A Probabilistic Approach to Inverse Problems

Michel Minoux, Riadh Zorgati

1 University P. & M. Curie, LIP6, 4 place Jussieu 75252 Paris Cedex 05 (France), michel.minoux@lip6.fr
2 EDF Lab Paris-Saclay, OSIRIS, 7 Bd Gaspard Monge 91120 Palaiseau (France), riadh.zorgati@edf.fr

Abstract

We introduce a new approach based on a probabilistic formulation for dealing with linear inverse problems. With this approach, solving a linear inverse problem is formulated as maximizing the probability to satisfy a set of two-sided (bilateral) random inequalities.

A key property needed for solving such problems concerns the concavity of the probability function. Various recent results concerning necessary and/or sufficient conditions for concavity are recalled, both for the case of single-sided (unilateral) random gaussian inequalities, and for the case of two-sided (bilateral) gaussian inequalities. It is shown how these results can be exploited in view of efficient resolution of linear inverse problems in the framework of the proposed probabilistic approach.

As a typical illustration, we discuss a series of computational results on random linear systems featuring some typical characteristics of inverse problems like ill-conditioning. The solutions obtained are compared in terms of quality with the outcome of several other classical methods such as those based on the generalized inverse, on regularization and on singular value decomposition, and the reported results suggest that the new approach appears to be competitive.

Keywords: Inverse Problems, Probability Maximization.

1 Introduction

In quite diverse application areas such as mechanics, medical imaging, astronomy, geophysics or non-destructive evaluation, many problems related to calibration, fitting or estimation of a large number of input parameters of a model from a small amount of output noisy data, can be cast as inverse problems. This type of problems frequently arises in physics when we want to estimate one or several parameter(s) not directly accessible to measurements. Such an estimation is then performed by data inversion involving a model, generally nonlinear and deterministic, called the direct problem. Let’s mention for example the estimation of parameters for pricing electricity, or the non-destructive evaluation of some components of nuclear power plants. Devising efficient solution approaches for such inverse problems is therefore of great practical interest.

An interesting class of inverse problems relates to the Fredholm equation of first kind :

$$\int \int \int_{\Omega} H\left(r, r'\right) x\left(r'\right) dr' = y\left(r\right)$$

(1)

where $r, r'$ are two vectors of $\mathbb{R}^3$, $H\left(r, r'\right)$ is the known kernel of the integral, $y\left(r\right)$ the measurements and $x$ the unknown object to reconstruct (a field, an image, a contrast). Because an imperfect model may generate impossible values of parameters while some results may not correspond to any plausible parameter because of measurement errors, an inverse problem doesn’t have an unique solution $x$ which is continuously dependent on the data for any measurement $y$ : a small variation of data may result in a huge variation of the parameters. Hence, due to noise data corruption, insufficient data and model errors, most inverse problems are ill-posed in Hadamard sense, i.e. existence, uniqueness and stability of the solution are not guaranteed [1].

In a discrete setting, such a model becomes a deterministic linear system of equations :

$$Hx = y$$

(2)
where $H \in \mathbb{R}^{m \times n}$ is the matrix of the physical system, $y \in \mathbb{R}^{m}$ is the vector of measurements and $x \in X$ is the unknown (in many applications, $X \subseteq \mathbb{R}^{n}$ can be assumed). For this model, given $H$ and $y$, finding $x$ by solving (2) is the inverse problem by contrast to the direct problem which consists in computing $y$ given $H$ and $x$. After discretization, the ill-posedness of operator $H$ results in a very high condition number of the associated matrix $H$, which can typically range in certain situations from $10^{9}$ to $10^{18}$. This condition number plays the role of an amplifier of perturbation of uncertainties on data during inversion and then, renders the inverse problem difficult to handle numerically.

A lot of work has been devoted to the development of a wide set of techniques for solving such inverse problems, from the classical least squares to Bayesian methods. Amongst the classical approaches, let’s first mention the generalized inverse approach introduced by Moore and Penrose, for finding a “solution” to the problem $Hx = y$ when $H$ is a singular matrix [2], [3]. The solution to (2) reads as $x = H^{+}b$ where $H^{+}$ is the generalized inverse of $H$ allowing to select a solution in the set of “least squares” solutions. In practice, this approach aims at solving the system of equation $Hx = y$ in a deterministic setting without any knowledge about the unknown $x$ using algebraic techniques. Satisfactory solutions are obtained with these methods only when the system is over-determined, noise is negligible and the matrix $H$ is well conditioned. Hence, if the formalism of generalized inverse solves the issues of existence and uniqueness, the stability issue remains critical. In fact, when a noise corrupts the data, the generalized inverse solution may be completely inappropriate. A deterministic model is not sufficient to determine acceptable solutions.

A more suitable approach for dealing with an inverse problem is based on a regularization process which consists in integrating an appropriate a priori information on the solution in view of obtaining a unique stable and acceptable solution compatible with the data and continuously depending on the data [4], [5]. Depending on how the available a priori information is modeled and integrated into the problem, inversion is performed within either a deterministic or a stochastic formalism. Among the many types of deterministic regularization methods, two most popular types are worth mentioning because of the ease in inversion is performed within either a deterministic or a stochastic formalism. The first type is based on a regularization by control of a dimension (leading for example to the Truncated Singular Value Decomposition). The second type is based on the Phillips-Twomey-Tikhonov (PTT) approach [6], [7] leading to an optimization problem of the type $\min_{x} J(x) = \| Hx - y \|^2_B + \Omega(x)$, $\lambda > 0$, where $\| . \|^2_B$ is a given norm chosen for measuring the residue $r = Hx - y$ on the solution and $\Omega(.)$ is a norm controlling the smoothness of the solution (generally chosen different from $B$). This popular approach aims at minimizing an optimality criterion that accounts for both the data misfit (measured via an appropriate norm) and the a priori information available on the solution, expressed mathematically through a functional $\Omega(x)$. The regularization parameter $\lambda$ expresses the trade-off between a measure of how closely the reconstructed solution matches the given data and the relative confidence we have in the a priori information. Even if formal criteria can be found in [4], [8], appropriately choosing the parameter $\lambda$ is often a source of difficulty.

We propose here a probabilistic approach for dealing with discrete linear inverse problems which, to the best of our knowledge, appears to be new. In this approach, it is assumed that uncertainty not only impacts the right-hand side vector $y$, but also the various entries of the $H$ matrix: thus, assuming knowledge of the probability distributions of $H$ and $y$, for any fixed $x \in X$, $Hx - y$ is a random vector $\in \mathbb{R}^{m}$ with known probability distribution. Now, for any given tolerance vector $\varepsilon$ with small positive components $\varepsilon_i$, $(i = 1, m)$, it is possible to compute for any $x \in X$ the probability $f(x) = \mathbb{P}[-\varepsilon < Hx - y < \varepsilon]$ which can be viewed as a likelihood measure for $x$ to be a relevant solution to the problem addressed. In this way, the inverse problem is reformulated in terms of maximizing $f$ over $X$.

In the framework of the present paper, multidimensional gaussian distributions will be assumed for the various rows of the $[H, -y]$ matrix, together with the stochastic independence of the corresponding random vectors; as a result, for any given $x$, each component of the $m$-vector $Hx = y$ is a gaussian random variable with known mean and variance, and the (joint) probability value $f(x)$ can be computed explicitly as the product of the $m$ (individual) probability values $\mathbb{P}[-\varepsilon_i < h_i^T x - y_i < \varepsilon_i]$. A key issue, in such an approach, is how to choose the tolerance vector $\varepsilon$. As will be confirmed by our computational experiments, the optimal solutions to the above model appear to be extremely stable (or ’robust’) when varying the $\varepsilon$ coefficients in a wide range of values, and a systematic way of choosing appropriate values in
this range (through the solution of an associated convex optimization problem) will be proposed. Finally, a series of preliminary computational experiments aimed at comparing the solutions to the inverse problem obtained with the new approach with those obtained with a number of standard approaches will be reported.

The paper is organized as follows. Section 2 presents our probabilistic approach to linear inverse problems. Section 3 reports preliminary numerical experiments. Some conclusions and perspectives are presented in Section 4.

2 Probabilistic Approach to Linear Inverse Problems

Let $H$ a $m \times n$ matrix and $y \in \mathbb{R}^n$ be given. A possible probabilistic formulation of the inverse problem: find $x \in X$ such that $Hx = y$, consists in considering $H$ and $y$ as realizations of a random matrix $H(\xi)$ and a random vector $y(\xi)$, and to find $x$ maximizing the probability that the various residues $(y(\xi) - H(\xi)x)_i$, \( i = 1, \ldots, m \) belong to the interval $[-\varepsilon_i, \varepsilon_i]$, for given small positive scalars $\varepsilon_i \in \mathbb{R}_+$. This leads to solve the following Probability Maximization Problem (PMP):

$$\max_{x \in X} \mathbb{P}[-\varepsilon \leq H(\xi)x - y(\xi) \leq \varepsilon]$$

where $X \subseteq \mathbb{R}^n$ is a closed convex set (we assume that $X \neq \emptyset$ and does not contain $0$).

Optimization problems of type (PMP) have been studied in [9] and a new necessary and sufficient condition for convexity of solutions sets defined by one-sided probabilistic constraints with Gaussian distributions has been derived. From this result, a sufficient condition of convexity for problems involving two-sided random gaussian inequalities was obtained. The case of $m$ independent bilateral probabilistic constraints, which is the case of interest for inverse problems, is investigated below and approximate solutions are obtained by considering the closely related max-min formulation, which can be efficiently solved by a sequence of Second-Order Conic Programs, a class of efficiently solved convex problems [10].

In §2.1 and §2.2, we recall some key ingredients on which our approach is based. In §2.3, using these ingredients, we derive the Max-Min optimization model used in our computational experiments in order to produce approximate optimal solutions to the probability maximization problem (3).

2.1 The Case of a Single Probabilistic Inequality

2.1.1 Necessary and Sufficient Condition for Concavity (1-Sided Case)

First, consider Probability Maximization Problems with single sided inequalities of the form:

$$\begin{align*}
\text{(PMP)} \quad & \max_{x \in X} f(x) = \mathbb{P}\{u^Tx \leq b\} = \Phi\left(\frac{b - u^T\mu}{\sqrt{\Sigma}}\right) \\
\text{s.t.} \quad & x \in X
\end{align*}$$

where:

- $f(x) = \mathbb{P}\{u^Tx \leq b\} = \Phi\left(\frac{b - u^T\mu}{\sqrt{\Sigma}}\right)$ for a given $x \in \mathbb{R}^n$ and given $b \in \mathbb{R}$, is the probability to satisfy the single-sided random inequality $u^Tx \leq b$, with $u \sim \mathcal{N}(\mu, \Sigma)$. We assume the covariance matrix $\Sigma$ symmetric positive-definite with eigenvalues $\lambda \in [\lambda_{\min}, \lambda_{\max}]$, $\lambda_{\min} > 0$.
- $\Phi$ denotes the c.d.f. for $\mathcal{N}(0,1)$ and $X$ is a polyhedron (in many practical applications, the condition $x \geq 0$ naturally arises, and thus $X$ can be assumed to be a subset of $\mathbb{R}_n^+$. This will be the case in our numerical experiments in Section 3).

In view of solving such problems, a key issue relates to the characterization of the domain(s) of (local) concavity of $f$. In [9], the following necessary and sufficient condition has been established:

**Theorem 2.1.** $\Phi\left(\frac{b - u^T\mu}{\sqrt{\Sigma}}\right)$ has negative semi-definite hessian in $x \neq 0$, such that $b - \mu^Tx > 0$,

if and only if:

$$ x^T\Sigma x \left(\mu^T\Sigma^{-1}\mu\right) \leq (b + \mu^Tx)^2 + b^2 \left(\frac{(b - \mu^Tx)^2}{x^T\Sigma x} - 3\right) \quad (4) $$

(for more details, see [9]).
2.2.2 Sufficient Condition for Concavity (2-Sided Case)

Theorem 2.1 above also turns out to be useful in analyzing the case of a 2-sided (bilateral) probabilistic constraint. For \( a < b \) and any given \( x \in \mathbb{R}^n, x \neq 0 \), the probability of satisfying \( a \leq x^T x \leq b \) where \( u \sim \mathcal{N}(\mu, \Sigma) \) is \( g(x) = \mathbb{P} \left[ a \leq u^T x \leq b \right] \) and can be expressed as:

\[
g(x) = \Phi \left( \frac{b - \mu^T x}{\sqrt{x^T \Sigma x}} \right) - \Phi \left( \frac{a - \mu^T x}{\sqrt{x^T \Sigma x}} \right) = \Phi \left( \frac{b - \mu^T x}{\sqrt{x^T \Sigma x}} \right) + \Phi \left( \frac{\mu^T x - a}{\sqrt{x^T \Sigma x}} \right) - 1
\]

**Corollary** [9]. We define \( \theta_a = \frac{(u^T x - a)^2}{(x^T \Sigma x)} \) and \( \theta_b = \frac{(b - u^T x)^2}{(x^T \Sigma x)} \). A sufficient condition for negative semi-definiteness of \( \nabla_x^2 g(x) \) in \( x \neq 0 \), \( a < u^T x < b \) is:

\[
(x^T \Sigma x) (\mu^T \Sigma^{-1} \mu) \leq \min \left\{ (b + \mu^T x)^2 + b^2 (\theta_b - 3); (a + \mu^T x)^2 + a^2 (\theta_a - 3) \right\}
\]

2.2 The Case of \( m \) Independent Bilateral Probabilistic Inequalities

This is the case of interest for probabilistic inversion.

2.2.1 Formulation

Typically, given \( a_i < b_i, (i = 1, \ldots, m) \), \( u_i \sim \mathcal{N}(\mu_i, \Sigma_i) \) with \( \Sigma_i > 0 \) and

\[
g_i(x) = \mathbb{P} \left[ a_i \leq u_i^T x \leq b_i \right] = \Phi \left( \frac{b_i - \mu_i^T x}{\sqrt{x^T \Sigma_i x}} \right) + \Phi \left( \frac{\mu_i^T x - a_i}{\sqrt{x^T \Sigma_i x}} \right) - 1
\]

We want to maximize the product \( \prod_{i=1}^{m} g_i(x) \) representing the joint probability of satisfaction of the \( m \) inequalities \( a_i \leq u_i^T x \leq b_i \), \( (i = 1, \ldots, m) \), or equivalently:

\[
(PMP') \quad \max_{x \in \mathbb{R}^n} \left\{ \sum_{i=1}^{m} \log (g_i(x)) \right\} \quad \text{s.t.} \quad x \in X \quad (5)
\]

This problem is nonconvex, hence belongs to a class of problems for which finding exact globally optimal solutions is computationally hard. In view of this, it is realized that if we want to handle large-size systems, the requirement for exact global optimality has to be relaxed, and we have to content ourselves with approximate solutions. We propose below a systematic way of generating approximate solutions to (PMP') which will be used in our computational experiments.

2.2.2 Approximating (PMP') : the Max-Min Formulation and the SOCP Approach

The Max-min formulation consists in replacing the product by the minimum, leading to the following problem:

\[
(PMP'') \quad \max_{x \in \mathbb{R}^n} \left\{ \min_{i=1,\ldots,m} \left\{ \frac{b_i - \mu_i^T x}{\sqrt{x^T \Sigma_i x}} \right\} ; \min_{i=1,\ldots,m} \left\{ \frac{\mu_i^T x - a_i}{\sqrt{x^T \Sigma_i x}} \right\} \right\} \quad (6)
\]

The efficient solvability of the above problem stems from the fact that it can be reduced to solving (for various values of the real parameter \( \rho > 0 \)) a sequence of problems of the form:

\[
(1) \quad \left\{ \begin{array}{l}
\max z \\
\text{s.t.} \\
(2) \quad \begin{array}{l}
b_i - \mu_i^T x \geq \rho \sqrt{x^T \Sigma_i x} + z \iff \sqrt{x^T \Sigma_i x} \leq \frac{1}{\rho} (b_i - \mu_i^T x - z), \quad (i = 1, \ldots, m) \\
\mu_i^T x - a_i \geq \rho \sqrt{x^T \Sigma_i x} + z \iff \sqrt{x^T \Sigma_i x} \leq \frac{1}{\rho} (\mu_i^T x - a_i - z), \quad (i = 1, \ldots, m) \\
x \in \mathbb{R}^n
\end{array}
\right.
\]

4
which are recognized as SOCP (Second-Order Cone Programming) problems, a class of very efficiently solvable convex optimization problems [10]. For any given \( \rho > 0 \), denoting \( z^*(\rho) \) the optimal solution value to (I), the optimal solution to (PMP) then corresponds to the maximum value \( \rho^* \) of \( \rho \) for which \( z^*(\rho) \geq 0 \). In view of this, the optimal value \( \rho^* \) can be determined by application of a simple dichotomic approach. In the sequel, the dichotomic approach will be implemented assuming that there exists \( x \in X \) such that \( a_i < \mu_i^2 x < b_i, \ i = 1, \ldots, m \). In that case, we know that \( \rho^* > 0 \). Therefore, the dichotomic approach can be carried out by iterative reduction of the initial interval of values \([0, \tilde{\rho}]\), where \( \tilde{\rho} = \min_{i=1, \ldots, m} \{ \frac{b_i - a_i}{2} \} / \min_{x \in X} \sqrt{x^T \Sigma_i x} \) (please note that computation of \( \min_{x \in X} \sqrt{x^T \Sigma \cdot x} \) can be efficiently done since this is again recognized as a SOCP problem). For each new value of \( \rho \) considered, \( \rho \) becomes the new lower bound of the confidence interval if \( z^*(\rho) \geq 0 \), and the new upper bound if \( z^*(\rho) < 0 \).

2.3 Application to Linear Inverse Problems

When uncertainties impact both matrix \( H \) and vector \( y \), problem (6) is adapted for dealing with problem (3) by considering the augmented matrix \( \hat{H} = [H, -y] \) and the associated augmented vector \( \hat{x} = \begin{bmatrix} x \\ x_{n+1} \end{bmatrix} \) with \( x_{n+1} = 1 \). In the sequel, \( \mu_i \) and \( \Sigma_i \) respectively represent the mean value and the variance-covariance matrix of \( H_i \), the \( i \)-th line of matrix \( H \). We note \( \mu_{y_i} \) and \( \sigma_{y_i} \) the mean value and the standard deviation of \( y_i \), respectively, for \( i = 1, \ldots, m \).

The solution of the inverse problem is finally looked for as the optimal solution \( \hat{x} \) to the problem of the form:

\[
(PMP''') \quad \max_{\hat{x} \in X} \left\{ \min_{i=1, \ldots, m} \left\{ \frac{\varepsilon_i - \mu_i^2 x + \mu_{y_i}}{\sqrt{\Sigma_i}} \right\} ; \min_{i=1, \ldots, m} \left\{ \frac{\mu_i^2 x - \mu_{y_i} + \varepsilon_i}{\sqrt{\Sigma_i}} \right\} \right\}
\]

The \( \varepsilon_i \) parameters \( i = 1, \ldots, m \) are small positive scalars, a possible way of choosing them will be explained into details in §3.1 below.

3 Numerical Experiments

We now turn to discussing computational experiments aimed at evaluating the capability of the probabilistic approach of solving hard, ill-conditioned, linear inverse problems. We first briefly describe the simulation framework considered in order to carry out these experiments.

3.1 The Inverse Problem Simulation Framework

The basic idea consists in starting with a known direct linear problem specified via a given \( m \times n \) matrix \( H \) and a given reference solution \( x_{ref} \in X \subseteq \mathbb{R}^n \); the corresponding right-hand side \( y \in \mathbb{R}^m \) is then deduced from the relation \( y = Hx_{ref} \) (the "direct system"). This linear system is supposed to constitute a perfect representation of the behavior of the physical system under consideration. In that way, \( x_{ref} \) may be viewed as an exact solution to the inverse problem: find \( x \in X \) satisfying \( Hx = y \).

In practice, for many reasons (e.g. because of errors in the computation of the \( h_{ij} \) coefficients, or because of measurement errors), \( H \) is only known approximately, as a perturbation \( \hat{H} \) of the exact matrix \( H \); similarly, the components of the right-hand side vector \( y \) are only known approximately (due to measurement errors) and only a perturbed vector \( \hat{y} \) is available. The inverse problem to be simulated is then to construct the best possible approximation of the \( x_{ref} \) vector, given \( \hat{H} \) and \( \hat{y} \). Since, in this context, the exact reference solution \( x_{ref} \) is known, the quality of a candidate solution \( \hat{x} \) can easily be evaluated using e.g. the (euclidian) norm of the difference \( \| x_{ref} - \hat{x} \| \).

Because linear inverse problems frequently feature ill-conditioning, the computation experiments below involve various matrices \( H \) chosen for their more or less severe ill-conditioning characteristics: the Wilson matrix [11], the Hilbert matrix [11], the Nagasaka matrix [12]. For any such matrix \( H \in \mathbb{R}^{m \times n} \), and for any given reference solution \( x_{ref} \in \mathbb{R}^n \), one or several inverse problem instances are generated by constructing \( \hat{H} \) from \( H \) using one of the following ways:

5
i) $\hat{H}$ is defined as $\hat{H} = H + \delta H$, where $\delta H$ is a given perturbation matrix featuring a few nonzero entries (in this case, the perturbed entries of $\delta H$ are specified);

ii) each entry $\tilde{h}_{ij}$ of $\hat{H}$ is drawn at random independently from the gaussian distribution $\mathcal{N}(h_{ij}, \tilde{\sigma}_{ij})$ where $\tilde{\sigma}_{ij} = \gamma \mid h_{ij}$ and $\gamma$ is a fixed real parameter chosen in $]0,1[$. This way of generating $\hat{H}$ from $H$ will be denoted $G(H, \gamma)$.

iii) each entry $\tilde{H}$ of $\hat{H}$ is drawn at random independently from the uniform distribution on the interval $[h_{ij} - \gamma \mid h_{ij}] ; h_{ij} + \gamma \mid h_{ij} [ ]$ where $\gamma$ is a fixed real parameter chosen in $]0,1[$. This way of generating $\hat{H}$ from $H$ will be denoted $U(H, \gamma)$.

In a similar way, the components $\tilde{y}_i$ of the perturbed right-hand side vector $\tilde{y}$ are drawn randomly and independently from either the normal distribution $\mathcal{N}(y_i, \gamma \mid y_i)$ (this is denoted $G(y, \gamma)$) or from the uniform distribution on the interval $[y_i - \gamma \mid y_i \mid ; y_i + \gamma \mid y_i || ]$ (this is denoted $U(y, \gamma)$).

Now, given $\hat{H}$ and $\tilde{y}$ resulting from the above instance generation procedure, the probabilistic approach for solving the inverse problem " find $\tilde{x}$ solving $H\tilde{x} = \tilde{y}$ " is implemented as follows. Each entry of $\hat{H}$ is considered as the mean value of a gaussian variable $u_{ij}$ with standard deviation $\sigma_{ij} = \alpha \mid h_{ij} |$ where $\alpha$ is a parameter chosen in $]0,1[$; and each component of the $\tilde{y}$ vector is considered to be the mean value of a random variable with standard deviation $\alpha \mid \tilde{h}_{ij} |$. It should be emphasized that, in this simulated reconstruction process, the exact matrix $H$ and the way $\hat{H}$ has been generated from $H$ are supposed to be unknown, therefore the method should work in a satisfactory way even in the case of a significant discrepancy between the $\sigma_{ij}$ and the $\tilde{\sigma}_{ij}$ values. This will be confirmed by the computational experiments reported below.

In our computational experiments, the $\varepsilon_i$ values ($i = 1,\ldots,m$) are deduced from the optimal solution to the following (SOCP) problem in variables $x$ and $\varepsilon_i$ ($i = 1,\ldots,m$):

$$\begin{array}{l}
\min_{x,\varepsilon} \sum_{i=1}^{m} \varepsilon_i^2 \\
\text{s.t.} \\
\sum_{i=1}^{m} h^+_{ij} x_j - y_i^- \leq \varepsilon_i (i = 1,\ldots,m) \\
\sum_{i=1}^{m} h^-_{ij} x_j - y_i^+ \geq -\varepsilon_i (i = 1,\ldots,m) \\
x \in X
\end{array} \quad (8)$$

where:

$$h^+_{ij} = \tilde{h}_{ij} + 2\alpha \mid \tilde{h}_{ij} |, \quad h^-_{ij} = \tilde{h}_{ij} - 2\alpha \mid \tilde{h}_{ij} |$$
$$y_i^+ = \tilde{y}_i + 2\alpha \mid \tilde{y}_i |, \quad y_i^- = \tilde{y}_i - 2\alpha \mid \tilde{y}_i |$$

The problem (7) is solved by carrying out 15 to 20 iterations of dichotomic search, each iteration requiring the solution of a (SOCP) problem of the form (I) (see Section 2.2.2).

An important point to mention regarding the probabilistic method, is the fact that the covariance matrix $\Sigma$ used for solving the inverse problem is different from the covariance matrix $\Sigma$ used for generating the random matrix $\hat{H} = \{h_{ij} + \sigma_{ij} \xi_{ij}\}$. For our simulations, covariance matrices $\tilde{\sigma}_{ij} = \gamma \mid h_{ij} |$ for generating $G(H, \gamma)$ and $U(H, \gamma)$ have been designed with and $\gamma = 0.01$, while the variance-covariance matrix $\Sigma$ for solving the inverse problem is designed by considering that each standard deviation is supposed to be a given percentage $\alpha$ of observed value $h_{ij} : \sigma_{ij} = \alpha \mid h_{ij} |$, with $\alpha = 0.15$ (i.e. we suppose that each coefficient is known with 15% of error). The mean matrix $M$ is taken as the observed matrix: $\mu_{ij} = h_{ij}$.

In the following subsections, we compare results obtained by our probabilistic method (noted PA) with some classical methods chosen amongst the different types: the Moore-Penrose Inverse (noted GIMP) and Kaczmarz (noted KACZ) as representatives of Inverse Generalized techniques, the Tikhonov (noted TIKO) as representative of PTT Regularization techniques and the Truncated Singular Value Decomposition (noted TSVD) as representative of techniques of regularization by control of a dimension.
Three test-cases have been chosen for illustrating the behavior of our approach with respect to moderate or severe ill-conditioning and to propagation of arithmetic errors.

For each matrix type tested, the results reported in the tables 1 to 3 will concern:

- one instance denoted $H + \delta H$ corresponding to a limited perturbation of the $H$ matrix, where $\delta H$ is fully specified by the list of its nonzero entries;
- 20 randomly generated instances obtained by applying the $G(H, \gamma)$ generation procedure explained above;
- 20 randomly generated instances obtained by applying the $U(H, \gamma)$ generation procedure explained above.

(In the last two cases, the figures displayed in the tables are the average values of the euclidian norm of $\| x_{ref} - \tilde{x} \|$ over the 20 instances).

3.2 Moderate Ill-Conditionning : The Wilson Matrix

The first test-case ([11]), based on the Wilson matrix, a non-singular ($\det(H) = 1$), $4 \times 4$ symmetric, positive-definite matrix with moderate condition number ($\text{cond}(H) = 2984$), is defined by:

$$H = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \quad x = \begin{pmatrix} 1 \\ . \\ . \\ 1 \end{pmatrix}, \quad y = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix}$$

The perturbed data are:

$$\tilde{H} = H + \delta H = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7.08 & 5.04 & 6 & 5 \\ 8 & 5.98 & 9.89 & 9 \\ 6.99 & 4.99 & 9 & 9.98 \end{pmatrix}, \quad \tilde{y} = \begin{pmatrix} 32.01 \\ 22.99 \\ 33.01 \\ 30.99 \end{pmatrix}$$

The Gaussian matrix and associated vector, for a value of $\gamma = 0.01$, are:

$$\tilde{H} = G(H, \gamma) = \begin{pmatrix} 10.05 & 7.12 & 7.81 & 7.06 \\ 7.02 & 4.93 & 5.97 & 5.01 \\ 8.28 & 6.16 & 9.86 & 9.27 \\ 7.05 & 4.99 & 9.06 & 9.97 \end{pmatrix}, \quad \tilde{y} = \begin{pmatrix} 31.96 \\ 23.34 \\ 33.46 \\ 31.43 \end{pmatrix}$$

The Uniform matrix and associated vector, for a value of $\gamma = 0.01$, are:

$$\tilde{H} = U(H, \gamma) = \begin{pmatrix} 10.03 & 6.93 & 8.05 & 7.06 \\ 7.02 & 5.02 & 6.02 & 4.98 \\ 8.02 & 5.96 & 10.04 & 8.91 \\ 6.96 & 4.95 & 8.92 & 10.06 \end{pmatrix}, \quad \tilde{y} = \begin{pmatrix} 32.12 \\ 22.91 \\ 33.29 \\ 30.71 \end{pmatrix}$$

Table 1 reports the obtained results for this test-case with the reference solution $x_{ref} = [1, 1, \ldots, 1]^T$.

When the matrix is partially perturbed, the probabilistic approach allows a reconstruction of $x_{ref}$ as good as Kaczmarz and regularized methods. Even when a few coefficients of the matrix are slightly perturbed, Moore-Penrose generalized inverse technique fails as expected. When all the coefficients of the matrix are subject to uncertainty, the probabilistic technique provides the best approximation of the reference solution vector $x_{ref}$. 
Wilson  |  GIMP  |  KACZ  |  TIKO  |  TSVD  |  PA (Prob) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$H + \delta H$</td>
<td>81.08</td>
<td>0.26</td>
<td>0.33</td>
<td>0.25</td>
<td>0.26 (0.9999)</td>
</tr>
<tr>
<td>$G(H, \gamma)$</td>
<td>81.08</td>
<td>0.27</td>
<td>0.33</td>
<td>0.25</td>
<td>0.22 (0.9999)</td>
</tr>
<tr>
<td>$U(H, \gamma)$</td>
<td>81.08</td>
<td>0.27</td>
<td>0.33</td>
<td>0.25</td>
<td>0.22 (0.9999)</td>
</tr>
</tbody>
</table>

Table 1: Instances corresponding to the Wilson matrix: comparison of the solutions produced by the various techniques in terms of the difference $\| \tilde{x} - \tilde{x}^{\text{ref}} \|$, where $\tilde{x}$ denotes the optimal solution to (7). PA refers to the probabilistic approach of the present paper (the value and average values of the probability function for the corresponding solutions are shown in parentheses). $\| \tilde{x} - \tilde{x}^{\text{ref}} \|$ with $\tilde{x}^{\text{ref}} = [1 \ldots 1]^T$.

### 3.3 Strong Ill-Conditioning: The Hilbert Matrix

The second test-case is based on the very ill-conditioned Hilbert matrix ([11]), a non-singular ($\det(H) = 1$), $n \times n$ symmetric, positive-definite matrix defined by $h_{ij} = \frac{1}{i+j-1}$. For $m = n = 15$, the matrix is:

$$
\begin{pmatrix}
1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{14} & \frac{1}{15} \\
\frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{14} & \frac{1}{16} \\
\frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \cdots & \frac{1}{15} \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{1}{14} & \frac{1}{15} & \frac{1}{16} & \frac{1}{17} & \frac{1}{18} & \frac{1}{19} \\
\frac{1}{15} & \frac{1}{16} & \frac{1}{17} & \frac{1}{18} & \frac{1}{19} & \frac{1}{20}
\end{pmatrix}
$$

Its condition number is very high: $\text{cond}(H) = 2.8 \times 10^{17}$, which renders difficult the numerical resolution of a linear system with this matrix. This level of condition number is typically the one observed on matrices arising from the discretization of a Fredholm equation of first kind.

The perturbed system differs from the original one of only four coefficients subject to perturbation of 1%:

- for the matrix: $\tilde{H}(3,4) = 0.1657$, $\tilde{H}(6,8) = 0.0779$, $\tilde{H}(12,12) = 0.0445$, $\tilde{H}(13,2) = 0.0704$ (instead of 0.1667, 0.0769, 0.0435, 0.0714 respectively);
- for the vector: $\tilde{y}(3) = 87.0979$, $\tilde{y}(6) = 61.3742$, $\tilde{y}(12) = 39.8648$, $\tilde{y}(13) = 37.7254$ (instead of 87.0969, 61.3732, 39.8658, 37.7244 respectively).

For the other instances, the value $\gamma = 0.01$ has been used.

Table 2 reports the obtained results. For the systems with Gaussian and Uniform matrices, we report below mean values of $\| \tilde{x} - \tilde{x}^{\text{ref}} \|$ for 20 random realizations.

<table>
<thead>
<tr>
<th>Hilbert</th>
<th>GIMP</th>
<th>KACZ</th>
<th>TIKO</th>
<th>TSVD</th>
<th>PA (Prob)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H + \delta H$</td>
<td>55367</td>
<td>12</td>
<td>16</td>
<td>48</td>
<td>1.71 (0.9999)</td>
</tr>
<tr>
<td>$G(H, \gamma)$</td>
<td>1734</td>
<td>11</td>
<td>16</td>
<td>87</td>
<td>3.56 (0.9999)</td>
</tr>
<tr>
<td>$U(H, \gamma)$</td>
<td>8911</td>
<td>12</td>
<td>16</td>
<td>75</td>
<td>2.44 (0.9999)</td>
</tr>
</tbody>
</table>

Table 2: Instances corresponding to the $15 \times 15$ Hilbert matrix. $\| \tilde{x} - \tilde{x}^{\text{ref}} \|$ with $\tilde{x}^{\text{ref}} = [1, 2, 3, \ldots, n]^T$.

### 3.4 Propagation of errors: The Nagasaka Matrix

The third test-case, based on the Nagasaka matrix [12], a tridiagonal $84 \times 84$ matrix with high condition number ($\text{cond}(H) \approx 10^{17}$), is defined by:
\[
H = \begin{pmatrix}
6 & 1 & 0 \\
8 & 6 & 1 \\
. & . & . \\
8 & 6 & 1 \\
0 & 8 & 6
\end{pmatrix}, \quad x = \begin{pmatrix}
1 \\
2 \\
. \\
83 \\
84
\end{pmatrix}, \quad y = \begin{pmatrix}
82 \\
23 \\
. \\
1238 \\
1168
\end{pmatrix}
\]

The Nagasaka matrix has been designed for showing the effect of propagation of errors on estimation of \(x\). Resolution of a linear system with Nagasaka matrix using the classical gaussian elimination leads to a desastrous estimation of the solution.

The perturbed system \(H + \delta H\) is obtained by applying a small change (by \(\delta = 0.01\)) of a few coefficients:

- the perturbed matrix is the Nagasaka matrix except for 8 coefficients which are defined as:
  \(\tilde{H}(3, 4) = H(3, 4) - \delta; \tilde{H}(6, 7) = H(6, 7) + \delta; \tilde{H}(21, 22) = H(21, 22) + \delta; \tilde{H}(43, 44) = H(43, 44) - \delta; \tilde{H}(52, 51) = H(52, 51) - \delta; \tilde{H}(66, 67) = H(66, 67) - \delta; \tilde{H}(81, 82) = H(81, 82) + \delta; \tilde{H}(82, 81) = H(82, 81) + \delta;\)
- the 8 coefficients of the perturbed vector are \(\tilde{y}(3) = y(3) + \delta; \tilde{y}(6) = y(6) + \delta; \tilde{y}(21) = y(21) - \delta; \tilde{y}(43) = y(43) + \delta; \tilde{y}(52) = y(52) + \delta; \tilde{y}(66) = y(66) - \delta; \tilde{y}(81) = y(81) - \delta; \tilde{y}(82) = y(82) - \delta.\)

For the other instances, the value \(\gamma = 0.01\) has been used.

Table 3 reports the obtained results. For the systems with Gaussian and Uniform matrices, we report below mean values of \(\|x_{\text{ref}} - \tilde{x}\|\) for 20 random realizations.

<table>
<thead>
<tr>
<th>Nagasaka</th>
<th>GIMP</th>
<th>KACZ</th>
<th>TIKO</th>
<th>TSVD</th>
<th>PA (Prob)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H + \delta H)</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>38</td>
<td>21 (0.9999)</td>
</tr>
<tr>
<td>(G(H, \gamma))</td>
<td>37</td>
<td>37</td>
<td>37</td>
<td>39</td>
<td>22 (0.9999)</td>
</tr>
<tr>
<td>(U(H, \gamma))</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>39</td>
<td>20 (0.9999)</td>
</tr>
</tbody>
</table>

Table 3: Instances corresponding to the 84 × 84 Nagasaka matrix. \(\|x_{\text{ref}} - \tilde{x}\|\) with \(x = [1, 2, 3, \ldots, n]^T\).

### 3.5 Sensitivity of the Solution w.r.t. the Choice of \(\alpha\) Parameter

By contrast with the Tikhonov or TSVD methods, where the choice of parameters (\(\lambda\), number of singular values to be kept) is difficult, an interesting point to notice is the robustness of the probabilistic approach with respect to the choice of the parameter \(\alpha\) which is the main parameter on which the method relies (the \(\varepsilon_i\) parameters being deduced by solving (8)).

Indeed, it can be observed that the quality of reconstruction remains the same for a wide range of values for \(\alpha\). As illustrated in Table 4 for the Wilson’s test-case with \(G(H, \gamma)\), the quality of estimation, measured as \(q = \|x_{\text{ref}} - \tilde{x}\|\), obtained with the probabilistic approach takes on the same value (\(\approx 0.25\)) whatever the value of alpha in the range \([0.25, 0.35]\).

<table>
<thead>
<tr>
<th>(\alpha)</th>
<th>0.25</th>
<th>0.26</th>
<th>0.27</th>
<th>0.28</th>
<th>0.29</th>
<th>0.30</th>
<th>0.31</th>
<th>0.32</th>
<th>0.33</th>
<th>0.34</th>
<th>0.35</th>
</tr>
</thead>
<tbody>
<tr>
<td>(q)</td>
<td>0.258</td>
<td>0.2506</td>
<td>0.2505</td>
<td>0.2504</td>
<td>0.2503</td>
<td>0.2502</td>
<td>0.2502</td>
<td>0.2502</td>
<td>0.2501</td>
<td>0.2501</td>
<td>0.2501</td>
</tr>
</tbody>
</table>

Table 4: Illustrating the stability of \(q = \|x_{\text{ref}} - \tilde{x}\|\) when varying \(\alpha\) in the range \([0.25, 0.35]\) for Wilson’s test-case with \(G(H, \gamma)\).
4 Conclusions, Perspectives

A probabilistic method has been proposed for dealing with linear inverse problems. Preliminary experiments on a series of simulated instances (some featuring the type of ill-conditioning commonly observed in hard inverse problems) appear to be quite encouraging:

- Performances of the proposed probabilistic approach are as good as known methods, often better, specially when facing strong ill-conditioning, propagation of errors and uncertainty on data;
- The method is robust w.r.t. the choice of parameters (essentially the value of $\alpha$ estimating the level of uncertainty on the coefficients of the matrix and the vector);
- Many inverse problems such as Fredholm type problems are known for their ill-conditioning. Our experiments have shown the potential of the new method for facing ill-conditioning, so we can be optimistic for its application to true inverse problems arising from physics.

Among the many perspectives opened by the present work, we can mention: the evaluation of the quality of the approximate solutions to (4) obtained as optimal solutions to the max-min problem by efficiently computing upper bound values; the search for improved, more powerful, global optimality tests addressing the case of joint two-sided chance constraints (ongoing research); the extension to nongaussian probability distributions and finally the application of our probabilistic method to real-life problems (nondestructive evaluation, mechanics, ...).

5 References