweber method is a linear filter

LANDWEBER METHOD: COMMENTS

- The trade-off between stability and fitting is realized by optimally stopping the iteration: high n means good fitting/bad stability; small n means bad fitting/high stability

- Tikhonov method and Landweber method behave pretty much the same

- There are partial results and many heuristics for fixing τ

PROJECTIONS

In many problems it is possible to a priori know that the solution belongs to a closed and convex subset C of the solution space

↓

Constrained least-squares problems:

$$\|Af - g\| = \min \quad f \in C$$

PROJECTIONS

Projected Landweber method:

$$f_{n+1} = P_C[f_n + \tau A^*(g - Af_n)] \quad f_0 = 0 \quad 0 < \tau < \frac{2}{\|A\|^2}$$

P_C is the projection operator onto the closed and convex subset C of the solution space

Remark: this projection is well-defined for two reasons:

- C is closed and convex
- the solution space has ‘good’ properties

PROJECTIONS

Super-resolution:

projections onto closed convex subsets of ‘good’ solution spaces implies the regularity of the Fourier transform of the regularized solution

Examples:

- compactly supported functions: the more the support is constrained the wider the band

- positive function: a general theorm (Paley-Wiener) guarantees the regularity of the Fourier transform

PROJECTIONS

Comments on the projected Landweber method:

- many open (interesting) issues concerned with convergence
- computational heaviness: preconditioned versions