

An online database of benchmark problems for verification of inverse problems computer codes

César C. Pacheco

Department of Mechanical Engineering, Federal Fluminense University (PGMEC/TEM),
Niterói, Brazil.

E-mail: cesarp@id.uff.br

Abstract. Verification & Validation (V & V) is a fundamental step in the development process of computer codes, particularly when seeking to numerically evaluate physical phenomena. Multiple methodologies can be applied, one of which is the use of benchmark solutions. The field of inverse problems is no exception, for not only the forward problem must be verified and validated, but also the inversion technique (minimization methods, Bayesian inference, level-set methods, etc.). To address the latter, this work outlines an online database of benchmark inverse problems, for the purposes of verifying computer codes for the inverse problem. This online database provides test problems with suggested values for the exact parameters and regularization parameters, as well as synthetic data and their respective estimated values. Both parameter and function estimation problems are available. This database will be continuously updated over time, with new problems and/or solutions via different techniques. Users are also free to contribute with problems/solutions of their own. This data being freely and easily accessible online will be of great value for young and upcoming researchers in the field of inverse problems. The database can be accessed in <http://inversetoolkit.uff.br/>.

1. Introduction

Code development for numerical simulation is often an intricate process, its complexity requiring extensive testing in order to ensure that the sought output is obtained. The American Society of Mechanical Engineers (ASME) states that “the objective of verification is to establish numerical accuracy, independent of the physical (modeling) accuracy that is the subject of validation” [1]. In other words, to verify a computer code is to evaluate if the selected solution method is being correctly applied. This same principle applies when dealing with inverse problems, particularly with both the estimation technique and the forward problem solution requiring verification.

Multiple approaches can be taken for code verification. Using analytical and manufactured solutions are the processes of choice. However, obtaining these might prove to be cumbersome or even impossible, in cases involving complex geometries or formulations. On the other hand, benchmark solutions can be found in scientific journals and technical books for a wide range of mathematical models. These solutions increase in relevance as more and more researchers and developers verify their codes using them and corroborate its results. At some point, one might collect and compile these benchmark solutions in an organized fashion, dramatically simplifying the verification process. Such is the case in the field of optimization, where the publications from Hock *et al.* [2] and Schittkowski *et al.* [3] presented an extensive list of test optimization

problems and their solutions, to be used when developing optimization codes, using new or already established methods.

In the particular case of inverse problems, however, no similar compilation is to be found, although some benchmark problems can be found in books and publications by some authors. A remarkable example is the “Regularization Tools MatLab Package” manual [4], in which 13 test problems are presented. The user must be able to run the codes though and also have MatLab.

In this work, an online database is proposed to address this issue. Test problems will be made available, as well as solutions of their forward and inverse problems via different techniques, reference (exact) parameters/functions and synthetic/experimental data. Particular stages of the solution, e.g. selecting the regularization parameter for Tikhonov’s regularization, will also be addressed. The database will be permanently kept online and its content freely available. Users will also be able to collaborate with their own problems, results and/or links to their respective publications, helping to expand the database extensiveness. Finally, solutions of the inverse problem will be shown for different algorithms, in order to enable discussions on code performance.

2. The ITK database

The database can be accessed via the web address <http://inversetoolkit.uff.br/>, where the user can scroll through the list of available test problems, divided in two categories: parameter and function estimation problems. The former includes mathematical models with few unknowns (compared to the number of datapoints), leading to an overdetermined problem. The latter considers problems where the unknown is a continuous function, with a discrete approximation thereof to be estimated. In such cases, as it is well-known, the ill-posedness must be addressed directly, via regularization. This can be done in a multitude of ways, e.g. by Tikhonov Regularization [5] or truncating the SVD solution [6]. More than one regularization approach can be made available for each test problem.

The main page is shown in Figs. 1 and 2, the first of which shows a welcome screen. In it, basic information of the database are to be found, such as the description of ITK, as well as administrator info. This page also states clearly that the database is open and to be used without cost. Finally, instructions for referencing the database are given, should the users cite it in their future publications. This particular piece of information will be properly updated at the time of publication of this paper.

In the second figure (i.e. Fig. 2), one observes the list of test problems. At this initial stage, the database presents at least four test problems, so that both linear/nonlinear and parameter/function estimation problems are addressed. These lists will be continuously updated in the future, whether with new test problems or new solutions to the already listed ones.

Each link in Fig. 2 leads to a page similar to the one shown in Fig. 3. In this page, the user will be able to download the test problem datasheet (an excerpt of which is shown in Fig. 4), containing the problem formulation, numerical values of the model parameters, number of data points considered etc. Furthermore, raw data from the solutions are also obtainable, so that users can perform a high-quality comparison between their and the database results. These comprises exact values considered, exact mapped values, synthetic (or experimental) measurements and estimates. It bears mentioning that, although only cases with synthetic measurements are presently available, experimental data will also be considered and uploaded in the future.

At last, if desired, any user is welcome to suggest test problems and improvements to be made on the already available ones or even provide some of their own data to be made available in the database, with due credits given in the datasheets.

Welcome

Welcome to the InverseToolKit (ITK) website. ITK is an opensource database for benchmark solutions for verification of inverse problems computer codes.

If this work has been useful for your research, please consider citing the following document on future publications:

Pacheco C.C. *InverseToolKit: An online database of benchmark problems for verification of inverse problems computer codes.*

Database administration:

Prof. César C. Pacheco, D.Sc. — [ORCID](#)

Department of Mechanical Engineering, Federal Fluminense University (PGMEC/UFF)

Niterói, RJ, Brazil

Email: cesarp@id.uff.br

Figure 1: Main page of the ITK database.

LIST OF TEST PROBLEMS

Parameter Estimation

[TC001P](#) – Linear – Free motion of a particle

[TC002P](#) – Nonlinear – Calibration of models of epidemiology

Function Estimation

[TC001F](#) – Linear – 1st order Fredholm integral with Gaussian Kernel – 0th order Tikhonov Regularization

[TC002F](#) – Nonlinear – Estimation of Temperature-Dependent thermal conductivity in a Heat Conduction Problem

Figure 2: Initial list of test problems.

TC001F

TC001F is a function estimation test problem. The forward problem is a 1st order Fredholm integral with Gaussian Kernel.

[Download Problem Data](#)

[Datasheet](#)

Figure 3: Page for test case TC001F.

2.1. Multiple solution techniques

One particular feature of this database is that solutions for inverse test problems through different techniques will be available. Furthermore, the necessary intermediate steps in such techniques will also be disclosed. For example, when using Tikhonov regularization in linear function estimation problems, different methods are available for regularization parameter selection, such as Morozov criterium [7], L-curve and Generalized Cross-Validation (GCV) [8]. Parameter selection will be made with at least one of such methods, but preferably with many. On the other hand, the same problem can be solved e.g. via the Truncated SVD method [4], on which one must select the number of singular values to be considered. In similar fashion,

<p>BASIC INFORMATION Name: TC001F Description: 1D Fredholm Integral Type: Function Estimation Unknowns: 100 Data Points: 100</p>
<p>FORWARD PROBLEM Problem Type: Linear Mathematical Model:</p> $y(s) = \int_0^1 K(s,t) f(t) dt.$ <p>Kernel type: Gaussian with $\gamma = 0.5$.</p> $K(s,t) = \frac{1}{\gamma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{t-s}{\gamma}\right)^2\right] \quad (3.1)$ <p>Numerical Solution: Trapezoidal rule; Independent Parameters: $s, t \in (0; 1]$; $t_i = i\Delta t$; $\Delta t = 10^{-3}$; Exact Parameters:</p> $f(t) = \begin{cases} 1 & t_0 < t < t_1 \\ 0 & \text{elsewhere} \end{cases}, \quad \begin{matrix} t_0 = 0.4 \\ t_1 = 0.6 \end{matrix}.$ <p>Plot: Cf. Fig. 3.1</p>
<p>EXPERIMENTAL DATA Type: Synthetic; Dataset size: $N = 1000$; Noise: Zero mean Gaussian with std $\sigma_y = 5 \times 10^{-2}$; Download of Synthetic Data: "TC001F_data.dat" file with $(t_i, f_i, y_i^{exact}, y_i)$. Plot: Cf. Fig. 3.2</p>

Figure 4: Excerpt of the datasheet of test problem TC001F.

nonlinear problems can be solved using classical techniques such as the Levenberg-Marquardt method [9, 10, 11], the Box-Kanemasu method [12] or many other techniques available in the literature.

Furthermore, solutions within different approaches to inverse problems can be considered. In particular, the Bayesian approach [13] will be used to provide extra solutions for the test problems (or even have test problems of their own). For example, Markov-Chain Monte Carlo (MCMC) solutions will be accompanied of information on the sampling technique (e.g. Metropolis-Hastings or Gibbs sampler), transition probability density functions, total number of chains, burn-in period etc. Filtering is also a important component within the Bayesian approach, comprising Kalman (KF) [14] and particle filters [15]. In the former, not only the classical filter will be featured, but also its extensions, i.e. Extended and Unscented KF [16]. Finally, it bears mentioning that further different approaches (e.g. level-set methods) are also valuable contributions and can also be added to the list of benchmark solutions.

3. The test problems

As mentioned above, the list of test problems will be increasing permanently, with new test problems and new solutions being included in the database. Nevertheless, two test problems will be briefly discussed here in order to provide a proper illustration of the contents of ITK. One parameter and one function estimation problems were selected for this purpose.

3.1. TC001P (Parameter estimation): Ballistic trajectory

This test problem was originally proposed and presented in [17]. A particle is thrown up, moving in the vertical only. From classical mechanics, the height y varies along time t according to Eq. (1), when air resistance is to be neglected. Thus, in this parameter estimation problem, the observations are given by the discrete set $\mathbf{y} = [y_1 \dots y_N]^T$, where $y_i = y(t_i)$ and $t_i = i\Delta t, i = 1, \dots, N$, with $t \in (0, 10]$, $\Delta t = 0.1$ and $N = 100$. Synthetic measurements were simulated by adding Gaussian noise to the reference values of y_i , with zero mean and standard deviation $\sigma = 1$. The unknowns are included in the vector of unknowns $\mathbf{x} = [p_1 \ p_2 \ p_3]^T$.

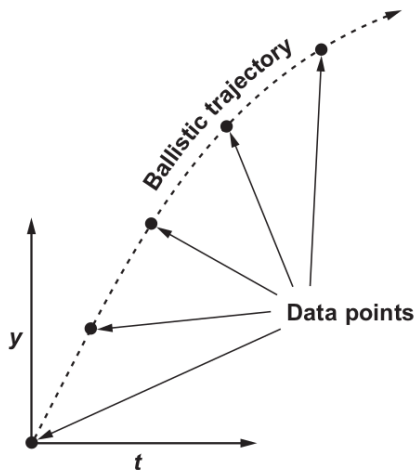


Figure 5: Illustration of the ballistic problem. Source: Aster *et al.* [17].

$$y(t) = p_1 + p_2 t - \frac{1}{2} p_3 t^2. \quad (1)$$

Due to the small number of parameters compared to the dataset size ($N = 100$), the ill-posedness of the inverse problem is not critical and the parameter vector can be estimated straight from the pseudoinverse, without any type of regularization. The linear system of equations is thus given by Eq. (2). The estimated parameters are presented in Tab. 1, together with the relative errors from the exact values. The $y(t)$ curves for the exact and estimated parameters are presented in Fig. 6, together with the residuals.

$$\begin{bmatrix} 1 & t_1 & -\frac{1}{2}t_1^2 \\ 1 & t_2 & -\frac{1}{2}t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_N & -\frac{1}{2}t_N^2 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \quad (2)$$

Table 1: Reference and estimated parameters for test problem TC001P.

	p_1	p_2	p_3
Exact	10	50	9.81
Estimated	10.472934476505358	49.796086097300176	9.7666818624341243
Rel. Error	4.7×10^{-2}	4.1×10^{-3}	4.4×10^{-3}

BASIC INFORMATION
Name: TC001P
Type: Parameter Estimation
Unknowns: 3
Data Points: 100

FORWARD PROBLEM
Problem Type: Linear
Mathematical Model:

$$y(t) = x_1 + x_2 t - \frac{1}{2} x_3 t^2.$$

Independent Parameters: $t \in (0; 10]$; $t_i = i\Delta t$; $\Delta t = 0.1$;

Exact Parameters:

$$\mathbf{x} = [10 \quad 50 \quad 9.81]^T.$$

INVERSE PROBLEM

LS Estimate for $\sigma_y = 1$:

$$\hat{\mathbf{x}} = [10.472934476505358 \quad 49.796086097300176 \quad 9.7666818624341243]^T.$$

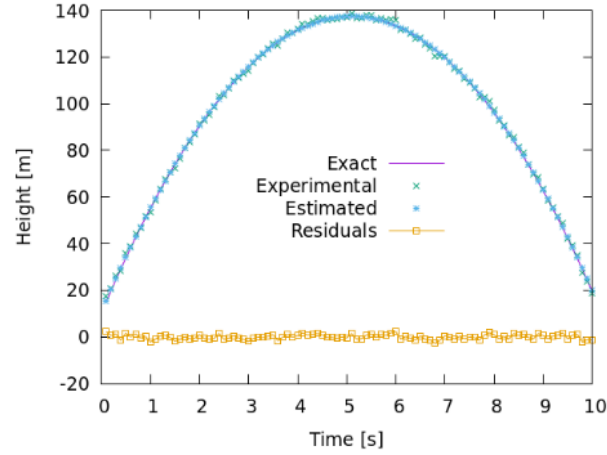


Figure 6: Excerpt of the datasheet of test problem TC001P, with exact, noisy mapping and estimated values for y . Residuals are also shown, in yellow

3.2. TC001F (Function estimation): 1st kind Fredholm Integral with Gaussian Kernel

The Fredholm integral of the 1st kind is an extremely versatile mathematical model, with applications in geophysics, tomography, image processing, heat/mass transfer, among others [17]. It is a continuous linear model, given by Eq. (3), to be discretized via numerical integration techniques, herein taken to be the Trapezoidal rule. The exact profile for $f(t)$ is given by Eq. (4) (with $t_0 = 0.4$ and $t_1 = 0.6$) and shown in Fig. 7a. In this test case, a Gaussian kernel is assumed (cf. Eq. (5)) with $K_0 = (\gamma\sqrt{2\pi})^{-1}$ and $\gamma = 0.1$. The independent variables $t, s \in [0, 1]$ are discretized in an uniform grid with $t_i = i\Delta t$, $s_i = i\Delta s$, $i = 1, \dots, N$, respectively.

Synthetic measurements were simulated by mapping the exact profile for $f(t)$ and adding Gaussian noise with zero mean and a standard deviation of $\sigma = 5 \times 10^{-2}$. These values are shown in Fig. 7b and are used in order to reconstruct the reference profile for $f(t)$. Selection of the regularization parameter is done through the L-curve, shown in Fig. 8.

$$y(s) = \int_0^1 K(s, t) f(t) dt; \quad (3)$$

$$f(t) = \begin{cases} 1, & t_0 < t < t_1 ; \\ 0, & \text{elsewhere} \end{cases} ; \quad (4)$$

$$K(s, t) = K_0 \exp \left[-\frac{1}{2} \left(\frac{t-s}{\gamma} \right)^2 \right]. \quad (5)$$

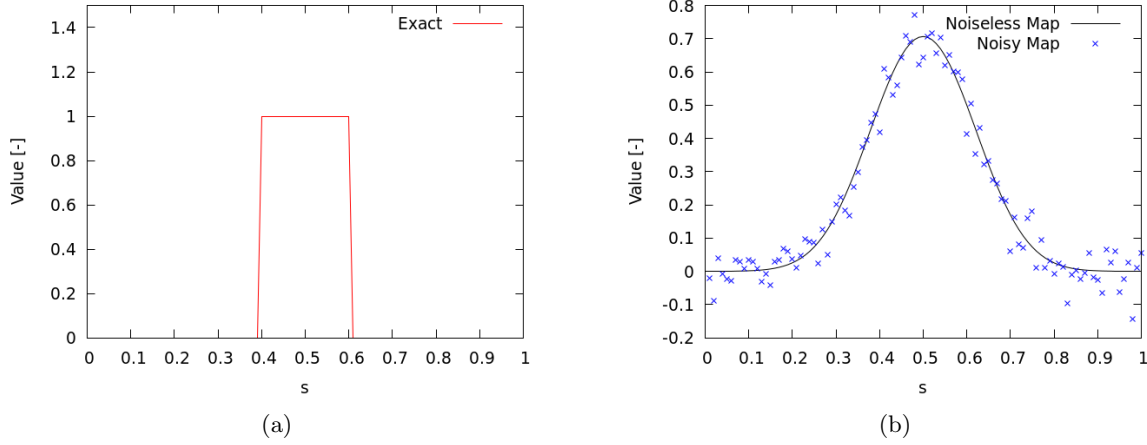


Figure 7: Reference values for test problem TC001F: (a) exact values of $f(t)$, given by Eq. 3; and (b) values of $y(s)$: with and without noise.

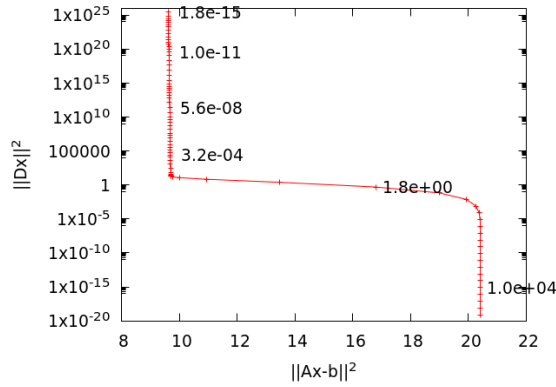


Figure 8: L-Curve for test problem TC001F.

Solving the regularization problems yields the profile shown in Fig. 9a, where a comparison with the exact values and the respective estimation errors are also present. The comparison of the synthetic values and the mapping of the estimated values for $f(t)$ are shown in 9b. The residuals are also shown, oscillating around zero and with the same order of magnitude of σ .

4. Conclusions

The ITK database brings a free, online and comprehensive collection of test inverse problems and their solutions. This information provides a positive and significant impact in the code

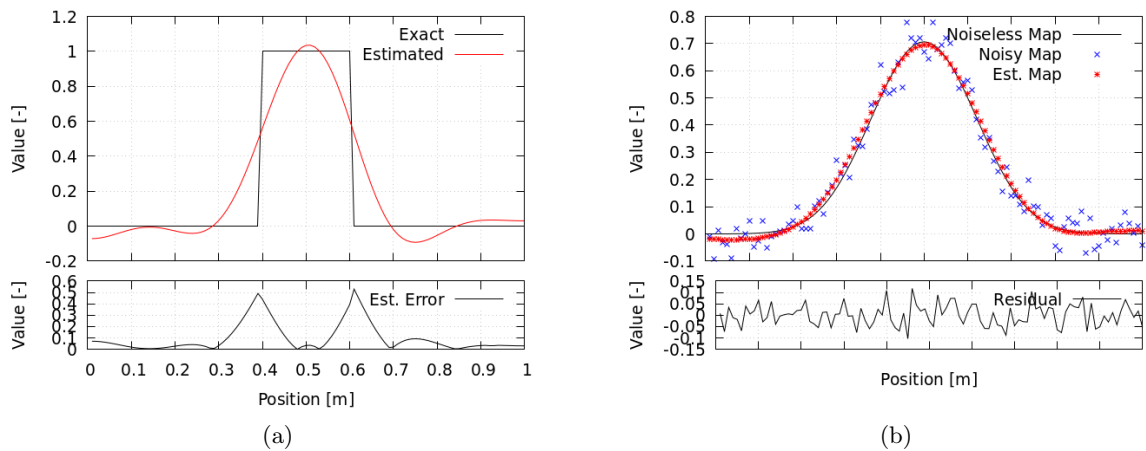


Figure 9: Results presented in the database for TC001F: (a) exact and estimated values of $f(t)$; and (b) reconstructed mapping and synthetic measurements.

verification process. Furthermore, it can be a first step towards standardization of the verification process for the specific case of inverse problems.

The improvement of the ITK database will be continuous, with new test problems and solutions being included over time, yielding

At some point, test inverse problems containing modelling errors will also be included, in order to provide test problems with no inverse crime [13]. These will be brought up by means of experimental data or purely numerical problems with different models and grid sizes used to generate the synthetic data and solve the inverse problems.

References

- [1] ASME 2009 *Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer* (New York: Standard No. ASME V&V 20-2009)
- [2] Hock W and Schittkowski K 1981 *Test Examples for Nonlinear Programming Codes* Lecture Notes in Economics and Mathematical Systems (Springer-Verlag Berlin Heidelberg) ISBN 978-3-540-10561-9
- [3] Schittkowski K 1987 *More Test Examples for Nonlinear Programming Codes* Lecture Notes in Economics and Mathematical Systems (Springer-Verlag Berlin Heidelberg) ISBN 978-3-540-17182-9
- [4] Hansen C 1994 Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems *Numerical Algorithms* **6** 1–35 ISSN 1017-1398, 1572-9265
- [5] Tikhonov A N and Arsenin V Y 1977 *Solution of Ill-Posed Problems* (V. H. Winston & Sons)
- [6] Hansen P C 1987 The truncatedSVD as a method for regularization *BIT Numerical Mathematics* **27** 534–553 ISSN 1572-9125
- [7] Morozov V 1966 Regularization of incorrectly posed problems and the choice of regularization parameter *USSR Computational Mathematics and Mathematical Physics* **6** 242–251 ISSN 0041-5553
- [8] Hansen C 1998 *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* (Society for Industrial and Applied Mathematics.)
- [9] Levenberg K 1944 A method for the solution of certain non-linear problems in least squares *Quarterly of Applied Mathematics* **2** 164–168 ISSN 0033569X, 15524485
- [10] Marquardt D W 1963 An Algorithm for Least-Squares Estimation of Nonlinear Parameters *Journal of the Society for Industrial and Applied Mathematics* **11** 431–441 ISSN 0368-4245
- [11] Ozisik M N and Orlande H R B 2000 *Inverse Heat Transfer: Fundamentals and Applications* (Taylor & Francis)
- [12] Beck J V and Arnold K J 1977 *Parameter Estimation in Engineering and Science* (John Wiley & Sons Australia, Limited) ISBN 978-0-471-06118-2
- [13] Kaipio J P and Somersalo E 2004 *Statistical and Computational Inverse Problems* (Springer Science+Business Media, Inc) ISBN 978-0-387-22073-4

- [14] Kalman R E 1960 A new approach to linear filtering and prediction problems *Transactions of the ASME–Journal of Basic Engineering* **82** 35–45
- [15] Arulampalam M S, Maskell S, Gordon N and Clapp T 2002 A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking *IEEE Transactions on Signal Processing* **50** 174–188 ISSN 1053-587X VO - 50
- [16] Simon D 2006 *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* (John Wiley & Sons, Inc.) ISBN 978-0-471-70858-2
- [17] Aster R C, Borchers B and Thurber C H 2013 *Parameter Estimation and Inverse Problems* 2nd ed (Elsevier Inc.)