# Regularized Posteriors in Linear Ill-posed Inverse Problems

Jean-Pierre Florens[1] Anna Simoni[2]

[1]Université de Toulouse I
GREMAQ and IDEI

[1]Université de Toulouse I
GREMAQ

8 August 2007

# Inverse Problems: definition

An Inverse Problem describes an *implicit relation* between the object of interest $x$ and the observed one $y$.

**General Formulation:** let $\mathcal{X}$ and $\mathcal{Y}$ be separable Hilbert Spaces, $x \in \mathcal{X}$, $y \in \mathcal{Y}$

$$y = Kx, \tag{1}$$

*i.e.* $y$ is linked to $x$ through a transformation made by the operator $K$ ($K : \mathcal{X} \to \mathcal{Y}$, $K$ compact) and inverse problems are concerned with determining causes for a desired or an observed effect. Usually, data are contaminated by errors, so that instead of the exact equation (1), we have an approximate equation $y \approx Kx$.

**Well-posed Inverse Problems** if (*Hadamard's conditions*)

- a solution **exists** (*i.e.* $y \in \mathcal{R}(K)$);
- the solution is **unique** (*i.e.* $\mathcal{N}(K) = \{0\}$);
- the solution **depends continuously on the data** (*i.e.* $K^{-1} : \mathcal{R}(K) \to \mathcal{X}$ is continuous).

Otherwise an Inverse Problem is said to be ill-posed.

**References:** Engle, Hanke & Neubauer (1996) [3], Kress (1989) [8].

# Inverse Problems: examples

**X-Ray Tomography (Radon Inversion)**: let $D \subseteq \mathbb{R}^2$ be a compact domain with a spatially varying density $f \Rightarrow$ the aim is to recover $f$ from X-ray measurements in the plane where $D$ lies.

**Inverse Problems in Physics**: time resolved fluorescence, inverse potential problems

**Inverse Problems in Signal and Image Processing**: deblurring images, extrapolating a band- or time-limited signal.

**Inverse Problems in Heat Conduction**: backwards heat equation, determining thermal parameters of the body material from temperature measurements, determining the temperature on an inaccessible part of the boundary.

**Inverse Problems in Statistical Learning Theory**: density estimation, conditional probability and conditional density estimation.

**Inverse Problems in Econometrics**: *Structural Models* $\Rightarrow$ there exists an implicit relation between the functional parameter of interest $\varphi$ and the *c.d.f.* $F$.

*Examples*: Generalized Method of Moments, Linear Regression (with possibly many regressors), Deconvolution, Nonparametric Instrumental Variable estimation, Additive Regression models, Game Theoretic model.

# Solution of ill-posed Inverse Problems

## Best Approximate Solution

- if $y \notin \mathcal{R}(K)$ we look for the *least-squares solution* $x$, *i.e.* $x$ is such that
$$||Kx - y|| = \inf\{||Kz - y||; z \in \mathcal{X}\};$$
- if $\mathcal{N}(K) \neq \{0\}$ we look for the solution of minimal norm;
- if $y \notin \mathcal{R}(K)$ and $\mathcal{N}(K) \neq \{0\}$ we look for the *best approximate solution* $x^\dagger$, *i.e.* the least-squares solution with minimal norm:
$$||x|| = \inf\{||z||; z \text{ is least-squares solution of } y = Kx\}.$$
The best approximate solution is the Moore-Penrose generalized inverse of $K$.

## Regularized solution

Even if a best approximate solution satisfies the first two Hadamard's conditions it can still not depend continuously on the data. Thus, *small errors in $y$ may cause errors of arbitrary size in $x^\dagger$ $\Rightarrow$ $x^\dagger$ **must be regularized**.*

**Example:** Tikhonov Regularization

$$x_\alpha = (\alpha I + K^* K)^{-1} K^* y;$$

*i.e.* $x_\alpha = \arg\min ||Kx - y||^2 + \alpha||x||^2$.

**Other Regularization scheme:** Spectral cut-off; Landweber-Fridman; Iterated Tikhonov.

# Bayesian Analysis of Inverse Problems

The *statistical inversion approach* is based on the following principles:

1. All variables included in the model are modelled as random variables;
2. the randomness describes our degree of knowledge regarding their realizations;
3. the information concerning these variables is coded in the probability distribution;
4. the solution of the inverse problem is the posterior probability distribution that is then used to produce single estimates.

**Remark:** the Bayesian method is an attempt to remove the ill-posedness by restating the inverse problem as a *well-posed extension* in a larger space of probability distributions.

**References:** Kaipio and Somersalo (2005) [7].

# The Classical Bayesian Analysis of Functional Equations

Consider Model (1), where we explicitly denote the noisy observations $\hat{Y}$ and restate an exact relation by adding an error term $U$:

$$\hat{Y} = Kx + U, \qquad x, \hat{Y}, U \in L^2.$$

## Examples in the literature:

- White noise model: $dY(t) = f(t)dt + \frac{\sigma^2}{\sqrt{n}} dB(t)$, $f \in L^2([-1,1])$, $B(t)$ is a Brownian Motion.
- Nonparametric Regression: $Y_i = f(X_i) + \varepsilon_i$, $X_i \in [0,1]$, $\mathbb{E}(\varepsilon_i) = 0$ and $f \in L^2([0,1])$.
- Density Estimation.

## How people deal with these problems?

- They define a sampling measure induced by $U$ with covariance operator proportional to the identity operator $I$;
- project the model on an orthonormal basis $\{\phi_j\}_{j=1}^{\infty}$ of $L^2$ and recover the Fourier coefficients $\theta_j =< Kx, \phi_j >$;
- put a prior Gaussian measure on $\ell_2 = \{\theta; \sum_j \theta_j^2 < \infty\}$, where $\theta = \{\theta_i\}$, with mean $\theta_0 \in \ell_2$ and covariance matrix $\Omega_0 = \ diag(\sigma_{jj}, j = 1, 2, \ldots)$ such that $\sum \sigma_{jj} < \infty$;
- compute the posterior distribution in the classical way by working with infinite dimensional matrices.

# The Classical Bayesian Analysis of Functional Equations

- **Drawback:** these specifications of the model do not well define the functional equation in $L^2$ because the trajectories of $U$ are not in $L^2$ with probability 1.
- **Nuclearity assumption:** to be sure the trajectories of $U$ and $x$ are in $L^2$ their ***covariance operators*** must be ***nuclear***.
- **Nonconvergent Bayes solution in the projected model:** If we put a Gaussian measure on $L^2$, satisfying the nuclearity assumption, and then project the model on an orthonormal basis we get non-invertible covariance matrices.

  $\hookrightarrow$ Regularization is accomplished by considering only a finite number of projections.

  $\hookrightarrow$ If we consider an infinite number of projection directions, there is not way to make the posterior mean and variance bounded.

  $\hookrightarrow$ It is possible to get consistency and boundedness only for particular assumptions on the *Singular Value Decomposition* and on the speed of convergence to zero of the *Singular Values* of $K$ and of the covariance operators.

**Remark:** in finite dimensional case (*i.e.* finite number of projections) and for centered gaussian, with spherical variances, prior and sampling distributions the *posterior mean estimator* of $x$ and the *Tikhonov solution* are the same, with regularization parameter $\alpha$ equal to the ratio of the noise and prior variances.

# Our Solution of Inverse Problems (1)

We do not use the normal projected model and *we make inference directly in function space*. We specify a Gaussian probability measure for the parameter of interest $x$.

**Mathematical Specification:**

$$Y = Kx, \qquad x \in L^2_\pi, \quad Y \in L^2_\rho$$

$K : L^2_\pi \to L^2_\rho$ is a known, injective, Hilbert-Schmidt operator. $K^*$ denotes the adjoint of $K$.

**Statistical Specification:**

$$\hat{Y} = Kx + U, \tag{2}$$

where $x$ and $U$ are Hilbert-random variables.
We assume that $x$ and $U$ induce Gaussian measures on $L^2_\pi$ and $L^2_\rho$: the *prior probability* $\mu$ and the *Sampling Probability* $P^x$, respectively. ⇒ see Assumptions 1 & 2 below.

↪ $\hat{Y}$ is a function of the whole observed sample: we transform a sample of discrete observations (*i.e.* a discrete object) in a continuous function of another variable (*i.e.* infinite dimensional object) (see the *Instrumental Regression* example below).

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\dots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} \ ).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\ldots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} \;).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\dots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} \;).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\dots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} ).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\dots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} \;).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\ldots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} \quad ).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Example in Econometrics: Instrumental Regression

$(G, Z, W) \in \mathbb{R} \times \mathbb{R}^p \times \mathbb{R}^q$, with *cdf* $F$. We define the *Instrumental Regression* $\varphi(Z)$ as

$$\varphi(Z) \in L^2_F(Z) \quad \text{such that} \quad G = \varphi(Z) + \varepsilon, \qquad \mathbb{E}(\varepsilon|W) = 0, \qquad Var(\varepsilon|W) = \sigma^2$$

*i.e.* $\varphi(Z)$ is solution of the *Fredholm Integral equation of I type:* $\mathbb{E}(G|W) = \mathbb{E}(\varphi(Z)|W)$.

1. projection on $L^2_F(Z)$ (to have asymptotic properties of the estimator of the LHS):

$$\mathbb{E}(\mathbb{E}(G|W)|Z) = \mathbb{E}(\mathbb{E}(\varphi(Z)|W)|Z).$$

2. $(G_i, Z_i, W_i)_{i=1,\dots,n}$, $\sim iid$ sample (**discrete object**)
   - $F$ **unknown**: nonparametric estimate of $\mathbb{E}(\mathbb{E}(G|W)|Z)$ (**infinite dimensional object**) and of the operator $K = \mathbb{E}(\mathbb{E}(\cdot|W)|Z) : L^2_F(Z) \to L^2_F(Z)$.
   - $F(\cdot, Z, W)$ **known**: nonparametric estimate of $\mathbb{E}(G|W)$ (**infinite dimensional object**).

(**Example**: *Nadaraya-Watson estimator*, with $K$ a generalized Kernel function

$$\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z) = \sum_j \frac{\sum_i G_i K(W_j - W_i)}{\sum_l K(W_j - W_l)} \frac{K(Z - Z_j)}{\sum_l K(Z - Z_l)} \; ).$$

3. The model becomes

$$\underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(G|W)|Z)}_{\hat{Y}} = \underbrace{\hat{\mathbb{E}}(\hat{\mathbb{E}}(\varphi(Z)|W)|Z)}_{\hat{K}\varphi} + U, \qquad \sqrt{n}\, U \to \mathcal{N}(0, \sigma^2 K).$$

# Our Solution of Inverse Problems (2)

## Assumption 1

Let $U$ be a measurable map on $(\Omega, \mathcal{A}, P)$ with values in an Hilbert space $L^2_\rho$ such that $\mathbb{E}(||U||^2_{L^2_\rho}) < \infty$. Moreover, $U$ is a zero-mean Gaussian Hilbert random variable, *i.e.* $\forall \psi \in L^2_\rho$, its characteristic function $\phi(\psi)$ has the form

$$\phi(\psi) = exp\{i < \psi, \mathbb{E}U >_{L^2_\rho} - \frac{1}{2} < \Sigma\psi, \psi >_{L^2_\rho}\}, \tag{3}$$

where $\Sigma := \mathbb{E}((U - \mathbb{E}U) \otimes (U - \mathbb{E}U))$ is the covariance operator and $\mathbb{E}U = 0$.

## Assumption 2

Let $x$ be a measurable map on $(\Omega, \mathcal{A}, P)$ with values in an Hilbert space $L^2_\pi$ such that $\mathbb{E}(||x||^2_{L^2_\pi}) < \infty$. Moreover, $x$ is a Gaussian Hilbert random variable with mean function $x_0$, *i.e.* $\forall \varphi \in L^2_\pi$, its characteristic function $\phi(\varphi)$ has the form

$$\phi(\varphi) = exp\{i < \varphi, x_0 >_{L^2_\pi} - \frac{1}{2} < \Omega_0\varphi, \varphi >_{L^2_\pi}\}, \tag{4}$$

where $\Omega_0 := \mathbb{E}((x - x_0) \otimes (x - x_0))$ is the covariance operator.

## Assumption 3

$\Sigma$ and $\Omega_0$ are injective operators.

# Our Solution of Inverse Problems (2)

## Assumption 1

Let $U$ be a measurable map on $(\Omega, \mathcal{A}, P)$ with values in an Hilbert space $L^2_\rho$ such that $\mathbb{E}(||U||^2_{L^2_\rho}) < \infty$. Moreover, $U$ is a zero-mean Gaussian Hilbert random variable, *i.e.* $\forall \psi \in L^2_\rho$, its characteristic function $\phi(\psi)$ has the form

$$\phi(\psi) = exp\{i < \psi, \mathbb{E}U >_{L^2_\rho} - \frac{1}{2} < \Sigma\psi, \psi >_{L^2_\rho}\}, \tag{3}$$

where $\Sigma := \mathbb{E}((U - \mathbb{E}U) \otimes (U - \mathbb{E}U))$ is the covariance operator and $\mathbb{E}U = 0$.

## Assumption 2

Let $x$ be a measurable map on $(\Omega, \mathcal{A}, P)$ with values in an Hilbert space $L^2_\pi$ such that $\mathbb{E}(||x||^2_{L^2_\pi}) < \infty$. Moreover, $x$ is a Gaussian Hilbert random variable with mean function $x_0$, *i.e.* $\forall \varphi \in L^2_\pi$, its characteristic function $\phi(\varphi)$ has the form

$$\phi(\varphi) = exp\{i < \varphi, x_0 >_{L^2_\pi} - \frac{1}{2} < \Omega_0\varphi, \varphi >_{L^2_\pi}\}, \tag{4}$$

where $\Omega_0 := \mathbb{E}((x - x_0) \otimes (x - x_0))$ is the covariance operator.

## Assumption 3

$\Sigma$ and $\Omega_0$ are injective operators.

# Our Solution of Inverse Problems (2)

## Assumption 1

Let $U$ be a measurable map on $(\Omega, \mathcal{A}, P)$ with values in an Hilbert space $L^2_\rho$ such that $\mathbb{E}(||U||^2_{L^2_\rho}) < \infty$. Moreover, $U$ is a zero-mean Gaussian Hilbert random variable, *i.e.* $\forall \psi \in L^2_\rho$, its characteristic function $\phi(\psi)$ has the form

$$\phi(\psi) = exp\{i < \psi, \mathbb{E}U >_{L^2_\rho} - \frac{1}{2} < \Sigma\psi, \psi >_{L^2_\rho}\}, \tag{3}$$

where $\Sigma := \mathbb{E}((U - \mathbb{E}U) \otimes (U - \mathbb{E}U))$ is the covariance operator and $\mathbb{E}U = 0$.

## Assumption 2

Let $x$ be a measurable map on $(\Omega, \mathcal{A}, P)$ with values in an Hilbert space $L^2_\pi$ such that $\mathbb{E}(||x||^2_{L^2_\pi}) < \infty$. Moreover, $x$ is a Gaussian Hilbert random variable with mean function $x_0$, *i.e.* $\forall \varphi \in L^2_\pi$, its characteristic function $\phi(\varphi)$ has the form

$$\phi(\varphi) = exp\{i < \varphi, x_0 >_{L^2_\pi} - \frac{1}{2} < \Omega_0\varphi, \varphi >_{L^2_\pi}\}, \tag{4}$$

where $\Omega_0 := \mathbb{E}((x - x_0) \otimes (x - x_0))$ is the covariance operator.

## Assumption 3

$\Sigma$ and $\Omega_0$ are injective operators.

# Our Solution of Inverse Problems (3)

- Since $\hat{Y}$ comes from the sample, it is linked to the sample size $n$. Thus, we add an index $n$ to the covariance operator of the noise in such a way that $\Sigma_n \to 0$.

Assumptions 1 & 2 imply that:

- $\Sigma$ and $\Omega_0$ are *bounded*, *self-adjoint*, *positive semi-definite*, *Hilbert-Schmidt* and *nuclear* operators
  ( *i.e.* $\exists$ a countable family $\{\lambda_j, \psi_j\}_j$ such that $\Sigma\psi_j = \lambda_j \psi_j$ and $||\Sigma||_{HS} := \sum_j \lambda_j^2 < \infty$ to $\Sigma$ be Hilbert-Schmidt and $\sum_j \lambda_j < \infty$ to $\Sigma$ be nuclear; a specular definition holds for $\Omega_0$).

- the *joint probability measure* $\Pi$ on $L^2_\pi \times L^2_\rho$ is a Gaussian measure and it is obtained by recomposing $\mu$ and $P^x$. It follows that $(x, \hat{Y})$ *is a joint Gaussian Process* and $\hat{Y}$ *has a Gaussian marginal distribution*( $\Rightarrow$ see below).

Notation: $\mu$ - *Prior probability*, $P^x = P(\hat{Y}|x)$ - *Sampling probability*,
$\mu^{\hat{Y}} = \mu(x|\hat{Y})$ (or $\mu^{\mathcal{F}}$) - *Posterior probability*
$\mathcal{E}$ - $\sigma-$*field over* $L^2_\pi$, $\mathcal{F}$ - $\sigma-$*field over* $L^2_\rho$.

# Joint and Marginal distribution of Gaussian Processes

- **Joint distribution** of $(x, \hat{Y})$ ($(x, \hat{Y}) \sim \mathcal{N}$ if and only if $< (x, \hat{Y}), (f, g) > \sim \mathcal{N}$, $\forall\, (f, g) \in L^2_\pi \times L^2_\rho$):

$$\begin{aligned} < (x, \hat{Y}), (f, g) >_{L^2_\pi \times L^2_\rho} &:= < x, f >_{L^2_\pi} + < \hat{Y}, g >_{L^2_\rho} \\ &= < x, f >_{L^2_\pi} + < Kx, g >_{L^2_\rho} + < U, g >_{L^2_\rho} \\ &= \mathcal{N}(< (x_0, Kx_0), (f, g) >, < V(f, g), (f, g) >), \end{aligned}$$

where $V$ is a covariance operator and can be decomposed in 4 operators:

$$\begin{aligned} V_{11} : L^2_\pi &\to L^2_\pi, \\ f &\to \Omega_0 f \end{aligned} \qquad \begin{aligned} V_{22} : L^2_\rho &\to L^2_\rho, \\ g &\to (\Sigma_n + K\Omega_0 K^*) g \end{aligned}$$

$$\begin{aligned} V_{12} : L^2_\pi &\to L^2_\rho, \\ f &\to K\Omega_0 f \end{aligned} \qquad V_{12} = V_{21}^* : L^2_\rho \to L^2_\pi$$

- **Marginal distribution** of $\hat{Y}$ ($\hat{Y} \sim \mathcal{N}$ if and only if $< \hat{Y}, g >_{L^2_\rho} \sim \mathcal{N}$, $\forall g \in L^2_\rho$):

$$< \hat{Y}, g >_{L^2_\rho} = < (0, \hat{Y}), (0, g) >_{L^2_\rho \times L^2_\rho} = \mathcal{N}(< Kx_0, g >, < (\Sigma_n + K\Omega_0 K^*) g, g >).$$

# Posterior Distribution: open question

- Does it exist a Gaussian distribution of $x|\hat{Y}$ (*i.e.* $x|\hat{Y} \sim \mathcal{N}(A\hat{Y} + b, V)$) such that, if recomposed with the marginal of $\hat{Y}$, gives the joint distribution of $(x, \hat{Y})$ (previously defined)?
- Making inference directly in functional spaces causes a difficulty in applying Bayes theorem and in defining the Posterior distribution of $x$.

- Identification of the operator $A : L^2_\rho \to L^2_\pi$: $\forall \varphi \in L^2_\pi$, $\psi \in L^2_\rho$

$$\begin{aligned} Cov(<x,\varphi>,<\hat{Y},\psi>) &= Cov(<\mathbb{E}(x|\hat{Y}),\varphi>,<\hat{Y},\psi>) \\ &= Cov(<A\hat{Y},\varphi>,<\hat{Y},\psi>) \\ &= Cov(<\hat{Y},A^*\varphi>,<\hat{Y},\psi>) \\ &= <(\Sigma_n + K\Omega_0 K^*)A^*\varphi,\psi>, \end{aligned}$$

and $Cov(<x,\varphi>,<\hat{Y},\psi>) = <V_{12}\varphi,\psi>$.

$\Rightarrow \quad A(\Sigma_n + K\Omega_0 K^*)\psi = \Omega_0 K^*\psi \quad \Rightarrow$ ill-posed problem (despite the prior)

( since $(\Sigma_n + K\Omega_0 K^*)$ is a compact operator).

- Does it means:
    - a *regular version of the conditional probability* $\mu^{\mathcal{F}}$ on $\mathcal{E}$ given $\mathcal{F}$ $\nexists$,

  or
    - the posterior mean $\mathbb{E}(x|\hat{Y})$ is not convergent?

  However, even in the case in which $\mu^{\mathcal{F}}$ exists, it could be not computable.

# Regularized Posterior Distribution

- Therefore, we ***define*** *the Posterior distribution as a Gaussian Process* and we ***guess*** *it is the solution of Statistical Inverse Problem (2):* $\hat{Y} = Kx + U$.

- Due to the infinite dimension of the problem, we are only able to compute a ***regularized version*** of it, denoted by $\mu_\alpha^{\mathcal{F}}$.

- Moreover, we prove (frequentist) consistency of this Regularized Posterior Distribution.

- The **Regularized Posterior Distribution** is defined through the Regularized operator $A$ (*Tikhonov regularization*):

$$A_\alpha = \Omega_0 K^* (\alpha I + \Sigma_n + K \Omega_0 K^*)^{-1}.$$

Then,

$$x|\hat{Y} \sim \mathcal{N}(A_\alpha \hat{Y} + b_\alpha, V_\alpha).$$

- $b_\alpha$ is identified by

$$\begin{aligned} \mathbb{E}(x) &= \mathbb{E}(\mathbb{E}(x|\hat{Y})) \\ x_0 &= A\mathbb{E}(\hat{Y}) + b \\ &= AKx_0 + b. \\ \Rightarrow \quad b_\alpha &= x_0 - A_\alpha K x_0. \end{aligned}$$

- Regularized posterior mean:

$$\begin{aligned} \mathbb{E}_\alpha(x|\hat{Y}) = & \overbrace{\Omega_0 K^*(\alpha I + \Sigma_n + K\Omega_0 K^*)^{-1}}^{A_\alpha} \overbrace{(Kx + U)}^{\hat{Y}} \\ & + \underbrace{(I - \Omega_0 K^*(\alpha I + \Sigma_n + K\Omega_0 K^*)^{-1} K)x_0}_{b_\alpha}. \end{aligned}$$

- Under homoskedasticity of $V$,

$$\begin{aligned} Var(x) &= \mathbb{E}(Var(x|\hat{Y})) + Var(\mathbb{E}(x|\hat{Y})) \\ \Omega_0 &= V + A(\Sigma_n + K\Omega_0 K^*)A^* \\ &= V + AK\Omega_0, \end{aligned}$$

- Regularized Variance:

$$V_\alpha = \Omega_0 - \underbrace{\Omega_0 K^*(\alpha I + \Sigma_n + K\Omega_0 K^*)^{-1}}_{A_\alpha} K\Omega_0.$$

# Asymptotic Analysis

We use a **frequentist** notion of **consistency**:
*we think the data as generated from a distribution characterized by the true value of the parameter and we check the accumulation of the posterior distribution in a neighborhood of this true value.*
⇒ we analyze convergence $P^x$-*a.s.* or in $P^x$-probability.

*Definition*: Posterior Consistency (Diaconis & Freedman (1986) [2])

- The pair $(x, \mu^{\mathcal{F}})$ is consistent if $\mu^{\mathcal{F}}$ converges weakly to $\delta_x$ as $n \to \infty$ under $P^x$-probability, or $P^x$-*a.s.*, where $\delta_x$ is the Dirac measure in $x$.
- The posterior probability $\mu^{\mathcal{F}}$ is consistent if $(x, \mu^{\mathcal{F}})$ is consistent for all $x$.

It is more difficult to obtain **consistency of the prior distribution**:
*we say that $(x, \mu)$ is consistent if and only if $\mu$ assigns a positive probability to every open interval around the true value $x$.*

# Convergence of Regularized Posterior (1)

Let $\mathcal{H}(\Omega_0)$ be the *Reproducing Kernel Hilbert Space* ($\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$) associated to $\Omega_0$, *i.e.*

$$\mathcal{H}(\Omega_0) = \{x : x \in L^2_\pi \quad \text{and} \quad ||x||^2_{\Omega_0} = \sum_{j=1}^{\infty} \frac{| < x, \varphi_j^{\Omega_0} > |^2}{\lambda_j^{\Omega_0}} < \infty\},$$

where $(\lambda_j^{\Omega_0}, \varphi_j^{\Omega_0})$ is the spectral decomposition of $\Omega_0$ and $|| \cdot ||_{\Omega_0}$ denotes the $\mathcal{H}(\Omega_0)$-norm.

**Remark:** $\mathcal{H}(\Omega_0)$ is the $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$ embedded in $L^2_\pi$.

(See Aronszajn (1950) [1] and Wahba (1990)[10] for a general theory on $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.$).

*Theorem 1* (Posterior Mean convergence)

If $tr\Sigma_n \sim \mathcal{O}_p(\frac{1}{n})$, $||\Sigma_n|| \sim \mathcal{O}_p(\frac{1}{n})$, $\alpha_n^2 n \to \infty$, $\alpha_n \to 0$ and $(x - x_0) \in \mathcal{H}(\Omega_0)$ then

$$\mathbb{E}_\alpha(x|\hat{Y}) \to^{\mathcal{P}^x} x$$

in $L^2_\pi$ sense.

Moreover, if $\Omega_0^{-\frac{1}{2}}(x - x_0) \in \mathcal{R}(\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ we have

$$||\mathbb{E}_\alpha(x|\hat{Y}) - x||^2 = \mathcal{O}_p(\alpha_n^\beta) + \mathcal{O}_p(\frac{1}{\alpha_n n}).$$

# Convergence of Regularized Posterior (2)

*Theorem 2* (Posterior Variance convergence)

If $tr\Sigma_n \sim \mathcal{O}_p(\frac{1}{n})$, $\alpha_n^2 n \to \infty$ and $\alpha_n \to 0$ then

$$||V_\alpha \varphi||^2 \to^{\mathcal{P}^x} 0.$$

Moreover, if $\Omega_0^{\frac{1}{2}}\varphi \in \mathcal{R}(\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$,

$$||V_\alpha \varphi||^2 = \mathcal{O}_p(\alpha_n^\beta + \frac{1}{\alpha_n^3 n^2}).$$

# Convergence of Regularized Posterior (3)

## The fastest global rate of convergence:

it is obtained when $\alpha_n^\beta = \frac{1}{\alpha_n n}$, that is when

$$\alpha_n \propto n^{-\frac{1}{\beta+1}};$$

$\hookrightarrow$ the speed of convergence of the regularized posterior mean is equal to $n^{-\frac{\beta}{\beta+1}}$.

## Convergence of the Posterior Distribution:

$\hookrightarrow$ for Gaussian posterior measure, posterior consistency is verified *iff* the posterior mean is consistent and the posterior variance converges to 0:

let $\tilde{x}$ be the true value of the parameter characterizing the DGP of $\hat{Y}$, by using *Chebyshev's Inequality* and for any sequence $M_n \to \infty$

$$\mu_\alpha^{\mathcal{F}}\{x : ||x - \tilde{x}||_{L^2_\pi} \geq M_n \varepsilon_n\} \quad \leq \quad \frac{||V_\alpha(x(t)|\hat{Y})||^2_{L^2_\pi} + ||\mathbb{E}_\alpha(x(t)|\hat{Y}) - \tilde{x}(t)||^2_{L^2_\pi}}{(M_n \varepsilon_n)^2}. \qquad (5)$$

It follows that, if $(\tilde{x} - x_0) \in \mathcal{H}(\Omega_0)$ the regularized posterior probability of the complement of any neighborhood of the true parameter $\tilde{x}$, $\mu_\alpha^{\mathcal{F}}\{x : ||x - \tilde{x}||_{L^2_\pi} \geq M_n \varepsilon_n\}$ goes to zero.

# A Result of Prior Inconsistency

- we have proved that to have posterior mean consistency the true value of the centered parameter of interest must satisfy the regularity condition: $(x - x_0) \in \mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.(\Omega_0)$.
- the support of the prior distribution is $\overline{\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.(\Omega_0)}$ (see Van der Vaart & Van Zanten [9]), but with $\mu-$ probability 1 the trajectories generated by this distribution are not in the $\mathcal{R}.\mathcal{K}.\mathcal{H}.\mathcal{S}.(\Omega_0)$.

  (**Remark:** the support is meant as the smallest closed set having probability one under the induced measure).
- it follows that **the prior distribution is not able to generate a trajectory $x$ that satisfies the necessary regularity condition**, as the following Lemma shows:

*Lemma*

Let $(x - x_0)$ a zero-mean Gaussian process with covariance operator $\Omega_0$.Then,

$$(x - x_0) \notin \mathcal{H}(\Omega_0) \ \mu\text{-a.s.}$$

$\Rightarrow$ Prior Inconsistency.

# Consistency of Bayes Estimates

Our result agrees with previous results in Bayesian non parametric estimation:

- **Doob's theorem** says that $(x, \mu^{\mathcal{F}})$ is consistent for $\mu$-almost all $x$, but it does not say anything on consistency at a particular true distribution of interest (this is the relevant case for the frequentist consistency) (see Ghoshal et alii (1998) [6]).
- Actually, nobody can be so sure about the prior distribution and **the true value of $x$ can have $\mu$-probability 0** (in particular when $x$ is infinite-dimensional).

$\Rightarrow$ there are a lot of **prior that are troublesome** (as for our case), however there are many reasonable prior for which prior consistency is verified (see Freedman (1965) [5]), namely, prior that are able to generate the true value of the parameter.

$\Rightarrow$ in our model the chosen prior is not good since it is not able to generate the true $x$, however the proposed posterior distribution is consistent for the true value of the parameter.

# The Case of Different Operators for each observation

Suppose to observe an $n$-sample of functions $\hat{Y}_1, ..., \hat{Y}_n$ such that

$$\hat{Y}_i = K_i x + U_i \qquad i = 1, \ldots, n \qquad U_i \sim iid\, \mathcal{GP}(0, \Sigma_n), \tag{6}$$

where

- each observation is a noisy transformation of the parameter of interest $x$ through an observation specific operator $K_i$;
- $K_i$, $i = 1, \ldots, n$ are known, non-random, H-S and injective operators;
- $x \sim \mathcal{GP}(x_0, \Omega_0)$;
- Assumptions 1, 2 & 3 are still valid;
- this is the classical Linear Regression Model with fixed regressors, where the operators play the role of explanatory variables;

**Notation:** $\hat{Y} = \overbrace{(\hat{Y}_1, \ldots, \hat{Y}_n)'}^{n \times 1}$, $U = \overbrace{(U_1, \ldots, U_n)'}^{n \times 1}$,

$K = \underbrace{(K_1, \ldots, K_n)'}_{n \times 1} : L^2_\pi \to (L^2_\rho)^n$, $K^* = \underbrace{(K_1^*, \ldots, K_n^*)}_{1 \times n} : (L^2_\rho)^n \to L^2_\pi$,

$((L^2_\rho)^n, \mathcal{F}^n)$ - product of $n$ measurable sample spaces,

$(P^\varepsilon)^n$ - associated sampling probability

$\Pi^n = \mu \otimes (P^\varepsilon)^n$ - joint probability measure

# Conditional and Marginal distribution of $\hat{Y}$

- $\hat{Y}$ is an $n$-dimensional Gaussian Process on $\mathbb{R}^n$.
- From the distributional assumption about $U_i$, the **conditional distribution** of $\hat{Y}$ is:

$$\hat{Y}_i | x \sim i\,\mathcal{GP}(K_i x, \Sigma_n) \qquad i = 1, \ldots, n.$$

- **Marginal distribution** of $\hat{Y}$:

$$\hat{Y} \sim \mathcal{GP}(K x_0, \Sigma_n I_n + K \Omega_0 K^*),$$

where $I_n$ is the identity matrix. We can re-write the **marginal covariance function** in matrix terms:

$$\mathrm{Var}(\hat{Y}) = \begin{pmatrix} \Sigma_n + K_1 \Omega_0 K_1^* & K_1 \Omega_0 K_2^* & \ldots & K_1 \Omega_0 K_n^* \\ K_2 \Omega_0 K_1^* & \Sigma_n + K_2 \Omega_0 K_2^* & & \\ \vdots & & \ddots & \vdots \\ K_n \Omega_0 K_1^* & \ldots & & \Sigma_n + K_n \Omega_0 K_n^* \end{pmatrix}.$$

# Marginalization through a Sufficient Statistic (SS)

Let $\mathcal{T} \subset \mathcal{F}^n$ be the sub-$\sigma$-field generated by a *sufficient statistic* $t$

$\hookrightarrow$ we consider the *marginalization on the sample space*, *i.e.* the restriction of probability $\Pi^n$ on $\mathcal{E} \otimes \mathcal{T}$, $\Pi^n_{\mathcal{E}\otimes\mathcal{T}}$, defined as $\Pi^n_{\mathcal{E}\otimes\mathcal{T}}(A) = \Pi^n(A)$, $\forall A \in \mathcal{E} \otimes \mathcal{T}$.

*Lemma*

Let $t = \sum_{i=1}^n K_i^* \hat{Y}_i$ and $\mathcal{T}$ be the $\sigma$-field generated by $t$. The statistic $t$ is sufficient for $x$ in the Bayesian experiment $(L^2_\pi \times (L^2_\rho)^n, \mathcal{E} \otimes \mathcal{F}^n, \Pi^n)$, *i.e.* $\mathcal{F}^n \perp\!\!\!\perp \mathcal{E} | \mathcal{T}$.

Sketch of the proof:

1. Project the model on an orthonormal bases of $(L^2_\rho)^n$ and consider a finite number $k$ of projections;
2. prove that $K^*\hat{Y}(k) := \sum_i K_i^* \hat{Y}_i(k)$ is a SS for the Bayesian projected experiment;
3. define $\mathcal{T}_k$ the sub-$\sigma$-field generated by $K^*\hat{Y}(k)$ and $\mathcal{T}_T = \sigma(K^*\hat{Y})$ the *tail $\sigma$-field*;
4. prove that, if $\mathcal{T}_k$ is sufficient for the projected model, the $\mathcal{T}_T$ is sufficient for the original model in infinite dimension.

■

$\Rightarrow \frac{1}{n}K^*\hat{Y}$ is a **sufficient statistic** for the infinite dimensional parameter $x$ and we make inference by only taking into account the information contained in $\mathcal{T} = \sigma(\frac{1}{n}K^*\hat{Y})$.

- The sampling probability restricted to $\mathcal{T}$ is a Gaussian Measure:

$$\frac{1}{n}K^*\hat{Y}|x \sim \mathcal{GP}(\frac{1}{n}K^*Kx, \frac{1}{n^2}K^*\Sigma K)$$

- The marginal distribution of $t$ is:

$$\frac{1}{n}K^*\hat{Y} \sim \mathcal{GP}(\frac{1}{n}K^*Kx_0, \frac{1}{n^2}(K^*\Sigma K + K^*K\Omega_0 K^*K))$$

- We define, as for the general case, the solution to ill-posed problem (6) to be the **regularized Gaussian Process** with **mean function**

$$\begin{aligned}\mathbb{E}_\alpha(x|K^*Y) &= \Omega_0\frac{1}{n}K^*K(\alpha_n I + \frac{1}{n^2}K^*\Sigma K + \frac{1}{n^2}K^*K\Omega_0 K^*K)^{-1} \\ &\quad \frac{1}{n}K^*[K(x-x_0)+U]+x_0\end{aligned}$$

and **Covariance operator**

$$W_\alpha = \Omega_0 - \Omega_0\frac{1}{n}K^*K(\alpha_n I + \frac{1}{n^2}K^*\Sigma K + \frac{1}{n^2}K^*K\Omega_0 K^*K)^{-1}\frac{1}{n}K^*K\Omega_0.$$

# Asymptotic Analysis for the different operators case

*Theorem 1bis* (Posterior Mean convergence for the different operator case)

If $\frac{tr(K^*\Sigma K)}{n^2} \sim \mathcal{O}_p(\frac{1}{n})$, $||\frac{K^*\Sigma K}{n^2}|| \sim \mathcal{O}_p(\frac{1}{n})$, $\frac{1}{n^2}K^*(\Sigma_n I_n + K\Omega_0 K^*)K \to Q$ with $Q$ a bounded operator, $\alpha_n^2 n \to \infty$, $\alpha_n \to 0$ and $(x - x_0) \in \mathcal{R.K.H.S.}(\Omega_0)$ then

$$\mathbb{E}_\alpha(x|\frac{1}{n}K^*\hat{Y}) \to^{(\mathcal{P}^x)^n} x$$

in $L^2_\pi$ sense. Moreover, if $\Omega_0^{-\frac{1}{2}}(x - x_0) \in \mathcal{R}(\frac{1}{n^2}\Omega_0^{\frac{1}{2}}K^*KK^*K\Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ we have

$$||\mathbb{E}_\alpha(x|\frac{1}{n}K^*\hat{Y}) - x||^2 = \mathcal{O}_p(\alpha_n^\beta) + \mathcal{O}_p(\frac{1}{\alpha_n n}) + \mathcal{O}_p(\frac{1}{\alpha_n^3 n^2}).$$

*Theorem 2bis* (Posterior Variance convergence for the different operator case)

If $||\frac{K^*\Sigma K}{n^2}|| \sim \mathcal{O}_p(\frac{1}{n})$, $\frac{1}{n^2}K^*(\Sigma_n I_n + K\Omega_0 K^*)K \to Q$ with $Q$ a bounded operator, $\alpha_n^2 n \to \infty$, $\alpha_n \to 0$ and if $\Omega_0^{\frac{1}{2}}\delta \in \mathcal{R}(\frac{1}{n^2}\Omega_0^{\frac{1}{2}}K^*KK^*K\Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ for any $\delta \in L^2_\pi$, then

$$||W_\alpha||^2 = \mathcal{O}_p(\frac{1}{\alpha_n^3 n^2} + \alpha_n^\beta).$$

$\Rightarrow$ *posterior consistency* but *prior inconsistency*

# Numerical Implementation of the general case

1. The measures $\pi$ and $\rho$ characterizing the two spaces $L^2_\pi$ and $L^2_\rho$ have been chosen equal to the uniform measure on $[0,1]$;
2. the DGP is:

$$\begin{aligned}\hat{Y} &= \int_0^1 x(s)(s \wedge t)ds + U, \qquad U \sim \mathcal{GP}(0, \Sigma_n) \\ \Sigma_n &= n^{-1} \int_0^1 \exp\{-(s-t)^2\} ds\end{aligned}$$

3. the true value of the parameter of interest is: $x(s) = -3s^2 + 3s$;
4. prior distribution:

$$\begin{aligned}x &\sim \mathcal{GP}(x_0, \Omega_0) \\ x_0 &= -2.8s^2 + 2.8s \\ \Omega_0\varphi &= 2\int_0^1 ((s \wedge t) - st)\varphi ds;\end{aligned}$$

5. $\alpha = 0.002$, $n = 1000$.
6. Monte Carlo experiment: $n = 100$, 50 Monte Carlo replications

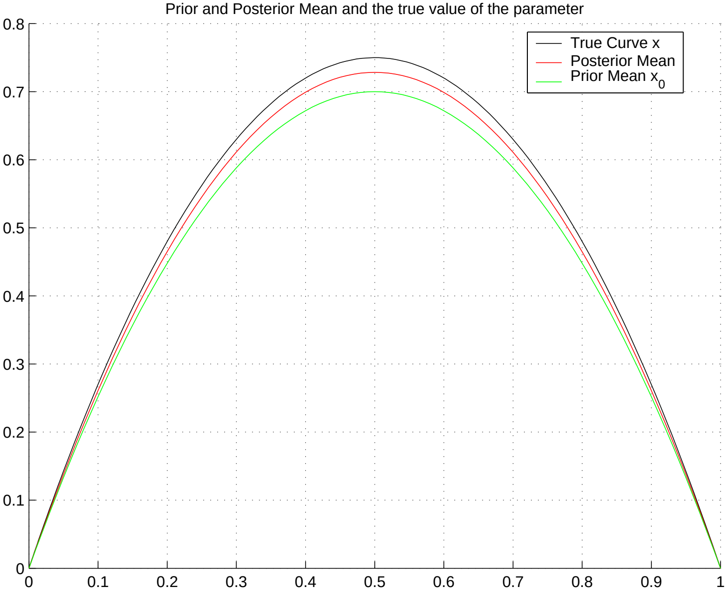


Figure: The true curve $x$ and the Regularized posterior mean

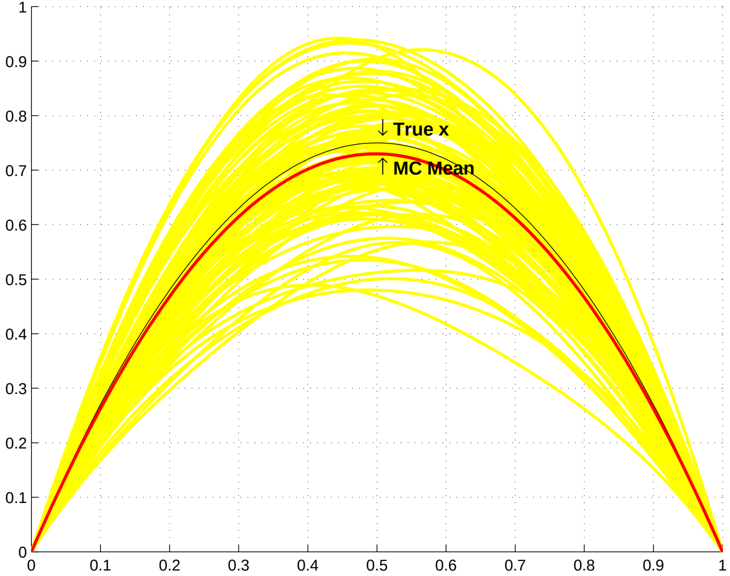


Figure: Monte Carlo simulation

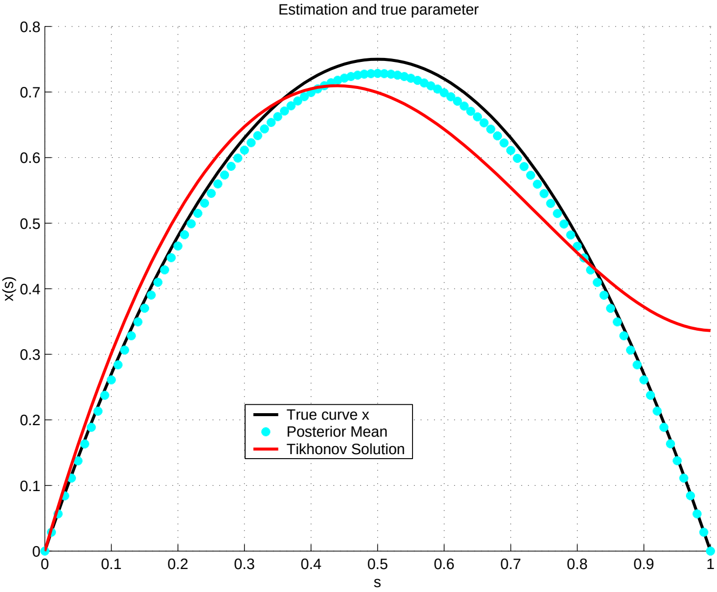


Figure: Regularized posterior mean and Tikhonov solution

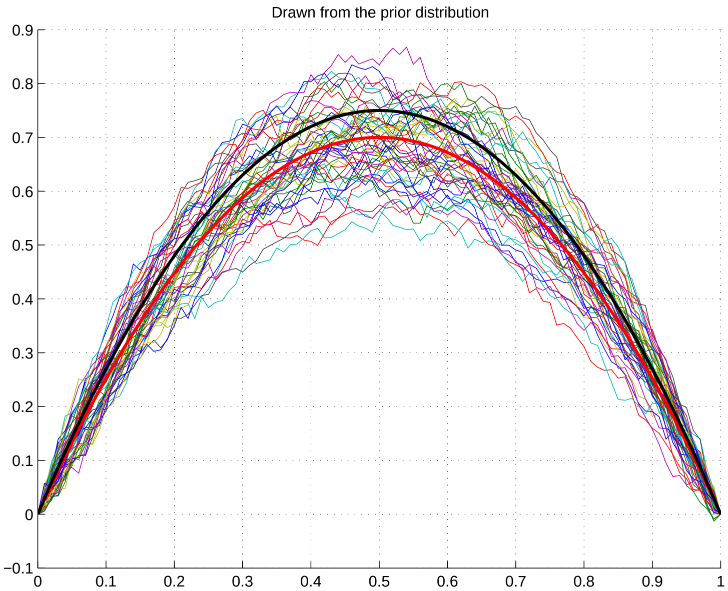


Figure: Drawn from the prior distribution with mean $x_0 = -2.8s^2 + 2.8s$ (in red) and variance $\Omega_0\varphi = 2\int_0^1((s \wedge t) - st)\varphi ds$

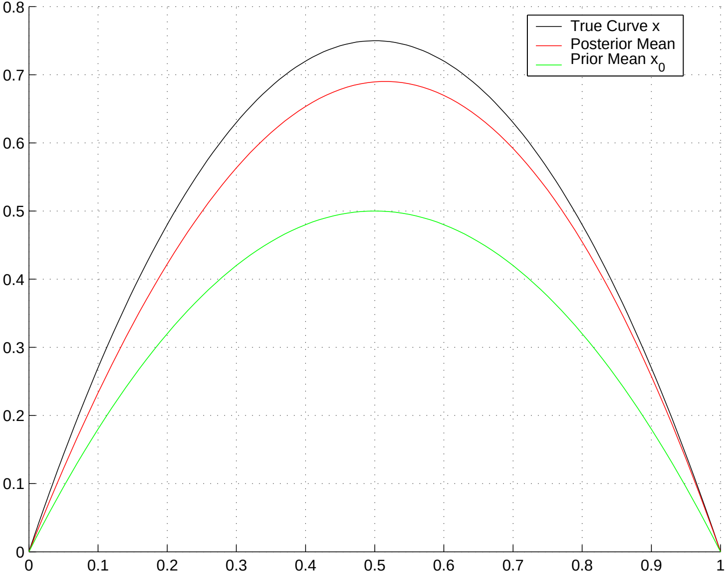


Figure: The true curve $x$ and the Regularized posterior mean for a prior mean $x_0 = -2s^2 + 2s$ and a prior variance $\Omega_0\varphi = 40\int_0^1((s \wedge t) - st)\varphi ds$

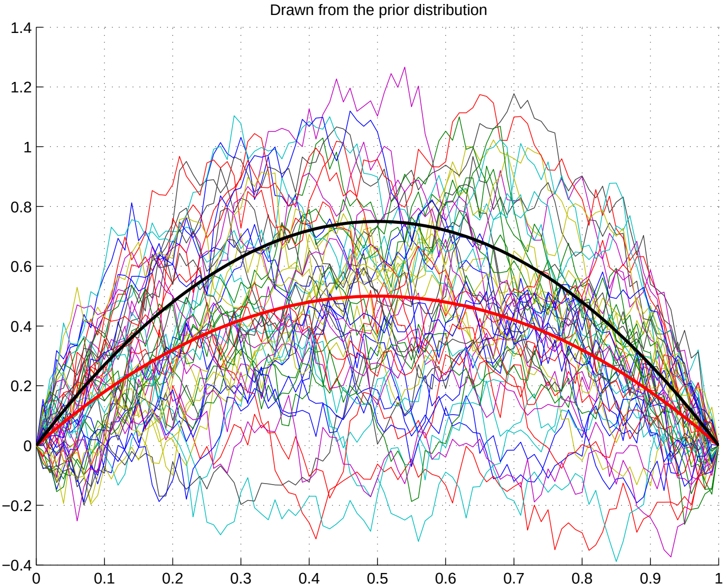


Figure: Drawn from the prior distribution with mean $x_0 = -2s^2 + 2s$ (in red) and variance $\Omega_0\varphi = 40\int_0^1((s \wedge t) - st)\varphi ds$

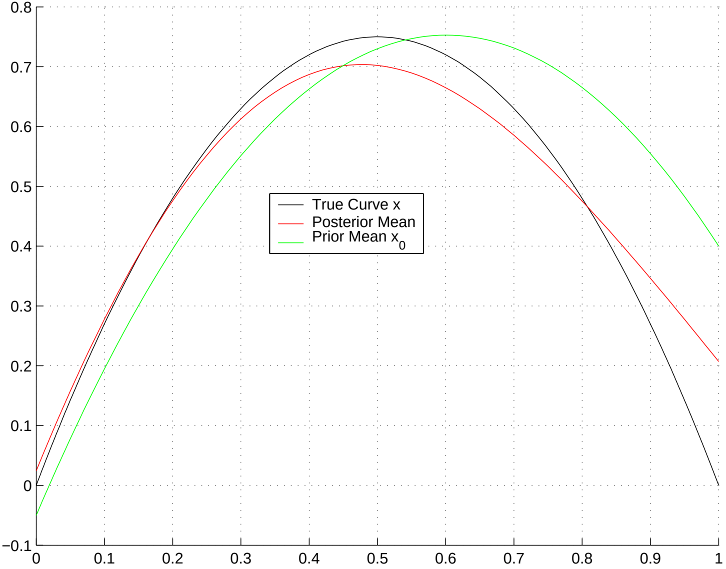


Figure: The true curve $x$ and the Regularized posterior mean for a prior mean $x_0 = -2.22s^2 + 2.67s - 0.05$ and a prior variance $\Omega_0\varphi = 600\int_0^1 (0.9(s-t)^2 - 1.9|s-t| + 1)\varphi ds$

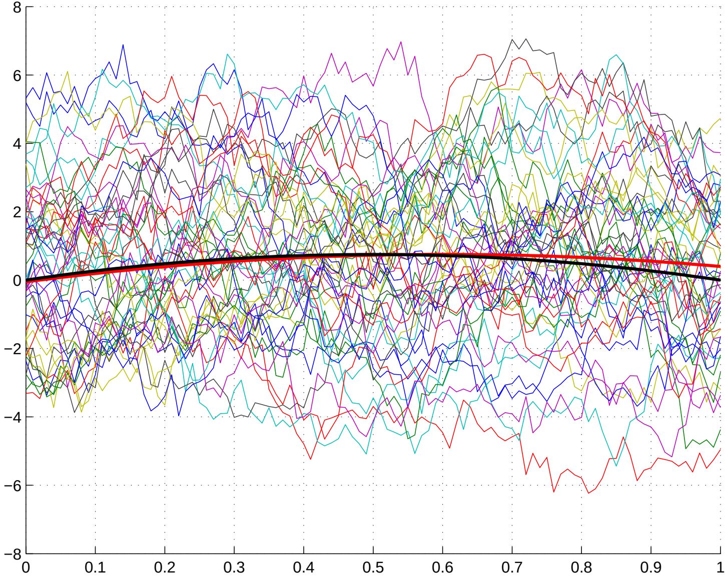


Figure: Drawn from the prior distribution with mean $x_0 = -2.22s^2 + 2.67s - 0.05$ (in red) and variance $\Omega_0\varphi = 600\int_0^1(0.9(s-t)^2 - 1.9|s-t| + 1)\varphi ds$

# Extensions and Conclusions

- Open questions: existence of a regular version of the conditional probability on the parameter space given the information in the sample (applicability of *Jirina Theorem*) when the spaces are infinite dimensional?

- Relation between *Bayes theorem* and *Ill posedness*? Applicability of *Bayes theorem* in functional spaces?

- Extension to the unknown operator $K$ case and study of the asymptotic properties of the solution $\mathbb{E}_\alpha(x|\hat{Y})$ with $K$ estimated (already done in Florens & Simoni (2007) [4]).

- Development of a data driven method to optimally select the parameter of regularization $\alpha$:
  - Rule 1: $\alpha = \arg\min \frac{1}{\alpha}||\hat{\varepsilon}^{\alpha}||^2$, where $\hat{\varepsilon}^{\alpha} = \hat{Y} - K\mathbb{E}_\alpha(x|\hat{Y})$.
  - Rule 2 (L-curve): graph $(||\hat{\varepsilon}^{\alpha}||, ||\mathbb{E}_\alpha(x|\hat{Y})||^2)$ for different $\alpha$ in log-log scale and select the corner.

# References


- Aronszajn, N. (1950), *Theory of Reproducing Kernels*, Trans.Amer.Math.Soc., **68(3)**, 337-404.
- Diaconis, F., and D., Freedman (1986), *On the Consistency of Bayes Estimates*, Annals of Statistics, **14**, 1-26.
- Engl, H. W., Hanke, M. and A., Neubauer (1996), *Regularization of Inverse Problems*, Kluwer Academic Publisher, Dordrecht.
- Florens, J.P. and A., Simoni (2007), *Bayesian Nonparametric Instrumental Regression*, working paper.
- Freedman, D. (1965), *On the Asymptotic Behavior of Bayes Estimates in the Discrete Case II.*, Ann. Math. Statist., **36**, 454-456.
- Ghosal, S., Ghosh, J.K., and A.,Van der Vaart (1998), *Convergence Rates of Posterior Distributions*, working paper.
- Kaipio, J. and E., Somersalo (2005), *Statistical and Computational Inverse Problems*, **160**, Springer-Verlag, New-York.
- Kress, R. (1999), *Linear Integral Equation*, Springer.
- Van der Vaart, A.W., and J.H., Van Zanten (2000), *Rates of Contraction of Posterior Distributions Based on Gaussian Process Priors*, Working paper.
- Wahba, G. (1990), *Spline Models for Observational Data*, CBMS-NSF Regional Conference Series. SIAM Society for Industrial and Applied Mathematics.