Journal Pre-proof

The curse of dimensionality in inverse problems

Juan L. Fernández-Martínez, Zulima Fernández-Muñiz

PII: S0377-0427(19)30576-X
DOI: https://doi.org/10.1016/j.cam.2019.112571
Reference: CAM 112571

To appear in: *Journal of Computational and Applied Mathematics*

Received date : 23 November 2018
Revised date : 1 October 2019

Please cite this article as: J.L. Fernández-Martínez and Z. Fernández-Muñiz, The curse of dimensionality in inverse problems, *Journal of Computational and Applied Mathematics* (2019), doi: https://doi.org/10.1016/j.cam.2019.112571.

Highlights (for review)

- Connection of curse of dimensionality to uncertainty analysis of inverse problems.
- Dependency of the sampling probability on ill-conditioning of the linear system.
- The bounds provided by linear analysis are very large for nonlinear inverse problems.
- Only 4-6 independent dimensions can be efficiently sampled by search methods.
- Model reduction techniques serve to increase the sampling probability.

The curse of dimensionality in inverse problems

Juan L. Fernández-Martínez[1*], Zulima Fernández-Muñiz[1]

[1]Group of Inverse Problems, Optimization and Machine Learning.

Department of Mathematics. University of Oviedo, Spain.

C/Federico García Lorca, 18, 33007 Oviedo

jlfm@uniovi.es, zulima@uniovi.es

Keywords: Inverse problems, Uncertainty analysis, Model reduction, Curse of dimensionality.

Abstract

Nonlinear inverse problems in real problems in industry have typically a very underdetermined character due to the high number of parameters that are usually needed to achieve accurate forward predictions. The corresponding inverse problem is ill-posed, that is, there exist many solutions which are compatible with the prior information, fitting the observed data within the same error bounds. These solutions are located in (one or several) flat curvilinear and disconnected valleys of the cost function topography. The random sampling of these equivalent models is impossible due to the curse of dimensionality and to the high computational cost needed to provide the corresponding forward predictions. This paper generalizes the curse of dimensionality to linear and nonlinear inverse problems outlining the main differences between them. With a simple 2D example we show that nonlinearities allow for a reduction in size of the nonlinear equivalence region that could be embedded in a linear hyperquadric with smaller condition number than the corresponding linearized equivalence region. We also analyze the effect of the regularization in the posterior sampling, and that of the dimensionality reduction, which is needed to perform efficient sampling of the region of uncertainty equivalence in high

dimensional problems. We hope that the additional theoretical knowledge provided by this research will help practitioners to design more efficient methods of sampling.

1. UNCERTAINTY ANALYSIS IN INVERSE PROBLEMS

Inverse problems are often encountered in many fields of technology, including engineering, science and mathematics. Solving an inverse problem entails the determination of certain model parameters from a set of observed data or measurements. The mapping of data to model parameters is done via a physical system in the case of an inverse problem, or through a regression model or a classifier in the case where the physics is unknown. For example, in the field of geophysics the model parameters, such as the electrical conductivity, the density, the magnetic permeability, the porosity or the seismic velocity are identified from some projections that are acquired on the surface of the earth (i.e. observed data) and are related to the model parameters through a forward model. More precisely, an inverse problem may be formulated in discrete form by defining the forward problem as follows:

$$\mathbf{F}(\mathbf{m}) = \mathbf{d} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{m} \in \mathbf{R}^n$ is the estimated geophysical model that belongs to a set of admissible models **M** defined in terms of some prior knowledge (e.g., geological interpretation), $\mathbf{d} \in \mathbf{R}^s$ are the observed data, and $\mathbf{F}(\mathbf{m}) = \left(f_1(\mathbf{m}), f_2(\mathbf{m}), ..., f_s(\mathbf{m}) \right)$, represents the forward model, with $f_j(\mathbf{m})$ being the scalar field that accounts for the *j*-th data. The term $\boldsymbol{\varepsilon}$ is introduced to explain that the relationship $\mathbf{F}(\mathbf{m}) = \mathbf{d}$ is not perfect, that is, this set of equations might not have any solution.

The inverse problem consists in finding **m,** given **F** and **d.** A classification problem can be cast in the same way, with **F** being the classifier that is built to emulate the physics

and **d** the set of observed classes. In geology, this second kind of problems is more atypical and involves machine learning techniques (Caté et al., 2017).

Both types can be referred to as parameter identification problems, which are commonly ill-posed, that is, there exist different kinds of model parameters sets **m** that predict the observed data with the same precision and are compatible with the prior information that is at disposal. That is to say, the geological/geophysical model that we hypothesize as the real one is not unique. This fact is usually called the model uncertainty and the discipline that tries to quantify it, model appraisal (Snieder and Tampert, 1999; Scales and Snieder, 2000).

Uncertainty exists in inverse problems because of a variety of factors, such as poor data calibration, contamination and noise in data measurements, discrete data coverage, approximated physics and conceptualization, discretization of continuous inverse problems, linearization and numerical approximations, model physical assumptions (e.g., isotropy, homogeneity, anisotropy, etc.), limited bandwidth, poor resolution, and so forth. Snieder (1998) studied the role of nonlinearity in inverse problems, introducing the relationship *Inversion=Estimation+Appraisal*, pointing out that non-uniqueness and error propagation are the main reasons for uncertainty assessment.

The problem of uncertainty has a natural interpretation in a Bayesian framework (see Scales and Tenorio, 2001). Bayes' rule (1763) states that a set of model parameters is more probable if it explains the observed data with a higher probability, that is, if the observed data are more likely to have happened:

$$P(\mathbf{m}/\mathbf{d}) = \frac{P(\mathbf{d}/\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}. \tag{2}$$

The term $P(\mathbf{d}/\mathbf{m})$ is called the likelihood and typically depends exponentially of the data misfit $\|\mathbf{F}(\mathbf{m})-\mathbf{d}\|_p$ in a certain norm *p*, being $P(\mathbf{m})$ the prior probability and

$P(\mathbf{d})$ the evidence that is usually considered as a normalization constant in these approaches.

Besides, assuming the likelihood and the prior probability in the family of Gaussians it is easy to prove the equivalence between the maximum likelihood method that finds the mode of $P(\mathbf{d}/\mathbf{m})$ and the deterministic least squares method with the regularization introduced by the prior information (see for instance Aster et al., 2005; Fernández-Martínez et al., 2013).

From the deterministic point of view, uncertainty assessment involves finding the family ($\mathbf{M}_{tol}$) of geophysical models, **m**, that are consistent with our prior knowledge and fit the observed data $\mathbf{d} \in \mathbf{R}^s$ (comprising all the observables) within the same relative misfit tolerance (E_{tol}):

$$\mathbf{m} \in \mathbf{M}_{tol} \ : \ \frac{\left\| \mathbf{F}(\mathbf{m}) - \mathbf{d} \right\|_2}{\left\| \mathbf{d} \right\|_2} \leq E_{tol}. \tag{3}$$

In the case of linear problems, the region of uncertainty will be called $\mathbf{L}_{tol}$, and it is defined as follows:

$$\mathbf{m} \in \mathbf{L}_{tol} \ : \ \frac{\left\| \mathbf{F}\mathbf{m} - \mathbf{d} \right\|_2}{\left\| \mathbf{d} \right\|_2} \leq E_{tol}. \tag{4}$$

Uncertainty analysis is important, since this ambiguity in the model parameters determination generates a risk in the decisions that can cause possible negative outcomes. Uncertainty analysis is quite well understood in linear inverse problems with linear algebra and linear inverse theory (see for instance Menke, 1984; Aster et al., 2005).

Nevertheless, in nonlinear inverse problems the uncertainty analysis has been intimately related to random sampling methods and Bayesian frameworks (Mosegaard

and Tarantola, 1995; Sambridge 1999; Sambridge and Mosegaard, 2002; Tarantola, 2005), due to the fact that the linearized inverse problem only provides a local linear approximation to the nonlinear variability (see for instance Alumbaugh, 2002; Fernández Álvarez et al., 2008; Fernández-Martínez et al., 2013). However, most of these stochastic sampling schemes depend too strongly on the dimension of the parameter space and can often require intractable numbers of forward solves (e.g., Haario et al., 2001). Scales and Snieder (2000) pointed that Monte Carlo sampling methods are not feasible for large-scale inverse problems, and developing an operational theory to account for the appraisal problem of nonlinear inverse problems with large number of parameters is one the biggest theoretical and practical challenges in inversion, which is much more important than establishing uniqueness proofs of idealized mathematical problems.

Fernández-Martínez et al. (2012, 2013) analyzed deterministically the uncertainty space of linear and nonlinear inverse problems through the cost function landscape (3) in the regions of low misfits. They showed that the region of equivalence $\mathbf{M}_{tol}$ in linear inverse problems is the part of the model space inside the hyperquadric surface of equivalence, whose axes depend on the error tolerance, E_{tol}, and on the ill-conditioning of the matrix of the linear system involved. This hyperquadric surface varies from a very oblong ellipsoid to an elliptical cylinder (rank-deficient systems). Least-squares method tries to determine the center of this surface. Obviously in the case of flat elongated valley (rank deficient system) this problem does not admit a unique solution. The determination of the center of the hyperquadric is very sensitive to the effect of noise and also to the type of regularization that it is used to stabilize the inversion (Fernandez-Martinez et al., 2014 a,b).

While the curse of dimensionality is often talked about in optimization and inverse problem settings, the blessings of dimensionality are less well-known or utilized (Donoho, 2000). Asymptotic methods in statistical physics allow derivation of results in very high dimensional settings that would be difficult in moderate dimensions. There may be potential for applying high-dimensional approximation theory and probability theory (e.g. Johnstone, 1998, 2000) to exploit the blessings of dimensionality for inverse problems.

Advantages and/or drawbacks of the dimensionality, in this paper we generalize the results shown in Tarantola (2006) about the curse of dimensionality in sampling, to the case of linear and nonlinear inverse problems, expanding the results that were already outlined in Fernández-Martínez (2015). In this previous work, Tarantola (2006) just showed that sampling inside an isotropic space (hypersphere) is almost impossible for more than 10 independent dimensions, trying to make aware practitioners that the model space is almost empty of good solutions, and the uncertainty analysis of the inverse solution is a very complicated problem.

In this paper, we generalize the results known as the curse of dimensionality (Bellman, 1961) that refers to the probability sampling within a hypersphere to the uncertainty analysis in inverse problems, relating the sampling probability to the ill-conditioning of the linear system and studying the effect of model reduction. This idea was briefly outlined in Fernández-Martínez (2015). The curse of dimensionality was used to explain the difficulty of sampling high dimensional model spaces by means of random sampling methodologies (Curtis A., Lomax A., 2001; Tarantola A., 2006). This paper has the novelty of explaining this fact by means of mathematical analysis. Besides, we show that the use of an orthonormal reduced basis set does not alter the ill-conditioning of the system matrix but serves to reduce considerably the dimensionality of the model space

because the solution is searched in a subspace. This analysis is generalized to the case of nonlinear problems either by reparameterization or by embedding the nonlinear region of equivalence within the linearized region of equivalence for a higher misfit. We show in a simple 2D example that the nonlinearities reduce the size of the nonlinear equivalence region compared to the linearized equivalence region, provoking an increase of the sampling probability for higher dimensions. This fact could explain why sampling is possible when an informative prior is adopted. Although this paper remains theoretical, it provides new insights about the directions that should be adopted to improve the uncertainty analysis in linear and nonlinear problems, and shows that for high-dimensional problems, brute force and/or fully random sampling approaches cannot be used to deal with the uncertainty problem.

2. THE CURSE OF DIMENSIONALITY IN INVERSE PROBLEMS

2.1 Linear problems

In the previous section, we have seen that the challenge consists mainly in sampling low misfit elongated valleys of the cost function in order to obtain representative samples of the linear/nonlinear uncertainty region. This sampling is hampered by the dimensionality of the problem, and the explanation is as follows: let us imagine that we want to sample inside the circle that it is inscribed in the square of side r. The conditional probability of throwing a dart inside the circle, knowing that it is inside the square is: $\frac{\pi\left(\frac{r}{2}\right)^2}{r^2}=\frac{\pi}{4}$. In this case, the square plays the role of our search space. If we increase one dimension, then the probability of sampling within the sphere that it is inscribed into a cube of side $2r$ is $\frac{\frac{4}{3}\pi\left(\frac{r}{2}\right)^3}{r^3}=\frac{\pi}{6}$. This probability goes very fast to zero

(see for instance Tarantola, 2006) when the number of dimensions increases and $n \geq 10$. In his own words: "*large-dimensional spaces tend to be terribly empty. Hitting by chance the circle inscribed in a square is easy. Hitting by chance the sphere inscribed in a cube is a little bit more difficult. When the dimension n of the space grows, the probability of hitting the hypersphere inscribed in a hypercube rapidly tends to zero (for n>10)".*

The deduction of the formulas concerning to the volumes of the hypersphere and the hyperellipsoid can be consulted in Wilson (2010). The volumes V_S and V_C of the hypersphere of radius *r* inscribed in a hypercube (of side *2r*) are:

$$\left.\begin{array}{l} V_S = \dfrac{2\pi^{n/2} r^n}{n\Gamma\left(n/2\right)}, \\ V_C = \left(2r\right)^n \end{array}\right\} \Rightarrow P_i = P(x \in S \,|\, x \in C) = \frac{\pi^{n/2}}{n2^{n-1}\Gamma\left(n/2\right)}. \quad (5)$$

In what follows $P_i = P(x \in S \,|\, x \in C)$ is the probability of sampling a point within the hypersphere, conditioned of being inside the hypercube. The subscript *i* stands for isotropic, that is, the uncertainty is the same in all the directions of the space (sampling inside the hypercube). Figure 1 shows the probability $P_i\left(n\right)$ as a function of the number of dimensions *n.* It can be observed that this probability approaches to 0 for *n* greater than 10. This simple fact known as the dimensionality curse (Curtis and Lomax, 2001) serves to explain why the random exploration of large-dimensional spaces is unfeasible. This would mean that no more than 10 dimensions could be handled to study an isotropic space of uncertainty. The phrase "curse of dimensionality" was probably first coined by Bellman (1961) in the context of optimization over a large number of variables.

Nevertheless, it has been shown that the uncertainty in inverse problems (and more generally in any decision problem) is not the same in all the directions of the model

space, that is, the region of equivalence is elongated and the uncertainty regions have an anisotropic nature (Fernández-Martínez et al., 2012). Therefore, it is easy to understand that the number of effective dimensions that we have to actually sample in this anisotropic uncertainty world is much less than the actual number of dimensions, depending on the valley eccentricity, which is related to the ill-conditioning (condition number) of the linear system matrix $\mathbf{F}$. We will also prove that this situation improves for nonlinear inverse problems due to the effect of the nonlinearities that serve to bound the size of the nonlinear equivalence region.

The demonstration in the case of linear problems is as follows: in the case of a purely overdetermined linear problem, $\mathbf{Fm}=\mathbf{d}$, where the matrix of the linear system $\mathbf{F}$ has spectrum with all non-null singular values $\alpha_1>\alpha_2>\cdots>\alpha_n>0$, then the condition number is defined as the ratio of the maximum and minimum singular values of $\mathbf{F}$, $\kappa=\frac{\alpha_1}{\alpha_n}$, and the linear region of equivalence $\mathbf{L}_{tol}$ is expressed as:

$$\frac{\|\mathbf{Fm}-\mathbf{d}\|_2}{\|\mathbf{d}\|_2}\le tol \Rightarrow (\Delta\mathbf{m})^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{F}(\Delta\mathbf{m})\le tol^2\|\mathbf{d}\|_2^2. \tag{6}$$

where $\Delta\mathbf{m}=\mathbf{m}-(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{d}$ is the model increment referred to the least-squares solution. To arrive to formula (6) we took account that the Moore-Penrose pseudoinverse of $\mathbf{F}$ writes: $\mathbf{F}^{\dagger}=(\mathbf{F}^{\mathrm{T}}\mathbf{F})^{-1}\mathbf{F}^{\mathrm{T}}$.

Considering the singular value decomposition of $\mathbf{F}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{T}$, and referring to the $\mathbf{V}$ base, we arrive at (Fernández-Martínez et al., 2012, 2013):

$$(\Delta\mathbf{m}_V)^{\mathrm{T}}\mathbf{\Sigma}^{\mathrm{T}}\mathbf{\Sigma}(\Delta\mathbf{m}_V)\le tol^2\|\mathbf{d}\|_2^2 \Leftrightarrow \sum_{k=1}^{n=rank(\mathbf{F})}\left(\frac{\Delta m_{Vk}}{1/\alpha_k}\right)^2=tol^2\|\mathbf{d}\|_2^2, \tag{7}$$

211 which is a hyperellipsoid with semi-axes $a_i = \frac{tol\,\|\mathbf{d}\|_2}{\alpha_i}$. The volume of a hyperellipsoid

212 with semi-axes $a_1, a_2, ..., a_n$ is $V_E = \frac{2\pi^{n/2} a_1 a_2 \cdots a_n}{n\Gamma(n/2)}$. Therefore the probability of

213 sampling inside $\mathbf{L}_{tol}$, conditioned to the fact that the point is inside of a hypercube of

214 side $H = 2\max(a_1, a_2, ..., a_n) = 2a_n = \frac{2tol\,\|\mathbf{d}\|_2}{\alpha_n}$ is:

215
$$\left.\begin{aligned} V_E &= \frac{2\pi^{n/2}\frac{\left(tol\,\|\mathbf{d}\|_2\right)^n}{\alpha_1\alpha_2\cdots\alpha_n}}{n\Gamma(n/2)} \\ V_C &= \left(\frac{2tol\,\|\mathbf{d}\|_2}{\alpha_n}\right)^n \end{aligned}\right\} \Rightarrow P(x\in E \,|\, x\in C) = \frac{\pi^{n/2}}{n2^{n-1}\Gamma(n/2)}\frac{\alpha_n^{n-1}}{\alpha_1\alpha_2\cdots\alpha_{n-1}}. \quad (8)$$

216 For instance, in the 2D space (Figure 2A) the probability of sampling inside the ellipse,

217 when the point is inside of the hypercube, is

218
$$P_a = \frac{\pi a_1 a_2}{(2a_2)^2} = \frac{\pi}{4}\frac{a_1}{a_2} = \frac{\pi}{4}\frac{\alpha_2}{\alpha_1} = \frac{\pi}{4}\kappa^{-1} = P_i\,\kappa^{-1}. \quad (9)$$

219 In the 3D space (Figure 2B), the probability P_a is

220
$$P_a = \frac{\frac{4}{3}\pi a_1 a_2 a_3}{(2a_3)^2} = \frac{\pi}{6}\frac{a_1 a_2}{a_3^2} = \frac{\pi}{6}\frac{\alpha_3^2}{\alpha_1\alpha_2} > \frac{\pi}{6}\frac{\alpha_3^2}{\alpha_1^2} = P_i\,\kappa^{-2}. \quad (10)$$

221 These results coincide with those that are obtained by applying equation (8) for *n=2,3.*

222 Denoting by $P_a = P(x\in E \,|\, x\in C)$ the anisotropic conditional probability, and

223 introducing the anisotropic constant $C_a = \frac{\alpha_n^{n-1}}{\alpha_1\alpha_2\cdots\alpha_{n-1}}$, we have:

224
$$P_a = P_i\,C_a, \quad (11)$$

225 that is, the anisotropic sampling probability is proportional to the isotropic sampling

226 probability through the anisotropic constant that depends on the ill-conditioning of **F.**

The following lower and upper bounds can be found for C_a

$$C_a > \frac{\alpha_n^{n-1}}{\alpha_1^{n-1}} = \kappa^{1-n}, \tag{12}$$

$$C_a = \frac{\alpha_n}{\alpha_1}\frac{\alpha_n}{\alpha_1}\frac{\alpha_1}{\alpha_2}\cdots\frac{\alpha_n}{\alpha_1}\frac{\alpha_1}{\alpha_{n-1}} = \kappa^{(1-n)}\frac{\alpha_1^{n-2}}{\alpha_2\cdots\alpha_{n-1}} < \kappa^{-1}. \tag{13}$$

Therefore, we have the lower and upper bounds of the anisotropic probability:

$$P_{\min} = P_i\,\kappa^{1-n} < P_a < P_i\,\kappa^{-1} = P_{\max}. \tag{14}$$

Figure 3 shows the upper bound ($P_{\max}$) of the sampling probability (P_a) inside of the linear region of equivalence, $\mathbf{L}_{tol}$, for linear systems with dimensions between $n=2$ and 12, and condition numbers $\kappa = 10, 10^2$ and 10^3, which are not extremely high. It can be observed that these probabilities are very small, and drop to zero for $\kappa = 10^3$ when *n* is greater than 3, for $\kappa = 10^2$ when *n* is greater than 7, and for $\kappa = 10$ when *n* is greater than 9. Therefore, ten independent dimensions can only be sampled in isotropic spaces which is not the case of the unceratinty analysis in linear inverse problems, where the sampling probability decreases dramatically fast when the condition number and/or the number of dimensions increases.

Figure 4 illustrates these results showing the anisotropic sampling probability P_a and the upper and lower bounds, $P_{\max}$ and $P_{\min}$, for different square random matrices with dimensions between 2 and 100, with their corresponding condition numbers. This plot is obtained by averaging the results obtained from 5000 random simulations. It can be observed that although these matrices are always full rank (due to their random generation that provides linear independence of their column vectors) their condition numbers vary from $10^{1.97}$ to $10^{3.5}$ and the (anisotropic) sampling probability within the

$\mathbf{L}_{tol}$ region is close to zero. The algorithm to generate these results is given in Appendix A.

It has been analytically shown that the regularization in linear inverse problems has the effect of improving their condition number, since it bounds the axes of the linear equivalence region with higher uncertainty, decreasing the condition number (Fernández-Martínez et al., 2013). The condition number of the system matrix is related to the minimum and maximum axes of the linear hyperquadric as follows:

$$\kappa = \frac{\alpha_1}{\alpha_n} = \frac{e_{\max}}{e_{\min}}, \tag{15}$$

with

$$e_{\max} = \frac{tol\,\|\mathbf{d}\|_2}{\alpha_n}, \quad e_{\min} = \frac{tol\,\|\mathbf{d}\|_2}{\alpha_1}. \tag{16}$$

Therefore, taking into account (14) and (15), the regularization increases the lower and upper bounds of the anisotropic probability P_a, and has the effect of improving the sampling. Besides, taking into account that P_i increases by reducing n (the number of dimensions), the model reduction is the simplest solution that can be adopted to improve the sampling. This analysis will be shown later in this paper.

Finally, the lower bound in (14), $P_{\min}$, can be used to establish the maximum number of dimensions and/or the amount of regularization needed to fulfil $P_a \geq P_{\min}$. For that purpose, we have to solve the nonlinear equation:

$$P_{\min} = P_i\,\kappa^{1-n} = \frac{\pi^{n/2}}{n2^{n-1}\Gamma\left(n/2\right)}\kappa^{1-n}. \tag{17}$$

Therefore, supposing that we know the condition number κ, the relationship (17) can be rewritten as follows:

$$\log\left(P_{\min}\right)=\frac{n}{2}\log\left(\pi\right)-\log\left(n\right)-\left(n-1\right)\log 2-\log\left(\Gamma\left(n/2\right)\right)+\left(1-n\right)\log\kappa. \quad (18)$$

Figure 5 shows the lower probability bounds for three different values of the condition number $\kappa=10,100$ and 1000 with dimensions from 1 to 10. It can be observed that the lower bound $P_{\min}$ drops very fast with the ill-conditioning of F. Therefore, without using any kind of regularization (and prior information) sampling very ill-conditioned problems via random sampling methodologies is literally impossible in high dimensions. This is a very important conclusion, which does not hamper the Bayesian approach of inverse problems (see for instance Scales and Tenorio, 2001; Rappel et al., 2018, 2019).

2.2 Application to gravimetric inversion

In this section we show the application of the previous results to a 1D geophysical model that accounts for the gravity anomaly $g_z\left(s_k\right)$ at the observation point s_k located in the surface, generated by a linear dense body with density distribution $\rho\left(x\right)$ located at a constant depth D (see Blakely, 1995):

$$g_z\left(s_k\right)=-G\int_a^b\frac{D}{\left(D^2+\left(s_k-x\right)^2\right)^{3/2}}\rho\left(x\right)dx. \quad (19)$$

This geophysical problem corresponds to a Fredholm integral equation of first kind (see Hansen, 2010).

The discrete inverse problem consists in expanding the density function $\rho\left(x\right)$ in a set of basis functions $\left\{\phi_i(x), i=1,...,n\right\}$ as follows (Fernández-Muñiz et al., 2015):

$$g_z(s_k) = \int_a^b K(x, s_k)\left(\sum_{i=1}^{n} \rho_i \phi_i(x)\right) dx = \sum_{i=1}^{n} \rho_i \left(\int_a^b K(x, s_k)\phi_i(x)\, dx\right) =$$

289 (20)

$$= \sum_{i=1}^{n} \rho_i \langle K(x, s_k), \phi_i(x)\rangle = \sum_{i=1}^{n} \rho_i G_{ki}, \quad k = 1, ..., m,$$

290 where $K(x, s_k) = -G\dfrac{D}{\left(D^2 + (s_k - x)^2\right)^{3/2}}$, and ρ_i are the coordinates of the unknown

291 density function $\rho(x)$ into the set of basis functions $\{\phi_i(x), i = 1, ..., n\}$, and

292 $$G_{ki} = \langle K(x, s_k), \phi_i(x)\rangle = \int_{x_{i-1}}^{x_i} K(x, s_k)\, dx, \tag{21}$$

293 is the projection of the kernel function $K(x, s_k)$ at the point s_k onto the basis function

294 ϕ_i.

295 In this case, we have used the piecewise continuous functions (pixel basis set):

296 $$\phi_i(x) = \begin{cases} 1/\sqrt{x_i - x_{i-1}}, & x \in [x_{i-1}, x_i] = \left[a + (i-1)\dfrac{b-a}{n-1}, a + i\dfrac{b-a}{n-1}\right], \\ 0 \quad \text{otherwise} \end{cases} \tag{22}$$

297 As result of the discretization, the discrete inverse problem can be written:

298 $$\mathbf{G\rho} = \mathbf{g_z}, \tag{23}$$

299 where $\mathbf{G} \in M_{m\times n}(\mathbf{R})$ is the matrix that contains the values of the projected kernel,

300 $\mathbf{G}(k, i) = G_{ki}$, $\boldsymbol{\rho} = (\rho_1, \rho_2, ..., \rho_n)$ and $\mathbf{g_z} = (g_z(s_1), g_z(s_2), ..., g_z(s_m))$. This linear system

301 turns to be very ill-conditioned, due to the low spatial resolution of the geophysical

302 kernel, $K(x, s_k)$.

303 Figure 6 shows the condition number for several examples of this simple linear

304 gravimetric inverse problem as a function of the number of data (*m*), and the number of

305 model parameters (*n*). In real problems the linear systems are typically under-

306 determined (or rank-deficient) since data acquisition has a cost that we try to minimize,

and the number of model parameters only depend on the number of pixel basis used for the discretization of the continuous inverse problem and has no cost. The rank deficient character implies that there exists a redundancy, both, in data acquisition and model parameterization. In this particular case, the rank is around 17 independently of the size of the linear discrete system.

It can be observed that the condition number (in log10 scale) is very high with an order of magnitude between 14 and 19 (condition numbers in the interval $[10^{14},10^{20}]$). This provides an idea of the very low sampling probabilities that can be achieved for *n>6*. This result highlights the importance of reducing the dimension to sample the equivalent models (using a reduced basis set, and it is relevant since it relates the curse of dimensionality in the uncertainty analysis of linear inverse problems with the ill-conditioning of their system matrix.

In Tarantola's words (Tarantola, 2006): "*sampling uncertainty in a nonlinear problem is like finding a needle (a curvilinear needle for nonlinear inverse problems) in a haystack*". In other words, in an anisotropic universe no more than 4-6 dimensions can be efficiently sampled, depending on the eccentricity of the linear equivalent region, which is related to the ill-conditioning of the linear system to be solved. This result also explains why the use of model reduction techniques and subspace methods is essential in quantifying uncertainty. The main question now resides in finding the right reduced dimensions to approach the uncertainty of any inverse problem.

In the case of rank-deficient linear systems, the zero-order Tikhonov's regularization has the effect of limiting the axes of the hyper ellipsoid to $1/\varepsilon$ in the directions of the **V** vectors spanning the null space of **F,** with ε^2 being the damping parameter used to stabilize the inversion: $\min_{\mathbf{m}} \|\mathbf{Fm}-\mathbf{d}\|_2^2 + \varepsilon^2 \|\mathbf{m}\|_2^2$. The zero-order regularization has the effect of fixing the minimum singular value of **F** to ε (Fernández-Martínez et al,

2014a). Therefore, the regularization improves the ill-conditioning of the system matrix, and also the sampling efficiency. It is important to remark that the linear hyperquadric in the case of rank deficient systems is a straight valley of infinite length in the directions spanning the kernel (or null-space) of **F**, that is, the ill-conditioning of these problems is infinite. The regularization modifies the condition number as follows: $\kappa = \frac{\alpha_1}{\varepsilon}$.

2.3 The effect of model reduction

The importance of model reduction in inverse problems has been analysed by Fernández-Martínez (2015). In this case, we try to study its effect in the uncertainty analysis.

Let us suppose now that the solution of the linear system $\mathbf{Fm} = \mathbf{d}$ is searched in a subspace of dimension *q*

$$\mathbf{m} = \sum_{k=1}^{q} \alpha_k \mathbf{v}_k = \mathbf{Q}\boldsymbol{\alpha}\,, \tag{24}$$

where $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_q\}$ is an orthogonal reduced basis of this subspace, $\mathbf{Q} = [\mathbf{v}_1\ \mathbf{v}_2 ... \mathbf{v}_q]$ is the orthogonal matrix that has $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_q\}$ as column vectors, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_q)$ are the coordinates of $\mathbf{m}$ in this subspace. Now the problem consists in finding $\boldsymbol{\alpha}$ such as $\|\mathbf{FQ}\boldsymbol{\alpha} - \mathbf{b}\|_2$ is minimum.

Taking into account the SVD of **F** we have**:**

$$\mathbf{FQ} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\mathrm{T}}\mathbf{Q} = \mathbf{U}\boldsymbol{\Sigma}\left(\mathbf{Q}^{\mathrm{T}}\mathbf{V}\right)^{\mathrm{T}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{P}^{\mathrm{T}}, \tag{25}$$

where $\mathbf{P} = \mathbf{Q}^{\mathrm{T}}\mathbf{V}$ is also an orthogonal matrix. In conclusion: $\kappa(\mathbf{FQ}) = \kappa(\mathbf{F})$. Therefore, the orthogonal model reduction has the effect of reducing the dimension from *n* parameters to *q* principal modes, but it does not alter the conditioning of the

linear system. Although the dimensionality reduction does not provide an improvement of the condition number of the system matrix, the sampling probability increases. Different methods to define the basis set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_q\}$ are available in the literature (see for instance Tompkins et al., 2011 a,b; 2013; Fernández-Martínez, 2015; Fernández-Martínez et al., 2017).

The model reduction techniques are crucial to design operational methods able to solve the uncertainty problem independent of the dimension of the model space and of the computational cost needed to solve the forward model, as requested by Scales and Snieder (2000). The underlying reason is that the model parameters must have an intrinsic correlation introduced by the physics of the forward problem which is needed to match the observed data. Therefore, these parameters should not be sampled independently: model reduction methods take advantage of these correlations to reduce the dimension and improve sampling. These methods are compatible with the Bayesian approach of uncertainty in inverse problems.

2.3 Nonlinear problems

In the case of nonlinear problems, the region of equivalence $\mathbf{M}_{tol}$ is a valley with curvilinear shape and might be composed of different disconnected basins (Fernández-Martínez et al., 2012).

We will show in a simple synthetic case that the results shown for the linear case can be applied to the nonlinear case to establish upper bounds for the sampling probabilities. The basic idea is that the nonlinear region of equivalence for a given error bound can be imbedded into the linearized region of equivalence that is deduced from the Jacobian of the forward model taken in the solution of the inverse problem. Nevertheless, it should be pointed out that the condition number of the linearized equivalence region (linear

hyperquadric) is much higher than the condition number of the embedding hyperquadric (calculated according to the expression (6)).

To show numerically this fact, let us consider a nonlinear regression problem of the kind:

$$y_k = f(x_k; a, b) = ae^{bx_k} + \varepsilon_k, \tag{26}$$

where y_k are the observed data in a set of observation points $\{x_k\}_{k=1,\ldots,s}$, ε_k the realization of the observational noise in x_k, and (a,b) the parameters that we would like to identify. In this case, we have generated a synthetic dataset with $(a,b)=(4,3)$ as true model. The observed values have been perturbed by 3% of Gaussian white noise in order to simulate the noise in the measurement data.

To determine the linearized region of equivalence the Jacobian matrix around the true model (a_T, b_T) for the discrete dataset $\{x_k, y_k\}_{k=1,\ldots,s}$ needs to be calculated as follows:

$$\begin{cases} \dfrac{\partial y_k}{\partial a} = e^{bx_k} \\ \dfrac{\partial y_k}{\partial b} = ax_k e^{bx_k} \end{cases}, \qquad \mathbf{JF}_{(a_T,b_T)} = \begin{pmatrix} e^{b_T x_1} & a_T x_1 e^{b_T x_1} \\ e^{b_T x_2} & a_T x_2 e^{b_T x_2} \\ \vdots & \vdots \\ e^{b_T x_s} & a_T x_s e^{b_T x_s} \end{pmatrix}. \tag{27}$$

The ellipse of equivalence in the plane (a,b) for a tolerance value *tol* referred to its principal axes writes:

$$\begin{aligned} &\frac{(a-a_T)^2}{e_1^2} + \frac{(b-b_T)^2}{e_2^2} = 1, \\ &e_k = tol\,\frac{\|d_k\|_2}{\alpha_k}, \quad k = 1, 2, \end{aligned} \tag{28}$$

where α_k are the singular values of the Jacobian matrix. This ellipse has to be rotated to its principal axis provides by the eigenvectors of the matrix $\mathbf{JF}^{\mathrm{T}}\mathbf{JF}$.

In this case there are two different ways of embedding the nonlinear equivalence region for a given tolerance *tol* into the linearized equivalence region:

1. The first way consists in using the logarithmic parametrization. Figure 7A shows the nonlinear (green color) and linearized uncertainty region for relative error tolerances of 5% and of 10%, respectively, in the (a,b) plane. Figure 7B shows the same regions in the $(\ln a,b)$ plane. Therefore, in this particular case, the results shown for linear problems can be applied to the sampling of $(\ln a,b)$.

2. The second way consists in using linearization techniques by computing the Jacobian. In the previous example, the condition number of the matrix $\mathbf{JF}^{\mathrm{T}}\mathbf{JF}$ is $\frac{e_2}{e_1}=87.3$ and the corresponding ellipse of equivalence for 10% relative misfit, which includes the nonlinear region, is too big compared to the nonlinear region. For that reason, in Figure 8 we have plotted the ellipse (red line) with numerical axes ($\frac{e_2}{e_1}=15$) that almost encompasses the entire nonlinear equivalence region (green line) for a condition number 6 times smaller than the theoretical one. Therefore, in this case the sampling probability of the nonlinear equivalence region is bounded by the sampling probability of a linear inverse problem in 2 dimensions with condition number $\kappa=\frac{e_2}{e_1}=15$.

3. Following the same idea, in other nonlinear cases it is always possible to find a linear hyperquadric for a higher tolerance error that contains the nonlinear region of value tol. Obviously in this case, some models inside the hyperquadric do not belong to the nonlinear equivalence region. Then the application of relationship (14) to the hyperquadric provides approximate bounds of the anisotropic sampling probabilities in the nonlinear case. A similar analysis was performed in a nonlinear 1D-DC inverse problem (Fernández-Alvarez et al., 2008) concerning the inversion of Vertical Electrical Soundings (VES) using the logarithmic parameterization.

CONCLUSIONS

In this paper, we have generalized the curse of dimensionality known for an isotropic world to uncertainty sampling which is intrinsically anisotropic. We have shown that the sampling depends not only in the dimensionality but also in the ill-conditioning of the linear system. Albert Tarantola was more than right and whether the needle is even thinner, or the haystack is huge, it does not matter. More than 4-6 independent dimensions cannot be efficiently sampled by exhaustive grid search in the case of well-conditioned linear inverse problems. This sampling probability drops to zero for very ill-conditioned linear systems. In the case of nonlinear problems, these results can be applied to the linearized inverse problem, but the nonlinearity has a positive role in reducing the size of the nonlinear uncertainty region with respect to the corresponding linearized equivalence region. These results are independent of the sampling algorithm that is used.

As a main conclusion, model reduction techniques are needed to efficiently sample the equivalence region in linear and nonlinear inverse problems. We have shown that model reduction techniques based in orthogonal basis sets do not alter the ill-conditioning of the system matrix but increases the anisotropic sampling probability because the dimensionality is drastically reduced. Otherwise, the only way to successfully performing the sampling is adopting very informative priors, but this is not the purpose of the uncertainty analysis, that has to be able to unravel other possible (or plausible) inverse solutions with a different structure than the one shown by the solution that has been adopted (see for instance Tompkins et al., 2011 a,b, 2013). Although the results shown in this paper are theoretical at this stage, we hope that the additional knowledge provided by this research will help practitioners to design more efficient methods of sampling. Particularly, the probability bounds given in this paper could be used to

estimate the number of reduced dimensions and the amount of regularization that is needed.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their careful reading of our manuscript and their insightful comments and suggestions, that helped us to improve the quality of this manuscript.

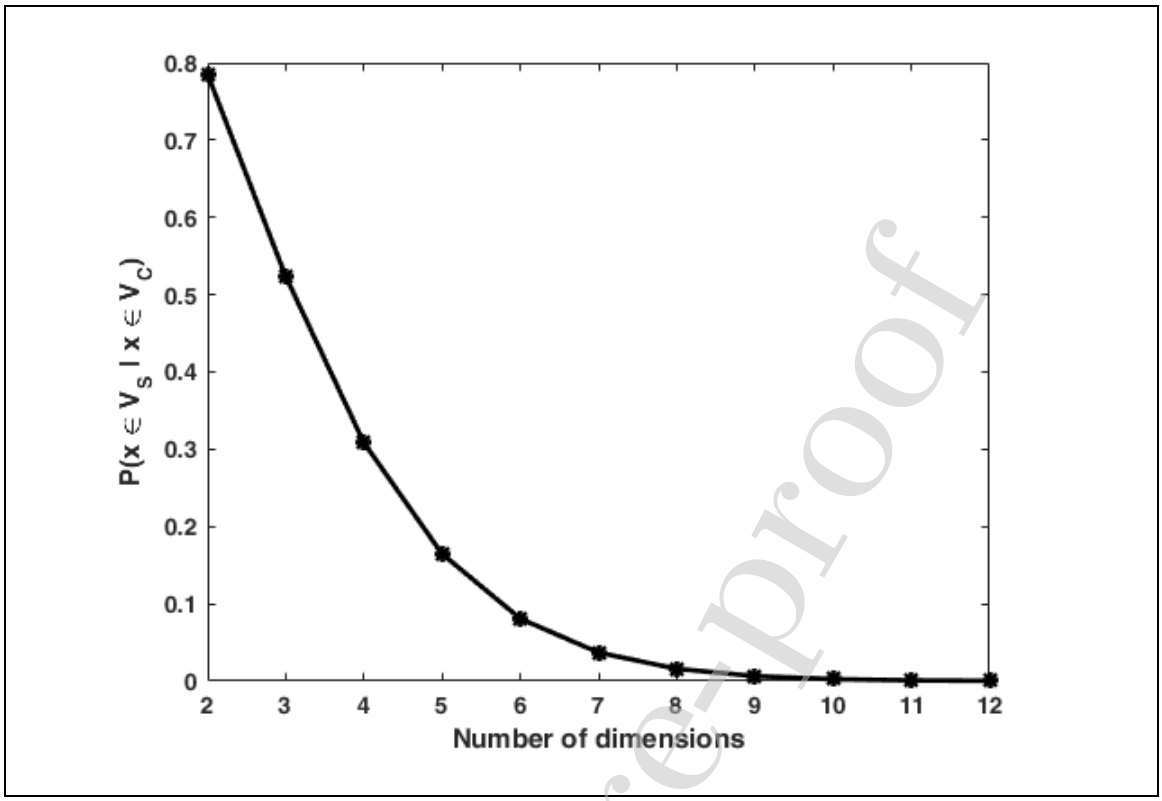
REFERENCES

1. Alumbaugh D.L., 2002. Linearized and nonlinear parameter variance estimation for two-dimensional electromagnetic induction inversion, Inverse Problems, 16, 1323-1341.
2. Aster R.C., Borchers B., Thurber C.H., 2005. Parameter Estimation and Inverse Problems. Elsevier Academic Press.
3. Bayes T., 1763. An Essay towards solving a Problem in the Doctrine of Chance, Philosophical Transactions of the Royal Society of London, 53, 370–418.
4. Bellman R., 1961. Adaptive Control Processes: A Guided Tour. Princeton University Press.
5. Blakely R.J., 1995. Potential Theory in Gravity and Magnetic Applications. Cambridge University Press.
6. Caté A., Perozzi L., Gloaguen E., Blouin M., 2017. Machine learning as a tool for geologists, The Leading Edge, 36 (3), 215–219.
7. Curtis A., Lomax A., 2001. Prior information, sampling distributions and the curse of dimensionality, Geophysics 66 (2), 372–378.
8. Donoho D., 2000. High-dimensional data analysis: The curses and blessings of dimensionality, Conf. American Math. Soc., Los Angeles, Aug. 6-11, 2000.
9. Fernández Álvarez J.P., Fernández Martínez J.L., Menéndez Pérez C.O., 2008. Feasibility analysis of the use of binary genetic algorithms as importance samplers. Application to a geoelectrical VES inverse problem, Mathematical Geosciences, 40, 375-408.
10. Fernández-Martínez J.L., Fernández-Muñiz M.Z., Tompkins M.J., 2012. On the topography of the cost functional in linear and nonlinear inverse problems, Geophysics 77 (1), W1–W15.
11. Fernández-Martínez J.L, Fernández-Muñiz Z., Pallero J.L.G., Pedruelo-González L.M., 2013. From Bayes to Tarantola: New insights to understand uncertainty in inverse problems, Journal of Applied Geophysics, 98, 62-72.

12. Fernández-Martínez J.L, Pallero J.L.G, Fernández-Muñiz Z., Pedruelo-González L.M., 2014a. The effect of the noise and Tikhonov's regularization in inverse problems. Part I: the linear case, Journal of Applied Geophysics, 108, 176-185.
13. Fernández-Martínez J.L, Pallero J.L.G, Fernández-Muñiz Z., Pedruelo-González L.M., 2014b. The effect of the noise and Tikhonov's regularization in inverse problems. Part II: the nonlinear case, Journal of Applied Geophysics, 108, 186-193.
14. Fernández-Martínez J.L., 2015. Model reduction and uncertainty analysis in inverse problems, The Leading Edge 34 (9), 1006-1016.
15. Fernández-Martínez J.L., Xu S., Sirieix C., Fernández-Muñiz Z., Riss J., 2017. Uncertainty analysis and probabilistic segmentation of electrical resistivity images: the 2D inverse problem, Geophysical Prospecting, 65 (S1), 112-130.
16. Fernández-Muñiz, Z., Fernández-Martínez J.L., Srinivasan S., Mukerji T., 2015. Comparative analysis of the solution of linear continuous inverse problems using different basis expansions, Journal of Applied Geophysics 113, 92-102.
17. Hansen P.C., 2010. Discrete Inverse Problems. Insight and Algorithms, SIAM. ISBN 978-0-898716-96-2.
18. Haario H., Saksman E., Tamminen J., 2001. An adaptive Metropolis algorithm, Bernoulli, 7 (2), 223–242.
19. Johnstone I., 1998. Oracle Inequalities and Nonparametric Functional Estimation, Documenta Mathematica ICM 1998 III, 267-278.
20. Johnstone I., 2000. On the distribution of the largest principal component. Technical Report, Department of Statistics, Stanford University. http://www-stat.stanford.edu/~imj/Reports/2000/largepc.ps
21. Menke W., 1984. Geophysical Data Analysis: Discrete Inverse Theory, Academic Press, San Diego.
22. Mosegaard K., Tarantola A., 1995. Monte Carlo Sampling of solutions to inverse problems, Journal of Geophysical Research, 100, B7, 12431–12447.
23. Rappel H., Beex L.A.A., Hale J.S., Bordas S.P.A., 2018. Bayesian inference to identify parameters in viscoelasticity, Mechanics of Time-Dependent Materials, 22 (2), 221-258.
24. Rappel H., Beex L.A.A., Hale J.S., Noels L., Bordas S.P.A., 2019. A tutorial o Bayesian inference to identify material parameters in solid mechanics, Archives of Computational Methods in Engineering, 1-25.
25. Sambridge M., 1999. Geophysical inversion with a neighborhood algorithm-II: appraising the ensemble, Geophysical Journal International, 138, 727-746.
26. Sambridge M., Mosegaard M., 2002. Monte Carlo methods in geophysical inverse problems, Reviews of Geophysics, 40, 3, 1–29.
27. Scales J.A., Snieder R., 1997. To Bayes or not to Bayes, Geophysics, 62, 4, 1045-1046.
28. Scales J.A., Snieder R., 2000. The anatomy of inverse problems, Geophysics, 65, 6, 1708-1710.
29. Scales J.A., Tenorio L., 2001. Prior information and uncertainty in inverse problems, Geophysics, 66 (2), 389-397.

30. Snieder R., 1998. The role of nonlinearity in inverse problems, Inverse Problems, 14, 387-404.
31. Snieder R., Trampert J., 1999. Inverse problems in geophysics, in Wavefield Inversion, Ed. A. Wirgin, Springer Verlag, New York, p. 119-190.
32. Scales J.A., Snieder R., 2000. The Anatomy of Inverse Problems, Geophysics 65, 1708-1710.
33. Tarantola A., 2006. Popper, Bayes and the inverse problem, Nature Physics 2, 492–494.
34. Tompkins M.J., Fernández-Martínez J.L., Alumbaugh D.L., Mukerji T., 2011a. Scalable Uncertainty Estimation for Nonlinear Inverse Problems Using Parameter Reduction, Constraint Mapping, and Geometric Sampling: Marine CSEM Examples, Geophysics, 76(4), F263-F281. doi: 10.1190/1.3581355.
35. Tompkins M.J., Fernández-Martínez J.L., Fernández Muñiz Z., 2011b. Marine electromagnetic inverse solution appraisal and uncertainty using model-derived basis functions and sparse geometric sampling, Geophysical Prospecting, 59 (5), 947-965. doi: 10.1111/j.1365-2478.2011.00955.x.
36. Tompkins M.J., Fernández-Martínez J.L., Fernández Muñiz Z., 2013. Comparison of sparse grid geometric and random sampling methods in nonlinear inverse solution uncertainty estimation, Geophysical Prospecting, 61(1), 28-41.
37. Wilson A.J., 2010. Volume of n-dimensional ellipsoid, Sciencia Acta Xaveriana, 1, 1, 101-106.

549 **LIST OF CAPTIONS**
550

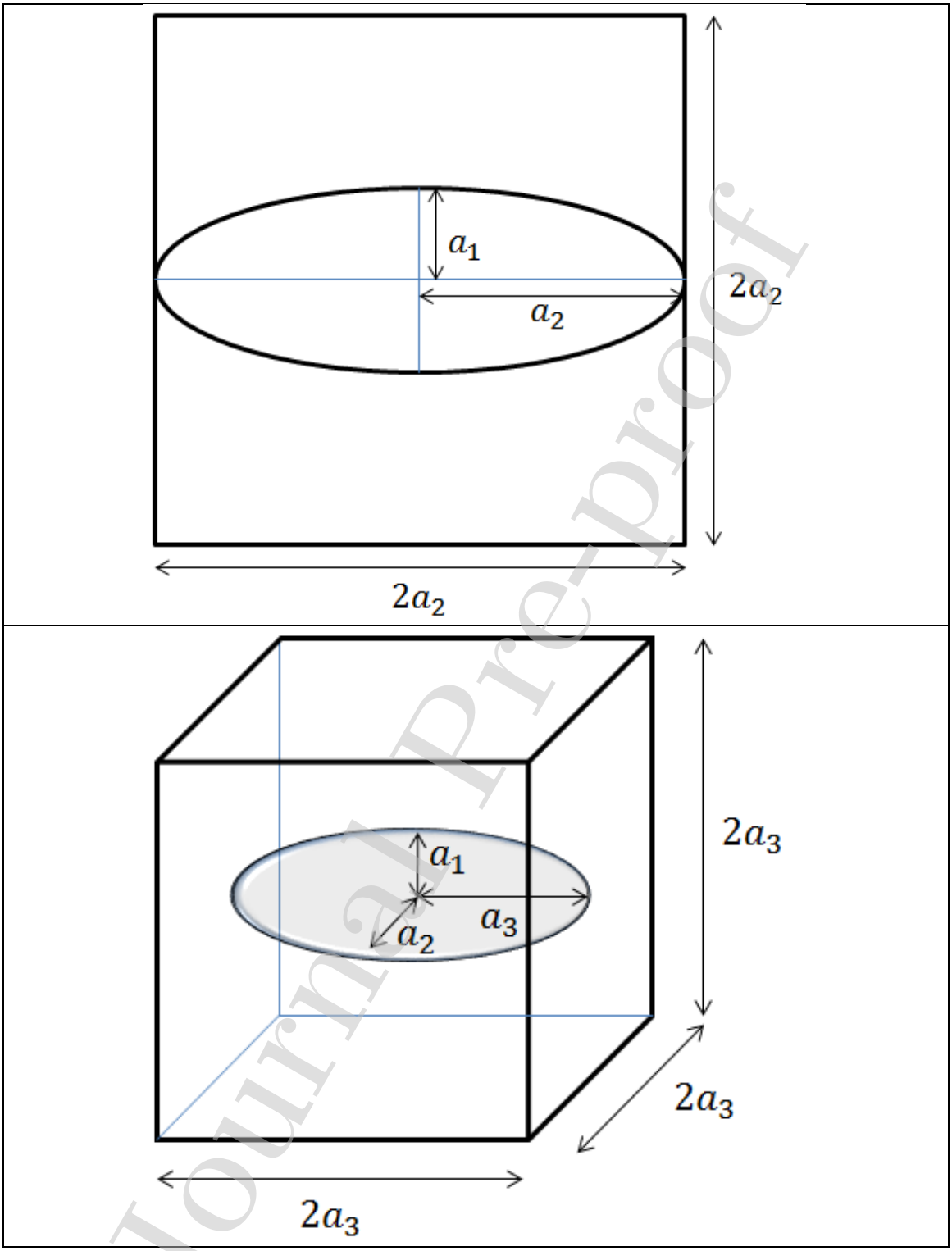


551

552 Figure 1: Isotropic sampling. The graph shows the conditional probability of sampling
553 within a hypersphere inscribed into a hypercube.

554

555



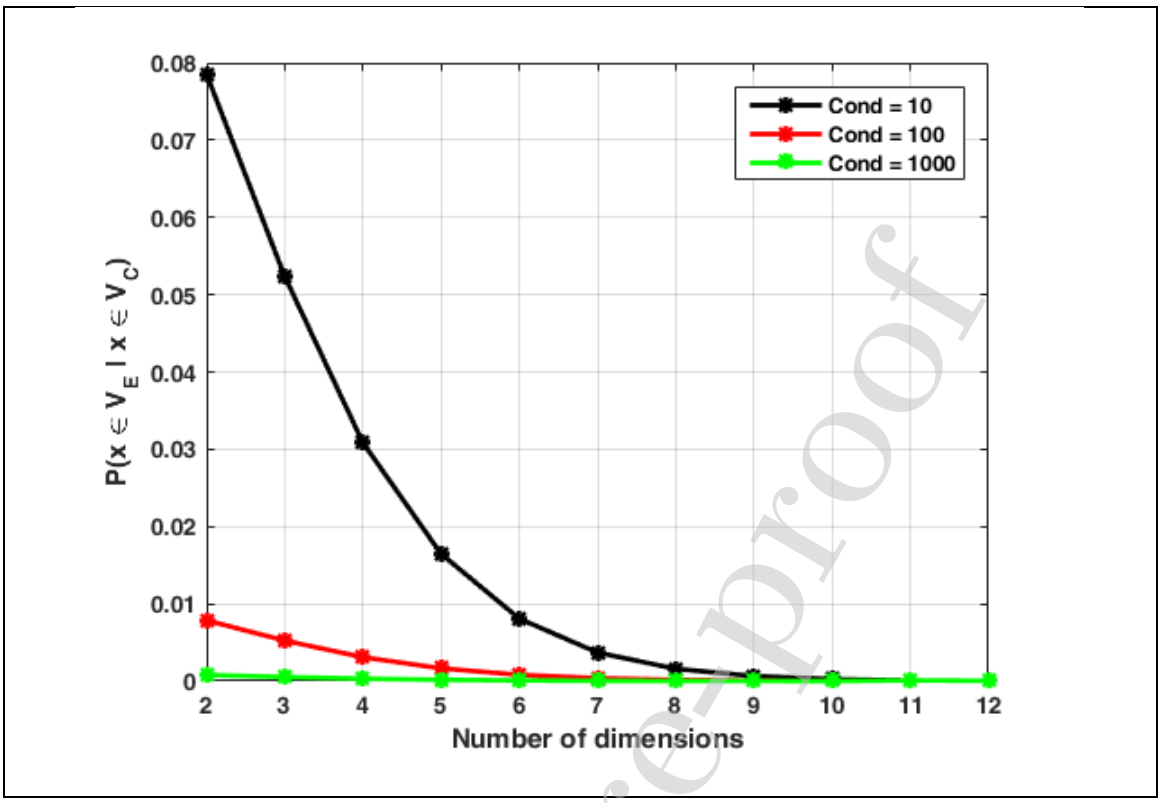
556

557 Figure 2: Anisotropic sampling in 2 and 3 dimensions. In the mathematical deduction,
558 the length of the hypercube coincides with the axis of maximum uncertainty.

559

560

561



562 Figure 3: Anisotropic sampling: conditional probability as a function of the number of
563 dimensions for different condition numbers of the system matrix. The low probabilities
564 indicate that no more than 5 to 9 dimensions can be efficiently sampled depending on
565 the condition number. For high condition numbers the sampling probability is close to
566 zero.

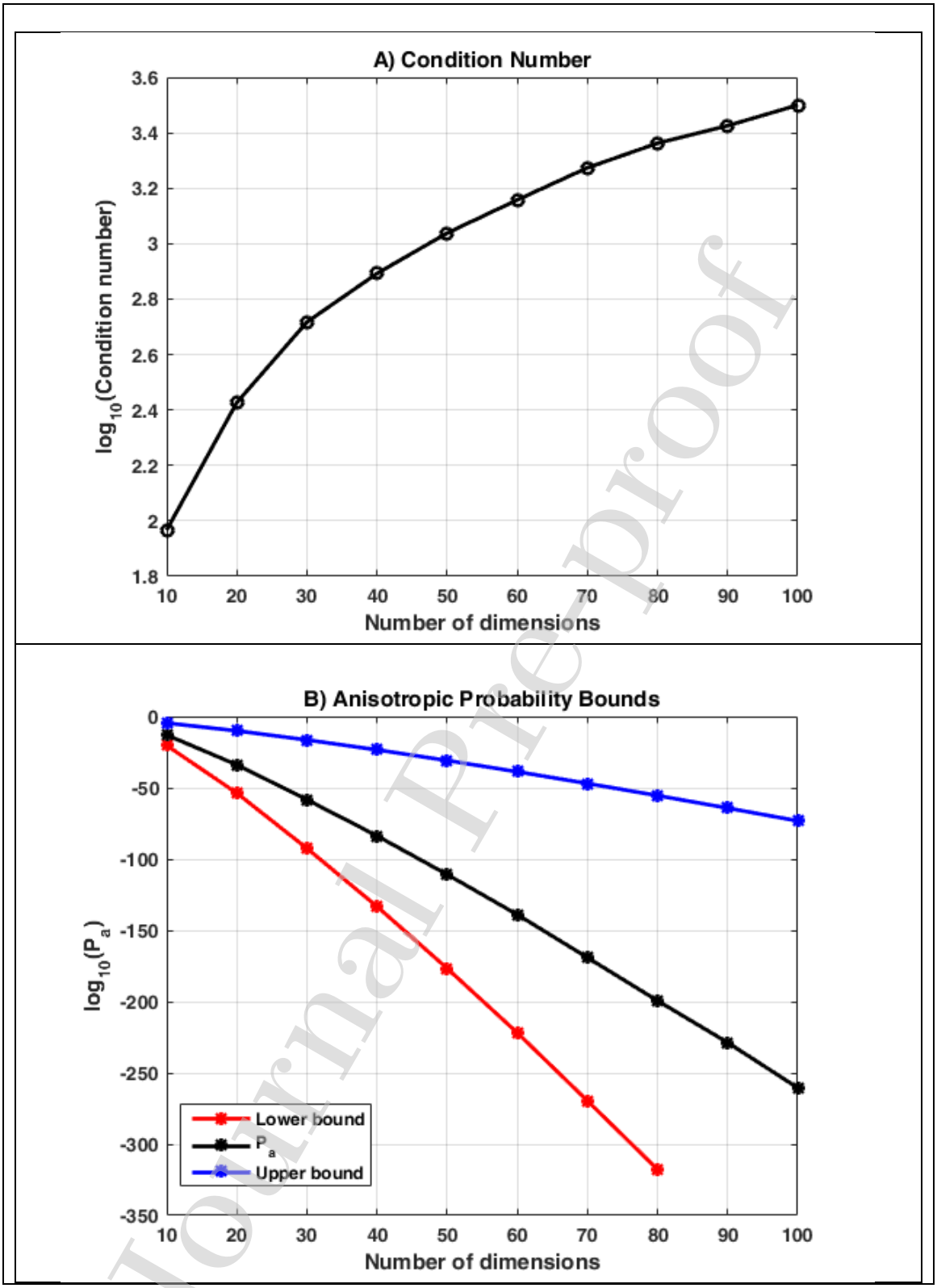


Figure 4: Anisotropic sampling probabilities (in log10 scale) of the region of linear equivalence for random matrices with dimensions between 10 and 100. A) Condition number. B) Anisotropic sampling probabilities and lower and upper bounds.

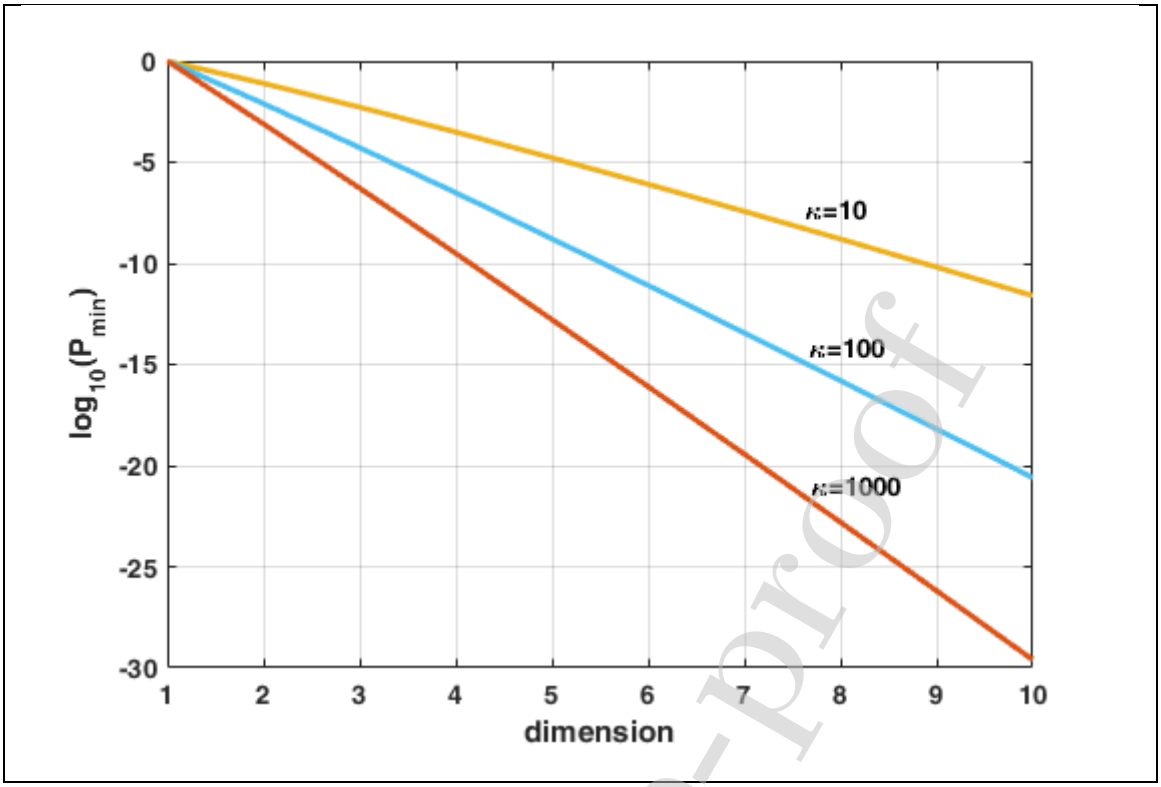


Figure 5: Variation of the values of the lower bound of the sampling probability (in logarithmic scale) with the number of dimensions (from 1 to 10), for different values of the condition number.

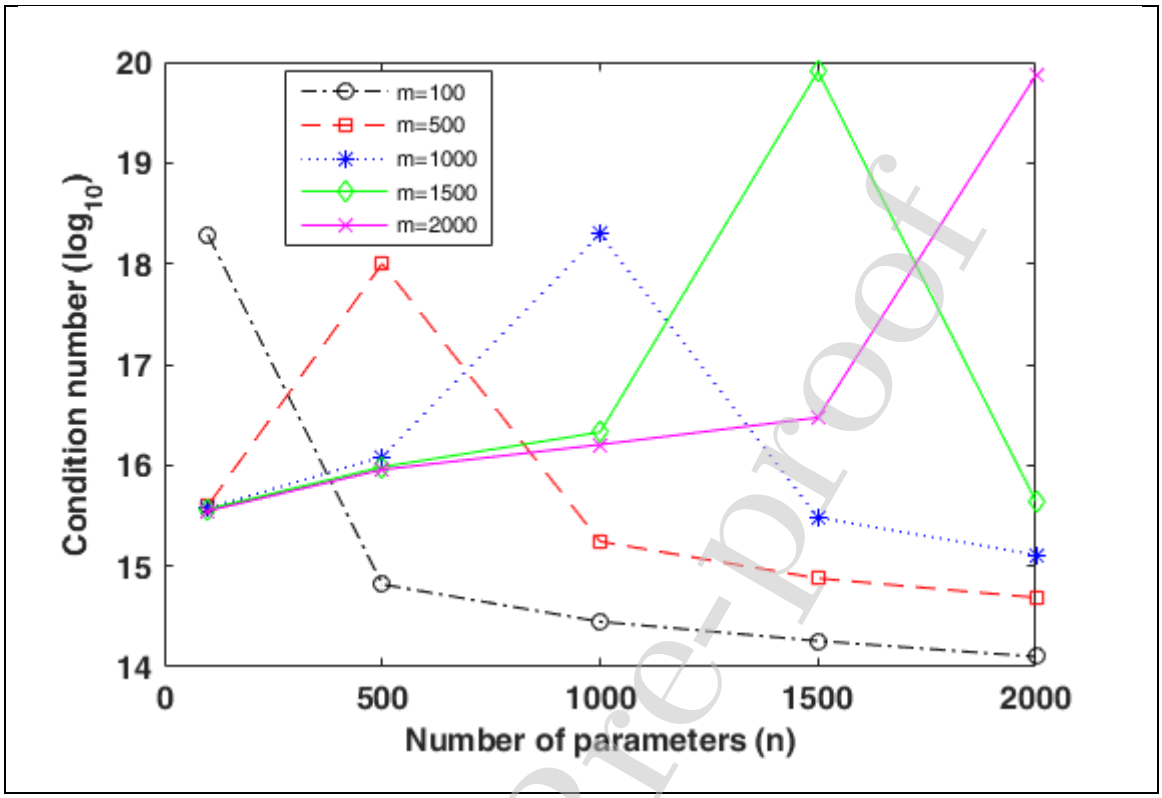


Figure 6: Condition number in logarithmic scales for different sizes of the discrete inverse gravimetric problem matrix. It can be observed that independent of its size the gravimetric problem is very ill-conditioned with condition numbers between 10^{14} and 10^{20}. In the figure m represents the number of data and n the number of parameters of the different matrices.

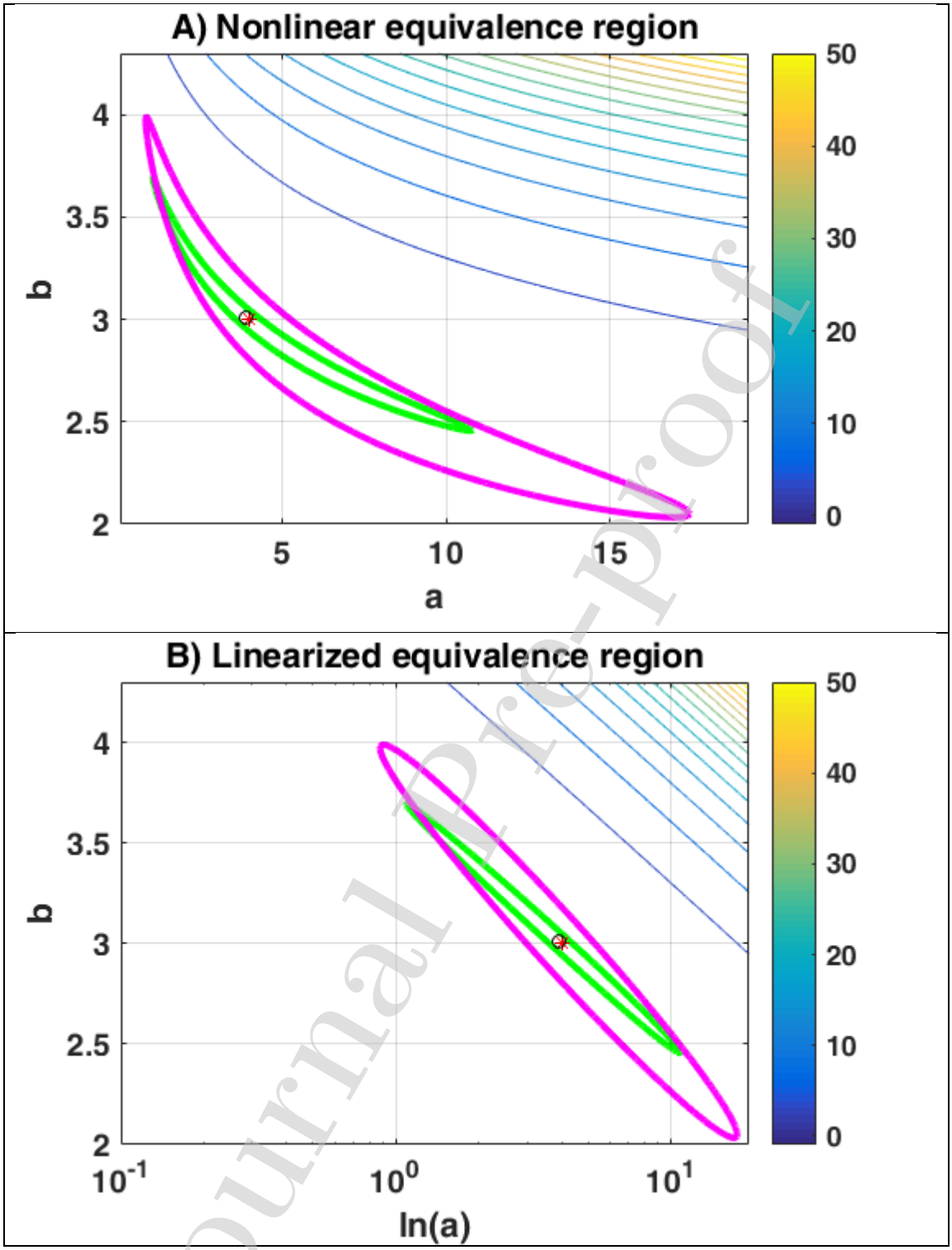


Figure 7: Nonlinear regression problem. A) Nonlinear equivalence regions for 5% (green curve) and 10% of relative error (pink curve). B) Linearized equivalence regions in the $(\ln a, b)$ plane, obtained by logarithmic reparameterization.

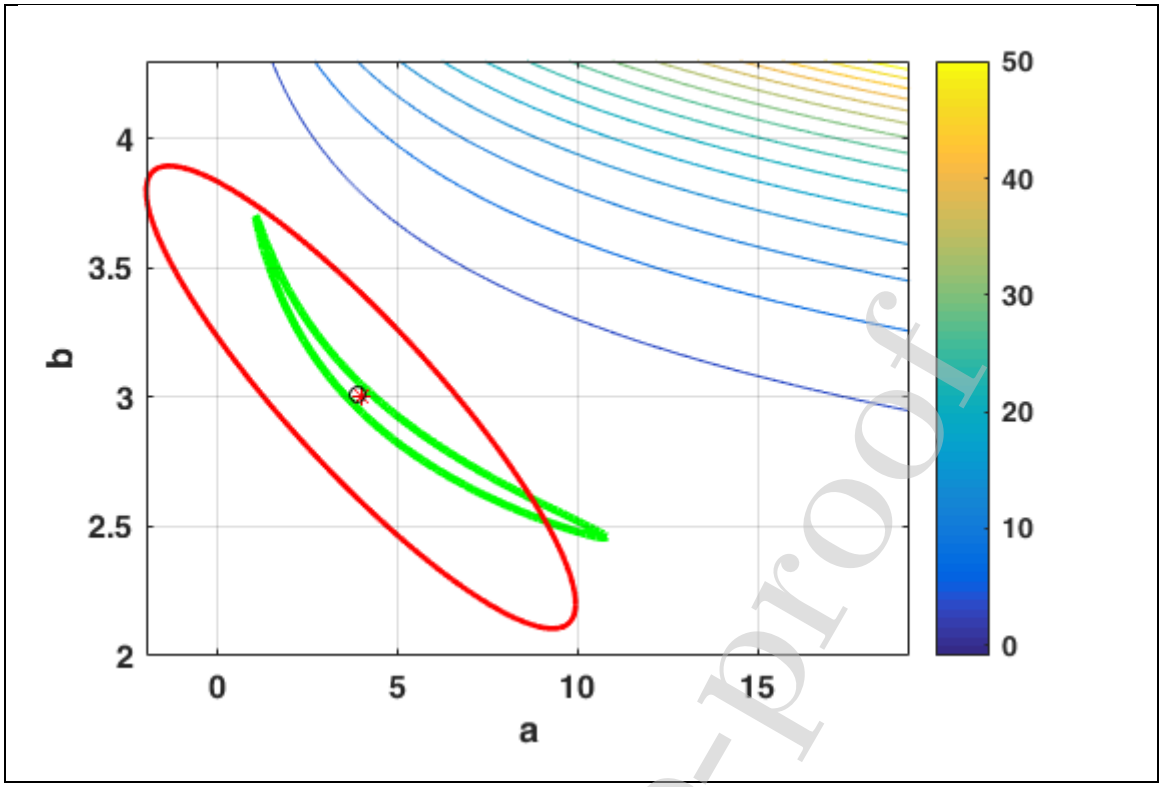


Figure 8: Nonlinear equivalent region (green line) of 10% and linearized linear region (red line) with $e_1 = 0.4, \quad e_2 = 6$.

Appendix A

Algorithm 1 Linear Problem

1. Select a matrix condition number: $[\kappa_1, \kappa_2, \ldots, \kappa_n]$
2. Select a number of matrix dimensions: $[n_1, n_2, \ldots, n_r]$
3. Choose the number of the matrices considered for each number selected in 2: N
4. **for** $i = [n_1, n_2, \ldots, n_r]$ **do**
5. $den = 2^{i-1} i \Gamma(i/2)$;
6. $P_i = \frac{\pi^{i/2}}{den}$;
7. **for** $j = 1:N$ **do** (for each matrix)
8. $\mathbf{matrix} = rand(i)$;
9. $s = svd(\mathbf{matrix})$;
10. $Ca = \frac{\min(s)^{i-1}}{prod(s)}$;
11. $\kappa = cond(\mathbf{matrix})$;
12. $CL = P_i \kappa^{1-i}$;
13. $CU = P_i \kappa^{-1}$;
14. $Pa = P_i Ca$;
15. **end for**
16. **end for**
17. Calculate: median (CL), median (CU), median (Pa)

In the case of nonlinear problems these calculations, can be performed in the Jacobian matrix $\mathbf{JF}(\mathbf{m_0})$ to estimate the number of reduced dimensions and the amount of the regularization that is needed to have a minimum sampling probability according to (18), taking into account that the condition number is the maximum singular value of $\mathbf{JF}(\mathbf{m_0})$ divided by ε, that is, the squared root of the damping parameter used in the zero-order Tikhonov regularization.