

# Regularization parameter estimation for underdetermined problems by the $\chi^2$ principle with application to $2D$ focusing gravity inversion

Saeed Vatankhah<sup>1</sup>, Rosemary A Renault<sup>2</sup> and Vahid E Ardestani,<sup>1</sup>

<sup>1</sup>Institute of Geophysics, University of Tehran, Tehran, Iran, <sup>2</sup> School of Mathematical and Statistical Sciences, Arizona State University, Tempe, USA

E-mail: svatan@ut.ac.ir, renaut@asu.edu, ebrahim@ut.ac.ir

**Abstract.** The  $\chi^2$ -principle generalizes the Morozov discrepancy principle to the augmented residual of the Tikhonov regularized least squares problem. For weighting of the data fidelity by a known Gaussian noise distribution on the measured data and, when the stabilizing, or regularization, term is considered to be weighted by unknown inverse covariance information on the model parameters, the minimum of the Tikhonov functional becomes a random variable that follows a  $\chi^2$ -distribution with  $m + p - n$  degrees of freedom for the model matrix  $G$  of size  $m \times n$ ,  $m \geq n$ , and regularizer  $L$  of size  $p \times n$ . Then, a Newton root-finding algorithm, employing the generalized singular value decomposition, or singular value decomposition when  $L = I$ , can be used to find the regularization parameter  $\alpha$ . Here the result and algorithm are extended to the underdetermined case,  $m < n$ , with  $m + p \geq n$ . Numerical results first contrast and verify the Generalized Cross Validation, Unbiased Predictive Risk Estimation and  $\chi^2$  algorithms when  $m < n$ , with regularizers  $L$  approximating zero to second order derivative approximations. The inversion of underdetermined  $2D$  focusing gravity data produces models with non-smooth properties, for which typical solvers in the field use an iterative minimum support stabilizer, with both regularizer and regularizing parameter updated each iteration. The  $\chi^2$  and unbiased predictive risk estimator of the regularization parameter are used for the first time in this context. For a simulated underdetermined data set with noise, these regularization parameter estimation methods, as well as the generalized cross validation method, are contrasted with the use of the L-curve and the Morozov Discrepancy principle. Experiments demonstrate the efficiency and robustness of the  $\chi^2$ -principle and unbiased predictive risk estimator, moreover showing that the L-curve and Morozov Discrepancy Principle are outperformed in general by the three other techniques. Furthermore, the minimum support stabilizer is of general use for the  $\chi^2$ -principle when implemented without the desirable knowledge of a mean value of the model.

AMS classification scheme numbers: 65F22, 65F10, 65R32

Submitted to: *Inverse Problems*

**Keywords:** Regularization parameter,  $\chi^2$  principle, unbiased predictive risk estimator, gravity inversion, minimum support stabilizer

## 1. Introduction

We discuss the solution of numerically ill-posed and underdetermined systems of equations,  $\mathbf{d} = G\mathbf{m}$ . Here  $G \in \mathcal{R}^{m \times n}$ , with  $m < n$ , is the matrix resulting from the discretization of a forward operator which maps from the parameter or model space to the data space, given respectively by the discretely sampled vectors  $\mathbf{m} \in \mathcal{R}^n$ , and  $\mathbf{d} \in \mathcal{R}^m$ . We assume that the measurements of the data  $\mathbf{d}$  are error-contaminated,  $\mathbf{d}_{\text{obs}} = \mathbf{d} + \mathbf{n}$  for noise vector  $\mathbf{n}$ . Such problems often arise from the discretization of a Fredholm integral equation of the first kind, with a kernel possessing an exponentially decaying spectrum that is responsible for the ill-posedness of the problem. Extensive literature on the solution of such problems is available in standard literature, e.g. [1, 3, 8, 29, 32, 34].

A well-known approach for finding an acceptable solution to the ill-posed problem is to augment the data fidelity term,  $\|W_{\mathbf{d}}(G\mathbf{m} - \mathbf{d}_{\text{obs}})\|_2^2$ , here measured in a weighted  $L_2$  norm $\ddagger$ , by a stabilizing regularization term for the model parameters,  $\|L(\mathbf{m} - \mathbf{m}_0)\|_2^2$ , yielding the Tikhonov objective function

$$P^\alpha(\mathbf{m}) := \|W_{\mathbf{d}}(G\mathbf{m} - \mathbf{d}_{\text{obs}})\|_2^2 + \alpha^2 \|L(\mathbf{m} - \mathbf{m}_0)\|_2^2. \quad (1)$$

Here  $\alpha$  is the regularization parameter which trades-off between the two terms,  $W_{\mathbf{d}}$  is a data weighting matrix,  $\mathbf{m}_0$  is a given reference vector of *a priori* information for the model  $\mathbf{m}$ , and the choice of  $L \in \mathcal{R}^{p \times n}$  impacts the basis for the solution  $\mathbf{m}$ . The Tikhonov regularized solution, dependent on  $\alpha$ , is given by

$$\mathbf{m}_{\text{Tik}}(\alpha) = \arg \min_{\mathbf{m}} \{P^\alpha(\mathbf{m})\}. \quad (2)$$

If  $C_{\mathbf{d}} = (W_{\mathbf{d}}^T W_{\mathbf{d}})^{-1}$  is the data covariance matrix, and we assume white noise for the mapped model parameters  $L\mathbf{m}$  so that the model covariance matrix is  $C_L = \sigma_L^2 I = \alpha^{-2} I = (W_L^T W_L)^{-1}$ , then (1) is

$$P^{\sigma_L}(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}_{\text{obs}}\|_{C_{\mathbf{d}}^{-1}}^2 + \|L(\mathbf{m} - \mathbf{m}_0)\|_{C_L^{-1}}^2. \quad (3)$$

Note we will use in general the notation  $\mathbf{y} \sim \mathbb{N}(\hat{\mathbf{y}}, C_{\mathbf{y}})$  to indicate that  $\mathbf{y}$  is normally distributed with mean  $\hat{\mathbf{y}}$  and symmetric positive definite (SPD) covariance matrix  $C_{\mathbf{y}}$ . Using  $C_L = \alpha^{-2} I$  in (3) permits an assumption of white noise in the estimation for  $L\mathbf{m}$  and thus statistical interpretation of the regularization parameter  $\alpha^2$  as the inverse of the white noise variance.

The determination of an optimal  $\alpha$  is a topic of much previous research, when  $m \geq n$ , and includes methods such as the L-curve (LC) [7], generalized cross validation (GCV) [4], the unbiased predictive risk estimator (UPRE) [13, 30], the residual periodogram (RP) [10, 27], and the Morozov discrepancy principle (MDP) [19], all of which are well described in the literature, see e.g. [8, 32] for comparisons of the criteria and

$\ddagger$  Here we use the standard definition for the weighted norm of the vector  $\mathbf{y}$ ,  $\|\mathbf{y}\|_W^2 := \mathbf{y}^T W \mathbf{y}$ .

further references. The motivation and assumptions for these methods varies; while the UPRE requires that statistical information on the noise in the measurement data be provided, for the MDP it is sufficient to have an estimate of the overall error level in the data. More recently, a new approach based on the  $\chi^2$  property of the functional (3) under statistical assumptions applied through  $C_L$ , was proposed by [14], for the overdetermined case  $m > n$  with  $p = n$ . The extension to the case with  $p \leq n$  and a discussion of effective numerical algorithms is given in [17, 24], with also consideration of the case when  $\hat{\mathbf{m}} = \mathbf{m}_0$  is not available. Extensions for nonlinear problems [16], inclusion of inequality constraints [18] and for multiple regularization parameters [15] have also been considered.

The fundamental premise of the  $\chi^2$  principle for estimating  $\sigma_L$  is that provided the noise distribution on the measured data is available, through knowledge of  $C_{\mathbf{d}}$ , such that the weighting on the model and measured parameters is as in (3), and that the mean value  $\hat{\mathbf{m}}$  is known, then  $P^{\sigma_L}(\mathbf{m}_{\text{Tik}}(\sigma_L))$  is a random variable following a  $\chi^2$  distribution with  $m + p - n$  degrees of freedom. Thus the expected value satisfies  $\hat{P}^{\sigma_L}(\mathbf{m}_{\text{Tik}}(\sigma_L)) = m + p - n$ . As a result  $P^{\sigma_L}(\mathbf{m}_{\text{Tik}}(\sigma_L))$  lies within an interval centered around its expected value, which facilitates the development of the Newton root-finding algorithm for the optimal  $\sigma_L$  given in [17]. This algorithm has the advantage as compared to other techniques of being very fast for finding the unique  $\sigma_L$ , provided the root exists, requiring generally no more than 10 evaluations of  $P^{\sigma_L}(\mathbf{m}_{\text{Tik}}(\sigma_L))$  to converge to a reasonable estimate. The algorithm in [17] was presented for small scale problems in which one can use the singular value decomposition (SVD) [5] for matrix  $G$  when  $L = I$  or the generalized singular value decomposition (GSVD) [20] of the matrix pair  $[W_{\mathbf{d}}G; L]$ . For the large scale case an approach using the Golub-Kahan iterative bidiagonalization based on the LSQR algorithm [21, 22] was presented in [24] along with the extension of the algorithm for the non-central distribution of  $P^{\sigma_L}(\mathbf{m}_{\text{Tik}}(\sigma_L))$ , namely when  $\mathbf{m}_0$  is unknown but may be estimated from a set of measurements. In this paper, the  $\chi^2$  principle is first extended to the estimation of  $\sigma_L$  for underdetermined problems, specifically for the central  $\chi^2$  distribution with known  $\mathbf{m}_0$ , with the proof of the result in Section 2 and examples in Section 2.4.

In many cases the smoothing that arises when the basis mapping operator  $L$  approximates low order derivatives is unsuitable for handling material properties that vary over relatively short distances, such as in the inversion of gravity data produced by localized sources, [11, 23, 31, 34]. A stabilizer that does not penalize sharp boundaries is instead preferable. This can be achieved by setting the regularization term in a different norm, as for example using the total variation, [26], already significantly studied and applied for geophysical inversion e.g. [25, 28]. Similarly, the minimum support (MS) and minimum gradient support (MGS) stabilizers provide solutions with non-smooth properties and were introduced for geophysical inversion in [23] and [34], respectively. We note now the relationship of the iterated MS stabilizer with non stationary iterated Tikhonov regularization [6], in which the solution is iterated to convergence with a fixed operator  $L$ , but updated residual and iteration dependent  $\alpha^{(k)}$  which is forced to

zero geometrically. The technique was extended for example for image deblurring in [2] with  $\alpha^{(k)}$  found using a version of the MDP and dependent on a good preconditioning approximation for the square model matrix. More generally, the iterative MS stabilizers, with both  $L$  and  $\alpha$  iteration dependent, are closely related to the iteratively reweighted norm approximation for the Total Variation norm introduced and analyzed in [33]. In this paper the MS stabilizer is used to reconstruct non-smooth models for the geophysical problem of gravity data inversion with the updating regularization parameter found using the most often applied techniques of MDP and L-curve, and more recently the GCV, [31]. Results are contrasted with those obtained for the first time in this framework using the UPRE, and the  $\chi^2$  principle. They also demonstrate that the  $\chi^2$  principle can be applied without knowledge of the prior  $\mathbf{m}_0$  for the MS iterative process. Initialization with  $\mathbf{m}_0 = 0$  is contrasted with an initial stabilizing choice determined by the spectrum of the operator.

The outline of this paper is as follows. In Section 2 the theoretical development of the  $\chi^2$  principle for the underdetermined problem is presented. We note that the proof is stronger than that used in the original literature when  $m \geq n$  and hence improves the general result. The algorithm uses the GSVD (SVD) at each iteration and leads to a Newton-based algorithm for estimating the regularization parameter. A review of the LC, MDP, GCV and UPRE methods for parameter estimation is presented in Section 2.3 and numerical examples contrasting these with the  $\chi^2$  approach are in Section 2.4. For completeness, all formulae which are necessary for implementation of the various regularization parameter estimates using the GSVD are provided in an Appendix. The MS stabilizer is described in Section 3 and numerical experiments contrasting the impact of the choice of the regularization parameter within the MS algorithm for the problem of 2D gravity inversion are described in Section 3.1. Conclusions and future work are discussed in Section 4.

## 2. Theoretical Development

Although the proof of the result on the degrees of freedom for the underdetermined case  $m < n$  effectively follows the ideas introduced [17, 24], the modification presented here provides a stronger result which can also strengthen the result for the overdetermined case,  $m \geq n$ .

### 2.1. $\chi^2$ distribution for the underdetermined case

We first assume that it is possible to solve the normal equations

$$(G^T W_d^T W_d G + L^T W_L^T W_L L) \mathbf{y} = G^T W_d^T W_d \mathbf{r}, \quad \mathbf{r} := \mathbf{d}_{\text{obs}} - G \mathbf{m}_0, \quad \mathbf{y} = \mathbf{m} - \mathbf{m}_0 \quad (4)$$

for the shifted system associated with (3). The **invertibility** condition for (4) requires that  $\tilde{L} := W_L L$  and  $\tilde{G} := W_d G$  have null spaces which do not intersect

$$\mathcal{N}(W_L L) \cap \mathcal{N}(W_d G) = 0. \quad (5)$$

Moreover, we also assume  $m+p \geq n$  which is realistic when  $L$  approximates a derivative operator of order  $\ell$ , then  $p = n - \ell$ , and typically  $\ell$  is small,  $n - \ell \geq 0$ .

Following [17] we first find the functional  $P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L))$  where  $\mathbf{m}_{\text{Tik}}(W_L) = \mathbf{y}(W_L) + \mathbf{m}_0$  and  $\mathbf{y}(W_L)$  solves (4), for general  $W_L$ , replacing  $W_L = \alpha^2 I$  in (1), and the argument within  $P^{W_L}$  indicates the functional evaluated for the given solution  $\mathbf{m}_{\text{Tik}}(W_L)$ . There are many definitions for the GSVD in the literature, differing with respect to the ordering of the singular decomposition terms, but all effectively equivalent to the original GSVD introduced in [20]. For ease of presentation we introduce  $\mathbf{1}_k$  and  $\mathbf{0}_k$  to be the vectors of length  $k$  with 1, respectively 0, in all rows, and define  $q = n - m \geq 0$ . We use the GSVD as stated in [1].

**Lemma 1** (GSVD). *Suppose  $H := [\tilde{G}; \tilde{L}]$ , where  $\tilde{G}$  has size  $m \times n$ ,  $m < n$ ,  $\tilde{L}$  has size  $p \times n$ ,  $p \leq n$  with both  $\tilde{G}$  and  $\tilde{L}$  of full row rank  $m$ , and  $p$ , respectively, and by (5) that  $H$  has full column rank  $n$ . The Generalized Singular Value Decomposition for  $H$  is*

$$[\tilde{G}; \tilde{L}] = [U\tilde{\Upsilon}X^T; V\tilde{M}X^T] \quad (6)$$

$$\tilde{\Upsilon} = [0_{m \times q} \quad \Upsilon], \quad \Upsilon = \text{diag}(v_{q+1}, \dots, v_p, \mathbf{1}_{n-p}) \in \mathcal{R}^{m \times m} \quad (7)$$

$$\tilde{M} = [M \quad 0_{p \times (n-p)}], \quad M = \text{diag}(\mathbf{1}_q, \mu_{q+1}, \dots, \mu_p) \in \mathcal{R}^{p \times p}, \quad (8)$$

$$0 < v_{q+1} \leq \dots \leq v_p < 1, \quad 1 > \mu_{q+1} \geq \dots \geq \mu_p > 0, \quad v_i^2 + \mu_i^2 = 1, \quad i = q+1 \dots p. \quad (9)$$

Matrices  $U \in \mathcal{R}^{m \times m}$  and  $V \in \mathcal{R}^{p \times p}$  are orthogonal,  $U^T U = I_m$ ,  $V^T V = I_p$ , and  $X \in \mathcal{R}^{n \times n}$  is invertible.

**Remark 1.** *The indexing in matrices  $M$  and  $\Upsilon$  uses the column index and we use the definitions  $\mu_i = 0$ ,  $i = p+1 : n$  and  $v_i = 0$ ,  $i = 1 : q$ . Moreover, by (7)-(8), we also define  $v_i = 1$ ,  $i = p+1 : n$ , and  $\mu_i = 1$ ,  $i = 1 : q$ . The generalized singular values are given by  $\gamma_i = v_i/\mu_i$ ,  $i = 1 : n$ . Of these  $n-p$  are infinite,  $m+p-n$  are finite and non-zero, and  $q$  are zero.*

We first introduce  $\tilde{\mathbf{r}} := W_{\mathbf{d}} \mathbf{r}$  and note the relations

$$\begin{aligned} \tilde{\Upsilon}^T \tilde{\Upsilon} + \tilde{M}^T \tilde{M} &= I_n, & \tilde{G}^T \tilde{G} + \tilde{L}^T \tilde{L} &= X X^T, & \tilde{\Upsilon} \tilde{\Upsilon}^T &= \Upsilon \Upsilon^T, & \text{and} \\ \mathbf{y} &= (X^T)^{-1} \tilde{\Upsilon}^T U^T \tilde{\mathbf{r}}, & \tilde{G} \mathbf{y} &= U \tilde{\Upsilon} \tilde{\Upsilon}^T U^T \tilde{\mathbf{r}}, & \tilde{L} \mathbf{y} &= V \tilde{M} \tilde{\Upsilon}^T U^T \tilde{\mathbf{r}}. \end{aligned}$$

Thus with  $\mathbf{s} = U^T \tilde{\mathbf{r}}$ , with indexing from  $q+1 : n$  for  $\mathbf{s}$  of length  $m$ ,  $s_i = \mathbf{u}_{i-q}^T \tilde{\mathbf{r}}$ ,

$$\begin{aligned} P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L)) &= \tilde{\mathbf{r}}^T U (I_m - \tilde{\Upsilon} \tilde{\Upsilon}^T) U^T \tilde{\mathbf{r}} = \sum_{i=q+1}^p \mu_i^2 s_i^2 = \|\mathbf{k}\|_2^2, \\ \mathbf{k} &= Q U^T W_{\mathbf{d}} \mathbf{r}, \quad Q := \text{diag}(\mu_{q+1}, \dots, \mu_p, \mathbf{0}_{n-p}). \end{aligned} \quad (10)$$

To obtain our desired result on  $P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L))$  as a random variable we investigate the statistical distribution of the components for  $\mathbf{k}$ , following [17, Theorem 3.1] and [24, Theorem 1] for the cases of a central, and non-central distribution, respectively, but with modified assumptions that lead to a stronger result.

**Theorem 2.1** (central and non-central  $\chi^2$  distribution of :  $P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L))$ ). Suppose  $\mathbf{n} \sim \mathbb{N}(0, C_d)$ ,  $\boldsymbol{\zeta} := (\mathbf{m} - \mathbf{m}_0) \sim \mathbb{N}(\hat{\mathbf{m}} - \mathbf{m}_0, C_m)$ ,  $L\boldsymbol{\zeta} := L(\mathbf{m} - \mathbf{m}_0) \sim \mathbb{N}(L(\hat{\mathbf{m}} - \mathbf{m}_0), C_L)$ , the invertibility condition (5), and that  $m + p - n > 0$  is sufficiently large that limiting distributions for the  $\chi^2$  result hold. Then for

(i)  $\mathbf{m}_0 = \hat{\mathbf{m}}$ :  $P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L)) \sim \chi^2(I_{m+p-n}, 0)$ .

(ii)  $\mathbf{m}_0 \neq \hat{\mathbf{m}}$ :  $P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L)) \sim \chi^2(I_{m+p-n}, c)$ ,  $c = \|QU^T W_d G(\hat{\mathbf{m}} - \mathbf{m}_0)\|_2^2 := \|\mathbf{c}\|_2^2$ .

Equivalently the minimum value of the functional  $P^{W_L}(\mathbf{m}_{\text{Tik}}(W_L))$  is a random variable which follows a  $\chi^2$  distribution with  $m + p - n$  degrees of freedom and non-centrality parameter  $c = \|QU^T W_d G(\hat{\mathbf{m}} - \mathbf{m}_0)\|_2^2$ .

*Proof.* By (10) it is sufficient to examine the components  $k_i$ ,  $i = q + 1, \dots, p$  to demonstrate  $\|\mathbf{k}\|^2$  is a sum of normally distributed components with mean  $c$  and then employ the limiting argument to yield the  $\chi^2$  distribution, as in [17, 24]. First observe that  $\mathbf{d} \sim \mathbb{N}(G\hat{\mathbf{m}}, GC_m G^T)$ , thus  $\mathbf{r} = \mathbf{d}_{\text{obs}} - G\mathbf{m}_0 = \mathbf{d} + \mathbf{n} - G\mathbf{m}_0 = G(\mathbf{m} - \mathbf{m}_0) + \mathbf{n} \sim \mathbb{N}(G(\hat{\mathbf{m}} - \mathbf{m}_0), C_d + GC_m G^T)$ ,  $W_d \mathbf{r} \sim \mathbb{N}(W_d G(\hat{\mathbf{m}} - \mathbf{m}_0), W_d(C_d + GC_m G^T)W_d^T)$ , and  $\mathbf{k} \sim \mathbb{N}(QU^T W_d G(\hat{\mathbf{m}} - \mathbf{m}_0), QU^T W_d(C_d + GC_m G^T)W_d^T UQ^T)$ . The result for the central parameter  $c$  is thus immediate. For the covariance we have

$$C_{\mathbf{k}} = QU^T W_d(C_d + GC_m G^T)W_d^T UQ^T = QQ^T + Q\tilde{Y}X^T C_m X\tilde{Y}^T Q^T. \quad (11)$$

By assumption,  $L$  has full row rank and  $C_m$  is SPD, thus for  $C_L := LC_m L^T$  we can define  $W_L := \sqrt{C_L^{-1}}$ . Therefore

$$I_p = W_L C_L W_L^T = W_L L C_m L^T W_L^T = \tilde{L} C_m \tilde{L}^T = V \tilde{M} X^T C_m X \tilde{M}^T V^T \quad \text{implies} \\ \tilde{M}(X^T C_m X) \tilde{M}^T = I_p. \quad (12)$$

Introduce pseudoinverses for  $\tilde{M}$  and  $\tilde{M}^T$ , denoted by superscript  $\dagger$ , and a dimensionally-consistent block decomposition for  $(X^T C_m X)$ , in which  $C_{11}$  is of size  $p \times p$ ,

$$X^T C_m X = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \quad \tilde{M}^\dagger = I_n \begin{pmatrix} M^{-1} \\ 0 \end{pmatrix} I_p, \quad (\tilde{M}^T)^\dagger = I_p \begin{pmatrix} M^{-1} & 0 \end{pmatrix} I_n. \quad (13)$$

Then applying to (12)

$$\tilde{M}^\dagger (\tilde{M}^T)^\dagger = \begin{pmatrix} M^{-2} & 0 \\ 0 & 0 \end{pmatrix} = \tilde{M}^\dagger \tilde{M} (X^T C_m X) \tilde{M}^T (\tilde{M}^T)^\dagger \quad \text{yielding} \\ \begin{pmatrix} M^{-2} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} (X^T C_m X) \begin{pmatrix} I_p & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} C_{11} & 0 \\ 0 & 0 \end{pmatrix}.$$

Moreover,

$$Q\tilde{Y} = \begin{pmatrix} M_{11} & 0 \\ 0 & 0_{n-p} \end{pmatrix} \begin{pmatrix} 0 & \Upsilon_{11} & 0 \\ 0 & 0 & I_{n-p} \end{pmatrix} = \begin{pmatrix} 0 & \Upsilon_{11} M_{11} & 0 \\ 0 & 0 & 0_{n-p} \end{pmatrix} \quad (14)$$

where  $M_{11} := \text{diag}(\mu_{q+1}, \dots, \mu_p)$ , and  $\Upsilon_{11} := \text{diag}(v_{q+1}, \dots, v_p)$ . Then with a block decomposition of  $(X^T C_{\mathbf{m}} X)$ , in which as compared to (13) now  $[C_{13}, C_{12}] := C_{12}$  and  $[C_{31}; C_{21}] = C_{21}$ , we have for (11)

$$\begin{aligned} C_{\mathbf{k}} &= Q Q^T + Q \tilde{\Upsilon} X^T C_{\mathbf{m}} X \tilde{\Upsilon}^T Q^T \\ &= \begin{pmatrix} M_{11}^2 & 0 \\ 0 & 0_{n-p} \end{pmatrix} + \begin{pmatrix} 0 & \Upsilon_{11} M_{11} & 0 \\ 0 & 0 & 0_{n-p} \end{pmatrix} \begin{pmatrix} I_{n-m} & 0 & C_{13} \\ 0 & M_{11}^{-2} & C_{12} \\ C_{31} & C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \Upsilon_{11} M_{11} & 0 \\ 0 & 0_{n-p} \end{pmatrix} \\ &= \begin{pmatrix} M_{11}^2 & 0 \\ 0 & 0_{n-p} \end{pmatrix} + \begin{pmatrix} \Upsilon_{11}^2 & 0 \\ 0 & 0_{n-p} \end{pmatrix} = \begin{pmatrix} I_{m+p-n} & 0 \\ 0 & 0_{n-p} \end{pmatrix}, \end{aligned}$$

as required to obtain the properties of the distribution for  $\|\mathbf{k}\|^2$ .  $\square$

**Remark 2.** Note that the result is exactly the same as given in the previous results for the overdetermined situation  $m \geq n$  but now for  $m < n$  with  $m + p \geq n$  and without the prior assumption on the properties for the pseudoinverse on  $C_L$ . Namely we directly use the pseudoinverse  $\tilde{M}^\dagger$  and its transpose hence, after adapting the proof for the case with  $m \geq n$ , this tightens the results previously presented in [17, 24].

**Remark 3.** We note as in [24, Theorem 2] that the theory can be extended for the case in which the filtering of the GSVD replaces uses  $v_i = 0$  for  $v_i < \tau$  for some tolerance  $\tau$ , eg suppose  $v_i < \tau$  for  $i \leq p - r$  then we have the filtered functional

$$\|\mathbf{k}(\sigma_L)\|_2^2 = \sum_{i=q+1}^{p-r} s_i^2 + \sum_{i=p-r+1}^p \frac{s_i^2}{\gamma_i^2 \sigma_L^2 + 1} := \sum_{i=q+1}^{p-r} s_i^2 + \|\mathbf{k}_{\text{FILT}}\|_2^2. \quad (15)$$

Thus we obtain  $\|\mathbf{k}(\sigma_L)_{\text{FILT}}\|_2^2 \sim \chi^2(I_r, c_{\text{FILT}})$ , where we use  $c_{\text{FILT}} = \|\mathbf{c}_{\text{FILT}}\|_2^2 = \|\tilde{I}\mathbf{c}\|_2^2$ , in which  $\tilde{I} = \text{diag}(\mathbf{0}_{m-r+p-n}, \mathbf{1}_r, \mathbf{0}_{n-p})$  picks out the filtered components only.

**Remark 4.** As already noted in the statement of Lemma 1 the GSVD is not uniquely defined with respect to ordering of the columns of the matrices. On the other hand it is not essential that the ordering be given as stated to use the iteration defined by (A.9). In particular, it is sufficient to identify the ordering of the spectra in matrices  $\Upsilon$  and  $M$ , and then to assure that elements of  $\mathbf{s}$  are calculated in the same order, as determined by the consistent ordering of  $U$ . This also applies to the statement for (15) and for the use of the approach with the SVD. In particular the specific form for (15) assumes that the  $\gamma_i$  are ordered from small to large, in opposition to the standard ordering for the SVD.

## 2.2. Algorithmic Determination of $\sigma_L$

As in [24] Theorem 2.1 suggests finding  $W_L$  such that  $\|\mathbf{k}(W_L)\|_2^2$  as closely as possible follows the  $\chi^2(I_{m+p-n}, c(W_L))$  distribution. Let  $\psi := \psi(W_L, \theta, m, n, p) = z_{\theta/2} \sqrt{2(m+p-n+2c(W_L))}$  where  $z_{\theta/2}$  is the relevant  $z$ -value for the  $\chi^2$  distribution

with  $m + p - n$  degrees of freedom. For given  $W_L$ ,  $m$ ,  $n$ , and  $p$ ,  $\theta$  defines the  $(1 - \theta)$  confidence interval for  $\|\mathbf{k}(W_L)\|_2^2$

$$(m + p - n + c(W_L)) - \psi \leq \|\mathbf{k}(W_L)\|_2^2 \leq (m + p - n + c(W_L)) + \psi, \quad (16)$$

and can be calculated for a given  $\theta$  in Matlab using `norminv`( $\theta/2, 1 - \theta/2$ ). In particular `norminv` provides the interval containing  $(1 - \theta)100\%$  of entries from a standard normal distribution. e.g. with  $\theta = .05$  we obtain the interval  $[-1.96, 1.96]$ , while for  $\theta = .95$  the interval is much tighter,  $[-.0627, .0627]$ . A root finding algorithm for  $c = 0$  and  $W_L = \sigma_L^{-2}I$  was presented in [17], and extended for  $c > 0$  in [24]. The general and difficult multi-parameter case was discussed in [15], with extensions for nonlinear problems in [16]. We collect all the parameter estimation formulae in Appendix A.

### 2.3. Related Parameter Estimation Techniques

In order to assess the impact of Theorem 2.1 in contrast to other accepted techniques for regularization parameter estimation we very briefly review key aspects of the related algorithms which are then contrasted in Section 2.4. Details can be found in the literature, but for completeness the necessary formulae when implemented for the GSVD (SVD) are given in Appendix A, here using as consistent with (1)  $\alpha := \sigma_L^{-1}$ .

The **Morozov Discrepancy Principle (MDP)**, [19], is a widely used technique for gravity and magnetic field data inversion.  $\alpha$  is chosen under the assumption that the norm of the weighted residual,  $\|\tilde{\mathbf{G}}\mathbf{y}(\alpha) - \tilde{\mathbf{r}}\|_2^2 \sim \chi^2(\delta, 0)$ , where  $\delta$  denotes the number of degrees of freedom. For a problem of full column rank  $\delta = m - n$ , [1, p. 67, Chapter 3]. But, as also noted in [16], this is only valid when  $m > n$  and, as frequently adopted in practice, a scaled version  $\delta = \rho m$ ,  $0 < \rho \leq 1$ , can be used. The choice of  $\alpha$  by **Generalized Cross Validation (GCV)** is under the premise that if an arbitrary measurement is removed from the data set, then the corresponding regularized solution should be able to predict the missing observation. The GCV formulation yields a minimization which can fail when the associated objective is nearly flat, creating difficulties to compute the minimum numerically, [7]. The **L-curve**, which finds  $\alpha$  through the trade-off between the norms of the regularization  $L(\mathbf{m} - \mathbf{m}_0)$  and the weighted residuals, [7, 9], may not be robust for problems that do not generate well-defined corners, making it difficult to find the point of maximum curvature of the plot as a function of  $\alpha$ . Indeed, when  $m < n$  the curve is generally smoother and it is harder to find  $\alpha_{\text{opt}}$ , [12, 31]. Unlike MDP, LC and GCV, the **Unbiased Predictive Risk Estimator (UPRE)** has apparently not been previously used for focusing gravity inversion. Moreover, while MDP, LC and GCV are presented in [9] for the underdetermined problem, we are also not aware of the use of the UPRE in more standard contexts when  $m < n$ . However, the formulation for the UPRE easily extends to the case with  $m < n$ , in the same way as for the GCV. In particular, as for the GCV, a nonlinear functional must be minimized. In this case it is designed to minimize the expected value of the predictive risk [32], and requires that information on the noise



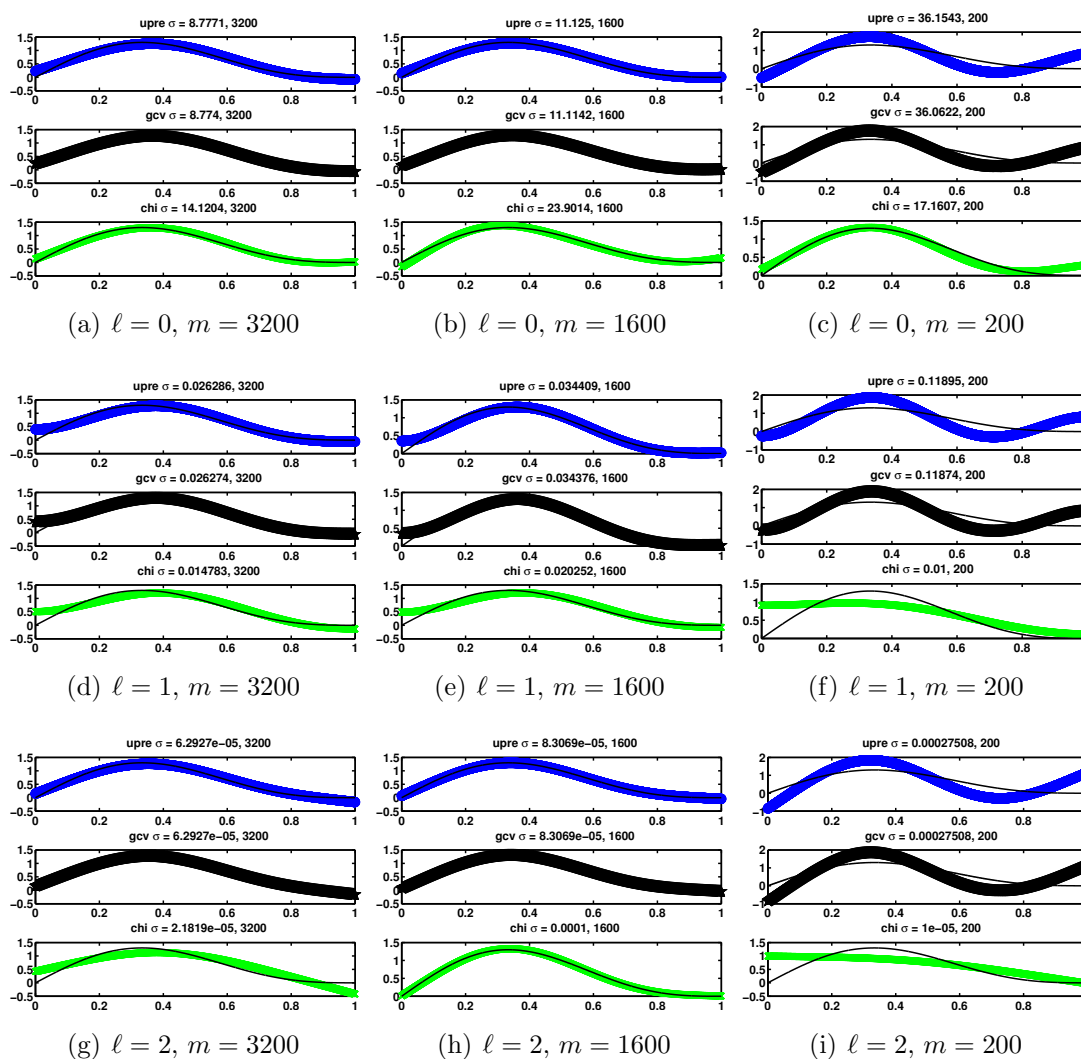
distribution in the data is provided. Apparently, there is no one approach that is likely successful in all situations. Still, the GSVD (SVD) can be used in each case to simplify the objectives and functionals e.g. [1, 9, 17, 32], hence making their repeat evaluation relatively cheap for small scale problems, and thus of relevance for comparison in the underdetermined situation with the proposed  $\chi^2$  method (A.9).

#### 2.4. Numerical Evaluation for Underdetermined Problems

We first assess the efficacy of using the noted regularization parameter estimation techniques for the solution of underdetermined problems, by presentation of some illustrative results using two examples from the standard literature, namely problems **gravity** and **tomo** from the Regularization Toolbox, [9]. Problem **gravity** models a 1-D gravity surveying problem for a point source located at depth  $z$  and convolution kernel  $K(s, t) = 1/z(z^2 + (s - t)^2)^{-1.5}$ . The conditioning of the problem is worse with increasing  $z$ . We chose  $z = .75$  as compared to the default  $z = .25$  and consider the example for data measured for the kernel integrated against the source function  $f(t) = \sin(\pi t) + 0.5 \sin(2\pi t)$ . Problem **tomo** is a two dimensional tomography problem in which each right hand side datum represents a line integral along a randomly selected straight ray penetrating a rectangular domain. Following [9] we embed the structure from problem **blur** as the source with the domain. These two test problems suggest two different situations for under sampled data. For **tomo**, it is clear that an undersampled problem is one in which insufficient rays are collected; the number of available projections through the domain are limited. To generate the data we take the full problem for a given  $n$ , leading to right hand side samples  $d_i, i = 1 : n$  and to under sample we take those same data and use the first  $m$  data points,  $d_i, i = 1 : m, m < n$ . For **gravity**, we again take a full set of data for the problem of size  $n$ , but because of the underlying integral equation relationship for the convolution, under sampling represents sampling at a constant rate from the  $d_i$ , i.e. we take the right hand side data  $d(1 : \Delta i : n)$  for a chosen integer sample step,  $\Delta i$ . Because the L-curve and MDP are well-known, we only present results contrasting UPRE, GCV and  $\chi^2$ .

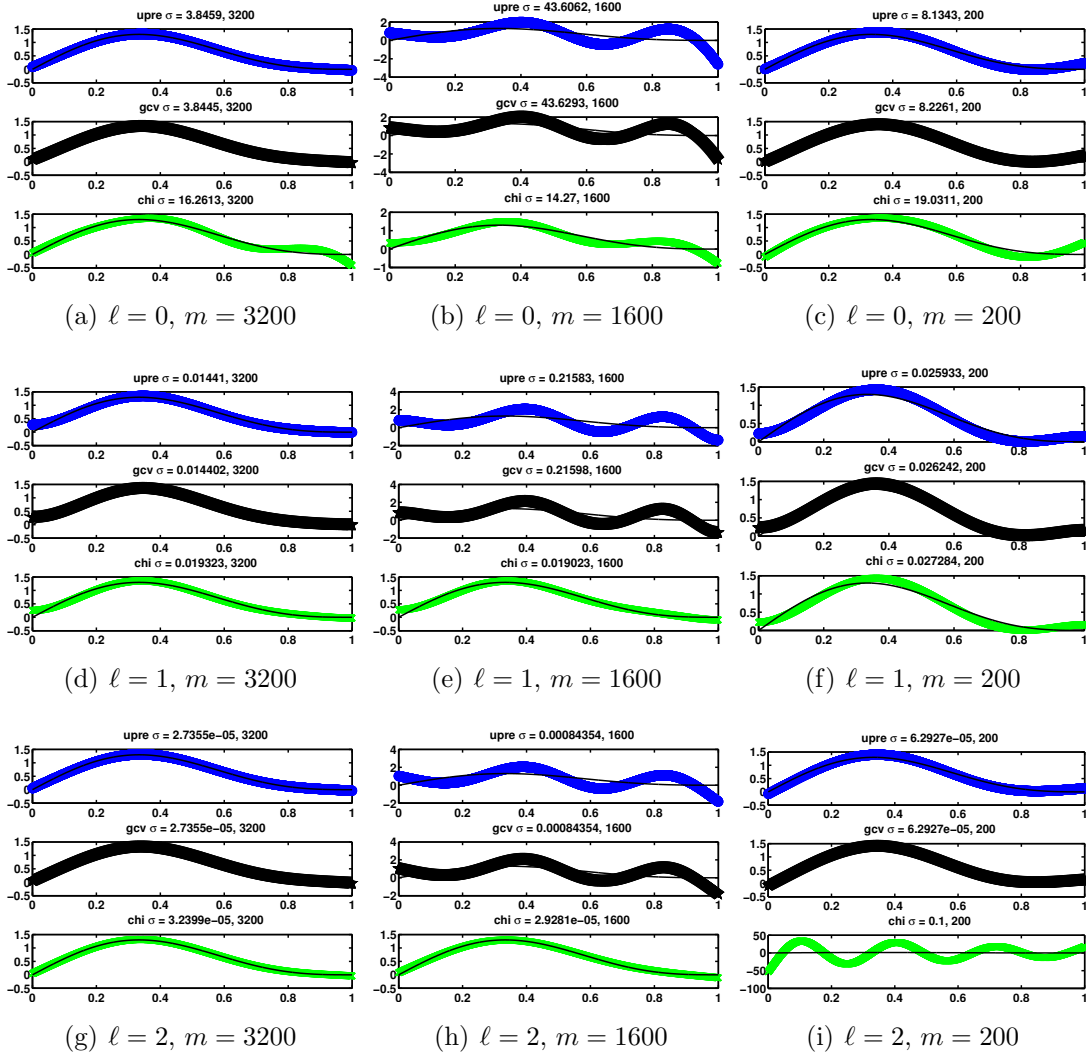
*2.4.1. Problem gravity* We take full problem size  $n = 3200$  and use sampling rates  $\Delta i = 1, 2, 4, 8$  and  $16$ , leading to problems of sizes  $m \times n, m = 3200, 1600, 800, 400, 200$ , so that we can contrast the solutions of the  $m < n$  case with those of the full case  $m = n$ . The mean and standard deviation of the relative error over 25 copies of the data are taken for noise levels  $\eta = 0.1$  and  $\eta = .01$ . Noisy data are obtained as  $\mathbf{d}^c = \mathbf{d} + \eta \max(\mathbf{d})\Theta^c, c = 1 : 25$ , with  $\Theta^c$  sampled from standard normal distribution using Matlab function **randn**. In downsampling,  $\mathbf{d}^c$  are found for the full problem, and downsampling is applied to each  $\mathbf{d}^c$ , hence preserving the noise across problem size. The UPRE and GCV algorithms use 200 points to find the minimum and the  $\chi^2$  is solved with tolerance determined by  $\theta = 0.90$  in (16), corresponding to satisfaction of the root finding condition to tolerance approximately  $.1257\sqrt{2(m + p - n)}$ . Noise levels  $\eta = .1$

and .01 correspond to white noise variance approximately .01, and .0001, respectively. Matrices are weighted by the assumption of white noise rather than colored noise. The results of the mean and standard deviation of the relative error for the 25 samples are detailed in Tables 1-2 for the two noise levels, all data sample rates, and for derivative orders in the regularization of order 0, 1 and 2. Some randomly selected illustrative results, at down sampling rates 1, 2 and 10 for each noise level are shown in Figures 1-2.



**Figure 1.** Illustrative Results for noise level .1 for randomly selected sample right hand side in each case, but the same right hand side for each method. The exact solutions are given by the thin lines in each plot.

The quantitative results presented in Tables 1-2, with the *best* results in each case in bold, demonstrate the remarkable consistency of the UPRE and GCV results. Application of the  $\chi^2$ -principle is not as successful when  $L = I$  for which the lack of a useful prior estimate for  $\hat{\mathbf{m}}$ , theoretically required to apply the central version



**Figure 2.** Illustrative Results for noise level .01 for randomly selected sample right hand side in each case, but the same right hand side for each method. The exact solutions are given by the thin lines in each plot.

of Theorem 2.1, has a far greater impact. On the other hand, for derivatives of order  $\ell = 1$  and  $\ell = 2$  this information is less necessary and competitive results are obtained, particularly for the lower noise level. Results in Figures 1-2 demonstrate that all algorithms can succeed, even with significant under sampling,  $m = 200$ , but also may fail even for  $m = 1600$ . When calculating for individual cases, rather than multiple cases at a time as with the results here, it is possible to adjust the number of points used in the minimization for the UPRE or GCV functionals. For the  $\chi^2$  method it is possible to adjust the tolerance on root finding, or apply filtering of the singular values, with commensurate adjustment of the degrees of freedom, dependent on analysis of the root finding curve. It is clear that these are worst case results for the  $\chi^2$ -principle because of the lack of use of prior information.

$m$	3200	1600	800	400	200
Method	Derivative Order 0				
UPRE	<b>.175(.088)</b>	<b>.218(.158)</b>	<b>.213(.082)</b>	<b>.239(.098)</b>	.331(.204)
GCV	<b>.175(.088)</b>	<b>.218(.158)</b>	<b>.213(.082)</b>	<b>.239(.098)</b>	.332(.205)
$\chi^2$	.223(.179)	.273(.234)	.331(.180)	.327(.186)	<b>.290(.161)</b>
	Derivative Order 1				
UPRE	.202(.084)	<b>.248(.151)</b>	<b>.238(.077)</b>	<b>.260(.088)</b>	.336(.201)
GCV	.202(.084)	<b>.248(.151)</b>	<b>.238(.077)</b>	<b>.260(.088)</b>	.337(.202)
$\chi^2$	<b>.190(.052)</b>	.260(.171)	.272(.093)	.286(.116)	<b>.305(.065)</b>
	Derivative Order 2				
UPRE	<b>.195(.111)</b>	<b>.246(.160)</b>	<b>.257(.087)</b>	.280(.094)	<b>.361(.188)</b>
GCV	<b>.195(.111)</b>	<b>.246(.160)</b>	<b>.257(.087)</b>	<b>.279(.093)</b>	<b>.361(.188)</b>
$\chi^2$	.226(.087)	.258(.084)	.430(.230)	.338(.161)	.397(.175)

**Table 1.** The mean and standard deviation of the relative error over 25 copies of the data with noise level .1. In each case  $n = 3200$  and downsampling is obtained by sampling at a sampling rate 1, 2, 4, 8 and 16. Best results for each case in boldface.

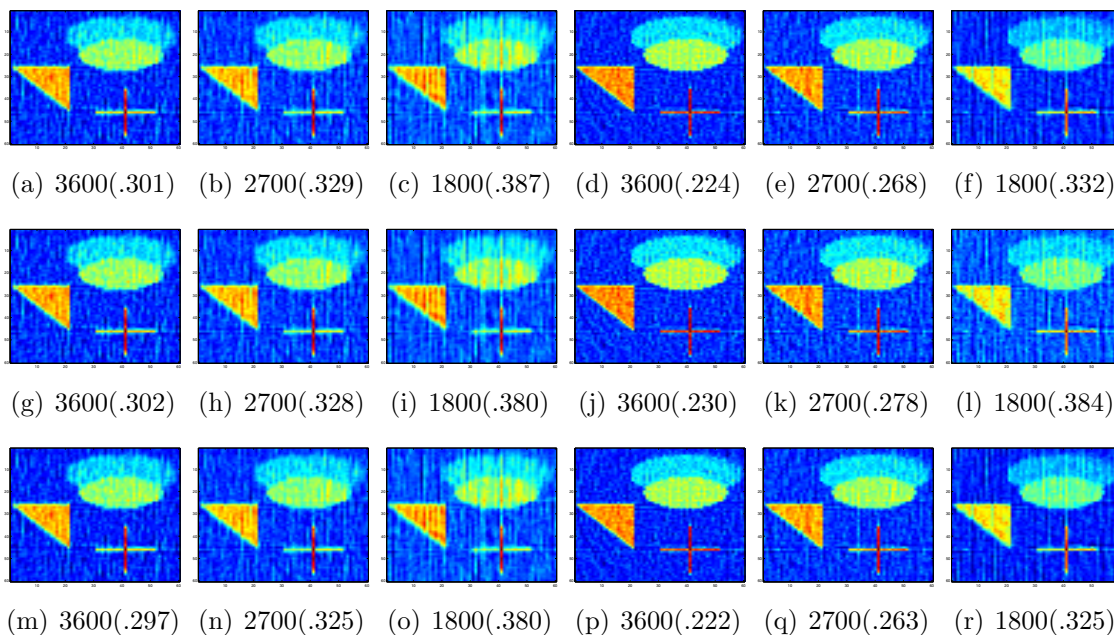
$m$	3200	1600	800	400	200
Method	Derivative Order 0				
UPRE	<b>.149(.205)</b>	<b>.075(.122)</b>	<b>.199(.301)</b>	<b>.120(.103)</b>	<b>.139(.081)</b>
GCV	<b>.149(.205)</b>	.075(.123)	<b>.199(.301)</b>	.120(.104)	<b>.139(.081)</b>
$\chi^2$	.255(.165)	.166(.130)	.300(.272)	.232(.120)	.267(.176)
	Derivative Order 1				
UPRE	.164(.197)	.108(.123)	.187(.258)	.164(.161)	<b>.155(.067)</b>
GCV	.164(.197)	.108(.123)	.187(.258)	.164(.161)	<b>.155(.067)</b>
$\chi^2$	<b>.151(.202)</b>	<b>.088(.030)</b>	<b>.137(.140)</b>	<b>.119(.058)</b>	.178(.197)
	Derivative Order 2				
UPRE	.125(.203)	.063(.122)	.104(.199)	.102(.110)	<b>.101(.063)</b>
GCV	.125(.203)	.063(.122)	.104(.199)	<b>.095(.103)</b>	<b>.101(.063)</b>
$\chi^2$	<b>.051(.034)</b>	<b>.045(.030)</b>	<b>.061(.040)</b>	.148(.209)	.187(.228)

**Table 2.** The mean and standard deviation of the relative error over 25 copies of the data with noise level .01. In each case  $n = 3200$  and downsampling is obtained by sampling at a sampling rate 1, 2, 4, 8 and 16. Best results for each case in boldface.

*2.4.2. Problem tomo* Figure 3 illustrates results for data contaminated by random noise with variance .0004 and .0001,  $\eta = .02$  and  $\eta = .01$ , respectively, with solutions obtained with regularization using a first order derivative operator, and sampled using 100%, 75% and 50% of the data. At these noise levels, the quality of the solutions when obtained with  $m = n$  are also not ideal, but do demonstrate that with reduction of sampling it is still possible to apply parameter estimation techniques to find effective solutions, i.e. all

methods succeed in finding useful regularization parameters, demonstrating again that these techniques can be used for under sampled data sets.

Overall the results for **gravity** and **tomo** demonstrate that GCV, UPRE and  $\chi^2$  algorithms for regularization parameter estimation can be successfully applied for problems with fewer samples than desirable.



**Figure 3.** Illustrative results in row 1 for the UPRE, in row 2 for the GCV and in row 3 for the  $\chi^2$  principle. From left to right problem size 3600, 2700 and 1800, for noise level .02 and then .01. The label gives the sample and the relative error  $m(\text{error})$ .

### 3. Algorithmic Considerations for the Iterative MS stabilizer

The results of Section 2.4 demonstrate the relative success of regularization parameter estimation techniques, while also showing that in general with limited data sets improvements may be desirable. Here the iterative technique using the MS stabilizing operator which is frequently used for geophysical data inversion is considered. Its connection with the iteratively regularized norm algorithms, discussed in [33], has apparently not been previously noted in the literature but does demonstrate the convergence of the iteration based on the updating MS stabilizing operator  $L$ :

$$L^{(k)} = (\text{diag}((\mathbf{m}^{(k-1)} - \mathbf{m}_0)^2) + \epsilon^2 I)^{-1/2}. \quad (17)$$

Note  $L^{(k)}$  is of size  $n \times n$  for all  $k$ , and the use of small  $\epsilon^2 > 0$  assures  $\text{rank}(L^{(k)}) = n$ , avoiding instability for the components converging to zero,  $\mathbf{m}_j - (\mathbf{m}_0)_j \rightarrow 0$ . With this  $L$  we see that  $L = L(\mathbf{m})$  and hence, in the notation of [34] (3) is of pseudo-quadratic form and the iterative process is required. The iteration to find  $\mathbf{m}^{(k)}$  from  $\mathbf{m}^{(k-1)}$ , as in [34],

replacing  $L$  by  $L^{(k)} := L(\mathbf{m}^{(k-1)})$  transforms (3) to a standard Tikhonov regularization, equivalent to the iteratively reweighted norm of [33], which can be initialized with  $\mathbf{m}^{(0)} = 0$ . Theorem 2.1 can be used with  $m$  degrees of freedom.

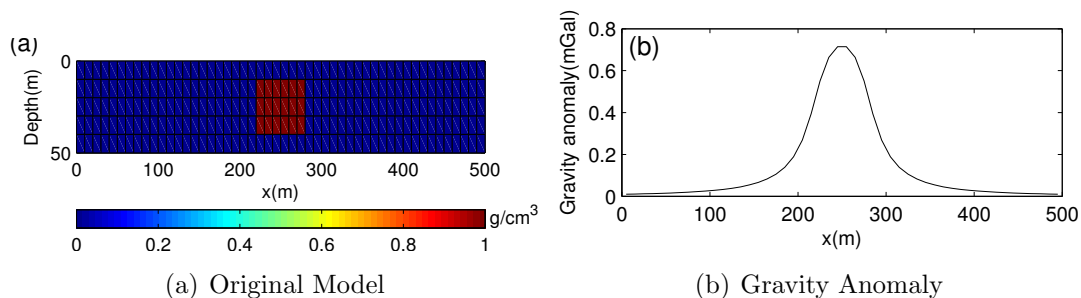
The regularization parameter  $\{\alpha^{(k)}\}$  needed at each iteration can be found by applying any of the noted algorithms, e.g. UPRE, GCV,  $\chi^2$ , MDP, LC, at the  $k^{\text{th}}$  iteration, using the SVD calculated for the matrix  $\tilde{G}(L^{(k)})^{-1}$ , here noting that solving the mapped right preconditioned system is equivalent to solving the original formulation, and avoids the GSVD. In particular (1) in the MS approach is replaced by

$$P^{\sigma_L}(\mathbf{m}) = \|G\mathbf{m} - \mathbf{d}_{\text{obs}}\|_{C_d^{-1}}^2 + (\alpha^{(k)})^2 \|L^{(k)}(\mathbf{m} - \mathbf{m}^{(k-1)})\|^2, \quad (18)$$

with  $L^{(k)}$  given by (17), and  $\{\alpha^{(k)}\}$  found according to a chosen algorithm, e.g. UPRE, GCV,  $\chi^2$ , MDP, LC. In our experiments for the 2D gravity model, we contrast the use of an initial zero estimate of the density with an initial estimate for  $\mathbf{m}_0$  obtained from the data and based on the generalized singular values for which the central form of the  $\chi^2$  iteration is better justified. For the case with prior information, the initial choice for  $\alpha^{(1)}$  is picked without consideration of regularization parameter estimation, and the measures for convergence are for the iterated solutions  $\mathbf{m}^{(k)}$ ,  $k \geq 1$ .

### 3.1. Numerical Results: A 2D Gravity Model

We contrasted the use of the regularization parameter estimations techniques on an underdetermined 2D gravity model. Figure 4(a)-4(b) shows this model and its gravity value. The synthetic model is a rectangular body, 60m  $\times$  30m, that has density



**Figure 4.** (a) Model of a body set in a grid of square cells each of size 10m, the density contrast of the body is  $1\text{gr}/\text{cm}^3$ . (b) The gravity anomaly due to the synthetic model.

contrast  $1\text{gr}/\text{cm}^3$  with an homogeneous background. Simulation data are calculated at 50 stations with 10m spacing on the surface. The subsurface is divided into  $50 \times 5$  cells with  $10\text{m} \times 10\text{m}$  dimension, hence in this case  $m = 50$  and  $n = 250$ . In generating noise-contaminated data we generate a random matrix  $\Theta$  of size  $m \times 50$ , with columns  $\Theta^c$ ,  $c = 1 : 50$ , using the MATLAB function `randn`. Then setting  $\mathbf{d}^c = \mathbf{d} + (\eta_1 \mathbf{d} + \eta_2 \|\mathbf{d}\|) \Theta^c$ , generates 50 noisy samples of the right-hand side vector

**d.** Results of this 20% under sampling are presented for 3 noise levels, namely ( $\eta_1 = 0.01, \eta_2 = 0.001; \eta_1 = 0.03, \eta_2 = 0.005$  and  $\eta_1 = 0.05, \eta_2 = 0.01$ ).

In the experiments we contrast not only the GCV, UPRE and  $\chi^2$  methods, but also the MDP and L-curve which are the standard techniques in the related geophysics literature. Details are as follows:

**Depth Weighting Matrix** Potential field inversion requires the inclusion of a depth weighting matrix within the regularization term. We use  $\beta = 0.6$  in the diagonal matrix  $(W_{\text{depth}})_{jj} = z_j^{-\beta}$  for cell  $j$  at depth  $z_j$ , and at each step form the column scaling through  $\tilde{G}^{(k)} = \tilde{G}(W_{\text{depth}})^{-1}(L^{(k)})^{-1}$ , where in  $L^{(k)} \epsilon = .02$ .

**Initialization** In all cases the inversion is initialized with a  $\mathbf{m}^{(0)} = \mathbf{m}_0$  which is obtained as the solution of the regularized problem, with depth weighting, and found for the fixed choice  $\alpha^{(0)} = (n/m) \max(\gamma_i) / \text{mean}(\gamma_i)$ , using the singular values of the weighted  $\tilde{G}$ . All subsequent iterations calculate  $\alpha^{(k)}$  using the chosen regularization parameter estimation technique.

**Stopping Criteria** The algorithm terminates for  $k = k_{\text{final}}$  when one of the following conditions is met, with  $\tau = .01$ ,

- (i) a sufficient decrease in the functional is observed,  $P^{\alpha^{(k-1)}} - P^{\alpha^{(k)}} < \tau(1 + P^{\alpha^{(k)}})$ ,
- (ii) the change in the density satisfies  $\|\mathbf{m}^{(k-1)} - \mathbf{m}^{(k)}\| < \sqrt{\tau}(1 + \|\mathbf{m}^{(k)}\|)$ ,
- (iii) a maximum number of iterations,  $K$ , is reached, here  $K = 20$ .

**Bound Constraints** Based on practical knowledge the density is constrained to lie between  $[0, 1]$  and any values outside the interval are projected to the closest bound.

**$\chi^2$  algorithm** The Newton algorithm used for the  $\chi^2$  algorithm is iterated to tolerance determined by a confidence interval  $\theta = .95$  in (16), dependent on the number of degrees of freedom, corresponding to 0.6271 for 50 degrees of freedom. The maximum degrees of freedom is adjusted dynamically dependent on the number of significant singular values, with the tolerance adjusted at the same time. We note that the number of degrees of freedom often drops with the iteration and thus the tolerance increases, but that the difference in results with choosing lower tolerance  $\theta = .90$ , leads to almost negligible change in the results.

**Exploring  $\alpha$**  At each step for the L-curve, MDP, GCV and UPRE, the solution is found at each iteration for 1000 choices of  $\alpha$  over a range dictated by the current singular values, see the discussion for the L-curve in e.g. [1, 9].

**MDP algorithm** To find  $\alpha$  by the MDP we interpolate  $\alpha$  against the weighted residual for 1000 values for  $\alpha$  and use the Matlab function `interp1` to find the  $\alpha$  which solves for the degrees of freedom. Here we use  $\delta = m$  so as to avoid the complication in the comparison of how to scale  $m$ , i.e. we use  $\rho = 1$ , see Section 2.3 and (A.10).

Tables 3-5 contrast the performance of the  $\chi^2$  discrepancy, MDP, LC, GCV and UPRE methods with respect to relative error,  $\|(\mathbf{m}_{\text{exact}} - \mathbf{m}^{(k_{\text{final}})})\|_2 / \|\mathbf{m}_{\text{exact}}\|_2$ , and the average regularization parameter calculated at the final iteration. We also record

the average number of iterations required to convergence. In Table 3 we also give the relative errors after one just one iteration of the MS and with the zero initial condition.

**Table 3.** Mean and standard deviation of the relative error measured in the 2–norm with respect to the known solution over 50 runs. Again the *best* results in each case are indicated by the boldface entries.

	Method				
Noise	UPRE	GCV	$\chi^2$	MDP	LC
$\eta_1, \eta_2$	Results after just one step, non zero initial condition				
0.01, 0.001	.331(.008)	<b>.325(.008)</b>	<b>.325(.008)</b>	.355(.009)	.447(.055)
0.03, 0.005	<b>.353(.019)</b>	.354(.042)	.361(.020)	.418(.025)	.374(.052)
0.05, 0.01	<b>.392(.034)</b>	.409(.062)	.416(.040)	.478(.043)	.463(.067)
	Results using the non zero initial condition				
0.01, 0.001	.323(.009)	.314(.010)	<b>.317(.009)</b>	.352(.011)	.489(.078)
0.03, 0.005	<b>.339(.022)</b>	.338(.040)	.359(.022)	.413(.026)	.369(.053)
0.05, 0.01	<b>.374(.041)</b>	.393(.068)	.414(.041)	.470(.046)	.460(.070)
	Results using the initial condition $\mathbf{m}_0 = 0$				
0.01, 0.001	.322(.001)	.312(.011)	<b>.315(.001)</b>	.359(.009)	.593(.014)
0.03, 0.005	<b>.333(.020)</b>	.334(.037)	.352(.021)	.425(.026)	.451(.067)
0.05, 0.01	<b>.357(.030)</b>	.440(.086)	.388(.034)	.477(.046)	.487(.082)

**Table 4.** Mean and standard deviation of  $\alpha^{k_{\text{final}}}$  over 50 runs.

Noise	Method				
$\eta_1, \eta_2$	UPRE	GCV	$\chi^2$	MDP	LC
0.01, 0.001	35.49(5.17)	14.93(9.85)	91.32(30.56)	53.42(7.43)	3.83(1.91)
0.03, 0.005	11.02(2.62)	4.18(2.58)	72.71(32.72)	30.90(8.21)	0.86(0.08)
0.05, 0.01	7.64(4.35)	3.28(3.05)	89.79(27.97)	25.81(14.87)	0.46(0.05)

With respect to the relative error one can see that the error increases with the noise level, except that the LC appears to solve the second noise level situation with more accuracy. In most regards the UPRE, GCV and  $\chi^2$  methods behavior similarly, with relative stability of the error (smaller standard deviation in the error), and increasing error with noise level. On the other hand, the final value of the regularization parameter is not a good indicator of whether a solution is over or under smoothed, contrast e.g. the  $\chi^2$  and GCV methods. The  $\chi^2$  method is overall cheaper, fewer iterations are required and the cost per iteration is cheap, not relying on an exploration with respect to  $\alpha$ , interpolation or function minimization. The illustrated results in Figure 5, for the second noise level,  $\eta_1 = .03$  and  $\eta_2 = .005$ , for a typical result, sample 37 and one of the



**Table 5.** Mean and standard deviation of the number of iterations  $k_{\text{final}}$  to meet the convergence criteria over 50 runs. Again the *best* results in each case are indicated by the boldface entries.

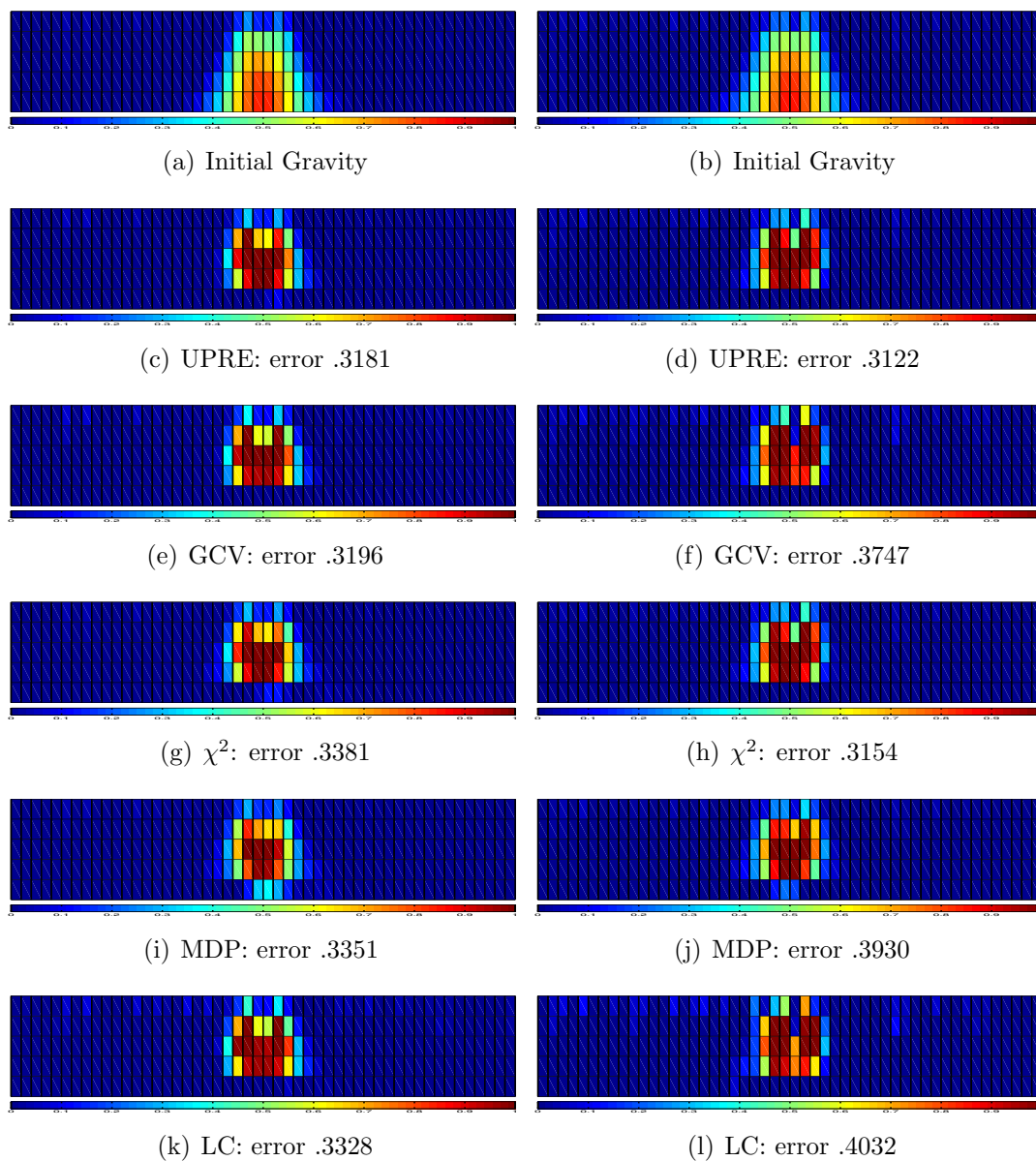
Noise	Method				
$\eta_1, \eta_2$	UPRE	GCV	$\chi^2$	MDP	LC
0.01, 0.001	18.94(0.31)	14.78(5.83)	16.26(3.00)	18.32(1.10)	<b>6.30(1.64)</b>
0.03, 0.005	11.90(2.76)	9.22(2.86)	<b>5.50(1.39)</b>	7.68(1.80)	7.90(2.48)
0.05, 0.01	7.82(1.73)	8.22(2.41)	<b>5.10(0.58)</b>	5.72(0.97)	7.84(2.41)

few cases from 50 with larger error, sample 22, demonstrate that all methods achieve some degree of acceptable solution with respect to moving from an initial estimate which is inadequate to a more refined solution. In all cases the geometry and density of the reconstructed models are close to those of the original model.

To demonstrate that the choice of the initial  $\mathbf{m}_0$  is useful for all methods, and not only the  $\chi^2$  method we show, in Figure 6, the same results as in Figure 5 but initialized with  $\mathbf{m}_0 = 0$ . In most cases the solutions that are obtained are less stable, indicating that the initial estimate is useful in constraining the results to reasonable values, however most noticeably not for the  $\chi^2$  method, but for the MDP and LC algorithms. We also illustrate the results obtained after just one iteration in Figure 7 with the initial condition  $\mathbf{m}_0$  according to Figure 5 to demonstrate the need for the iteration to generally stabilize the results. These results confirm the relative errors shown in Table 3 for averages of the errors over the 50 cases.

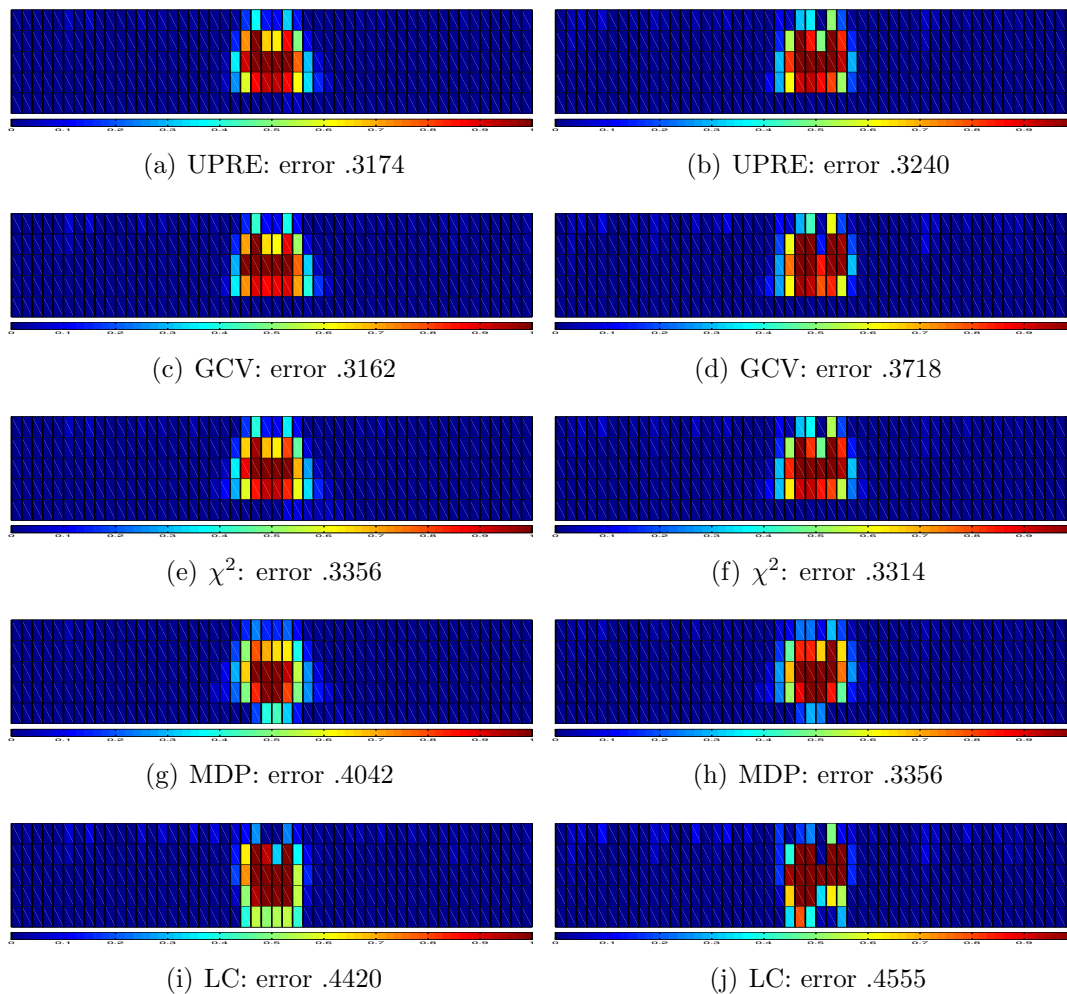
#### 4. Conclusions

The  $\chi^2$  algorithm for estimating a regularization parameter in the context of underdetermined Tikhonov regularization has been developed and investigated, extending the  $\chi^2$  method discussed in [14, 15, 16, 17, 18]. It has been contrasted for standard problems from the Regularization Toolbox [9], as well as with the extension of the UPRE to the underdetermined case, and with the GCV which is already given in [9]. UPRE and  $\chi^2$  techniques require that an estimate of the noise distribution in the data measurements is available, while ideally the  $\chi^2$  also requires a prior estimate of the mean of the solution in order to apply the central version of the  $\chi^2$  algorithm. Results demonstrate that UPRE, GCV and  $\chi^2$  techniques are useful for under sampled data sets, with UPRE and GCV yielding very consistent results. The  $\chi^2$  is more useful in the context of the mapped problem where prior information is not required. On the other hand, we have shown that the use of the iterative MS stabilizer provides an effective alternative to the non-central algorithm suggested in [18] for the case without prior information. The UPRE, GCV and  $\chi^2$  generally outperform LC and MDP methods to find the regularization parameter in the context of the iterative MS stabilizer for  $2D$  gravity inversion. Moreover, with regard to efficiency, the  $\chi^2$  generally requires fewer



**Figure 5.** Density model obtained from inverting the noise-contaminated data. The regularization parameter was found using the UPRE in 7(a)-7(b), the GCV in 7(c)-7(d), the  $\chi^2$  in 7(e)-7(f), the MDP in 7(g)-7(h), and the L-Curve in 7(i)-7(j). In each case the initial value  $\mathbf{m}_0^{(0)}$  is illustrated in 5(a)-5(b), respectively. The data are two cases with noise level,  $\eta_1 = .03$  and  $\eta_2 = .005$ . The example on the left is from a typical result; namely sample 37 from the 50 noisy samples. On the right is one of the few cases out of the 50 noisy samples for which the error is larger; here sample 22. One can see that results are overall either consistently good or consistently poor, except that the  $\chi^2$  and UPRE results are not bad in either case.

iterations, and is also cheaper to implement for each iteration because there is no need to sweep through a large set of  $\alpha$  values in order to find the optimal value. These results are useful for the development of approaches for solving larger 3D problems of gravity

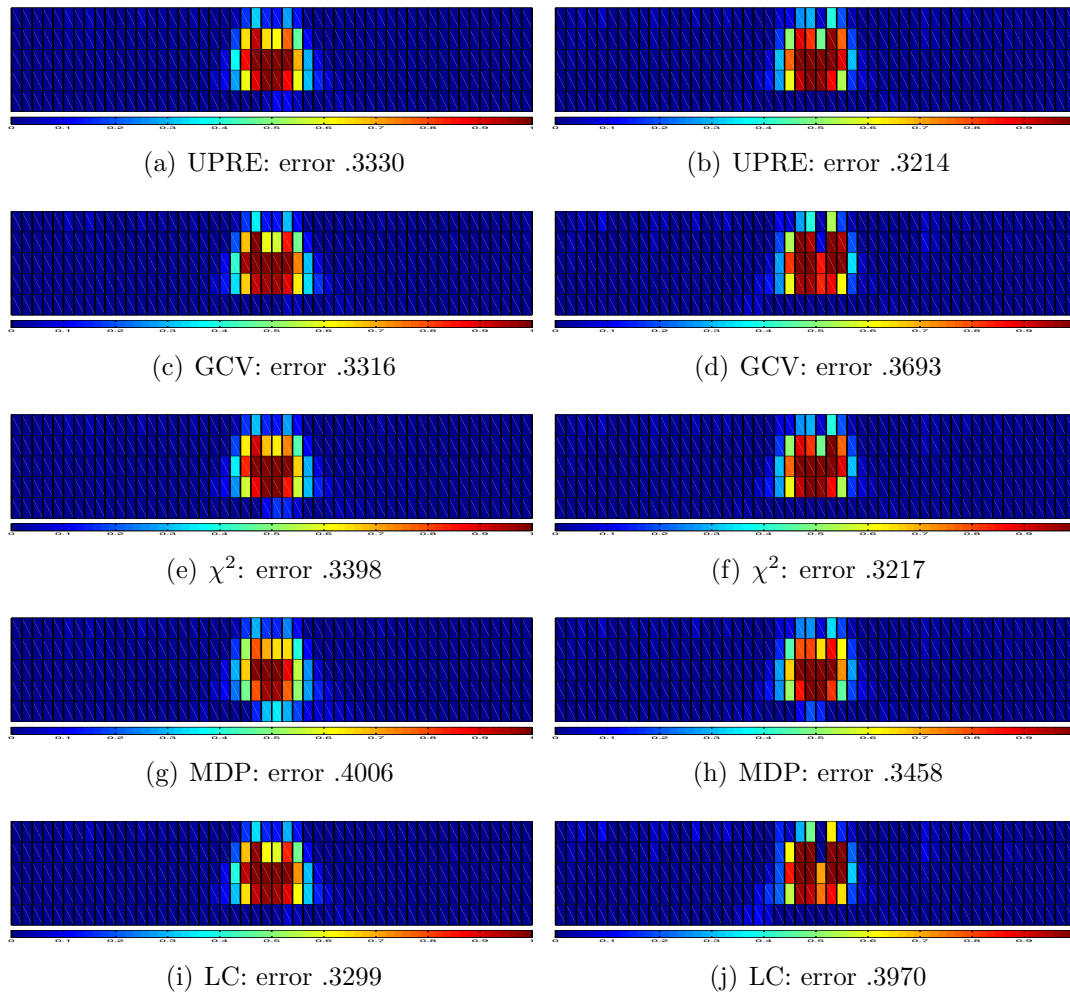


**Figure 6.** Density model obtained from inverting the noise-contaminated data, as in Figure 5 except initialized with  $\mathbf{m}_0 = 0$

inversion, which will be investigated in future work. Then, the ideas have to be extended for iterative techniques replacing the SVD or GSVD for the solution.

### Acknowledgments

Rosemary Renaut acknowledges the support of AFOSR grant 025717: “Development and Analysis of Non-Classical Numerical Approximation Methods”, and NSF grant DMS 1216559: “Novel Numerical Approximation Techniques for Non-Standard Sampling Regimes”. She also notes conversations with Professor J. Mead concerning the extension of the  $\chi^2$ -principle to the underdetermined situation presented here.



**Figure 7.** Density model obtained from inverting the noise-contaminated data, as in Figure 5 after just one step of the MS iteration.

## Appendix A. Parameter Estimation Formulae

We assume that the matrices and data are pre weighted by the covariance of the data, and thus use the GSVD of Lemma 1 for the matrix pair  $[\tilde{G}; L]$ . We also introduce inclusive notation for the limits of the summations, that are correct for all choices of  $(m, n, p, r)$ , where  $r \leq \min(m, n)$  determines filtering of the least  $p - r - \tilde{q}$  singular values  $\gamma_i$ ,  $\tilde{q} = \max(n - m, 0)$ . Then  $\mathbf{m}(\sigma_L) = \mathbf{m}_0 + \mathbf{y}(\sigma_L)$  is obtained for

$$\mathbf{y}(\sigma_L) = \sum_{i=\tilde{q}+1}^p \frac{v_i}{v_i^2 + \sigma_L^{-2} \mu_i^2} s_i \mathbf{z}_i + \sum_{i=p+1}^n s_i \mathbf{z}_i = \sum_{i=q+1}^p f_i \frac{s_i}{v_i} \mathbf{z}_i + \sum_{i=p+1}^n s_i \mathbf{z}_i, \quad (\text{A.1})$$

where  $Z := (X^T)^{-1} = [\mathbf{z}_1, \dots, \mathbf{z}_n]$ ,  $f_i = \left( \frac{\gamma_i^2}{\gamma_i^2 + \sigma_L^{-2}} \right)$  are the filter factors and  $s_i = \mathbf{u}_{i-\tilde{q}}^T \tilde{\mathbf{r}}$ ,  $s_i = 0$ ,  $i < q$ . Orthogonal matrix  $V$  replaces  $(X^T)^{-1}$  and  $\sigma_i$  replaces  $\gamma_i$ , when applied for the singular value decomposition  $\tilde{G} = U \Sigma V^T$  with  $L = I$ .

Let  $\tilde{s}_i(\sigma_L) = s_i/(\gamma_i^2\sigma_L^2 + 1)$ , and note the filter factors with truncation are given by

$$f_i = \begin{cases} 0 & \tilde{q} + 1 \leq i \leq p - r \\ \frac{\gamma_i^2}{\gamma_i^2 + \sigma_L^2} & p - r + 1 \leq i \leq p \\ 1 & p + 1 \leq i \leq n \end{cases} \quad (1 - f_i) = \begin{cases} 1 & \tilde{q} + 1 \leq i \leq p - r \\ \frac{1}{\gamma_i^2\sigma_L^2 + 1} & p - r + 1 \leq i \leq p \\ 0 & p + 1 \leq i \leq n \end{cases} . \quad (\text{A.2})$$

Then, with the assumption that if a lower limit is higher than a higher limit on a sum the contribution is 0,

$$\begin{aligned} \text{trace}(I_m - G(\sigma_L)) &= m - \sum_{i=\tilde{q}+1}^{\min(n,m)} f_i = (m - (n - (p - r))) + \sum_{i=p-r+1}^{\min(n,m)} (1 - f_i) \\ &= (m + p - n - r) + \sum_{i=p-r+1}^p \frac{1}{\gamma_i^2\sigma_L^2 + 1} := T(\sigma_L) \end{aligned} \quad (\text{A.3})$$

$$\|(I_m - G(\sigma_L))\tilde{\mathbf{r}}\|_2^2 = \sum_{i=p-r+1}^p (1 - f_i)^2 s_i^2 + \sum_{i=n+1}^m s_i^2 + \sum_{i=\tilde{q}+1}^{p-r} s_i^2 \quad (\text{A.4})$$

$$= \sum_{i=p-r+1}^p \tilde{s}_i^2(\sigma_L) + \sum_{i=n+1}^m s_i^2 + \sum_{i=\tilde{q}+1}^{p-r} s_i^2 := N(\sigma_L). \quad (\text{A.5})$$

Therefore we seek in each case  $\sigma_L$  as the root, minimum or corner of a given function.

**UPRE:** Minimizing  $(\|\tilde{G}\mathbf{y}(\sigma_L) - \tilde{\mathbf{r}}\|_2^2 + 2 \text{trace}(G(\sigma_L)) - m)$  we may shift by constant terms and minimize

$$U(\sigma_L) = \sum_{i=p-r+1}^p (1 - f_i)^2 s_i^2 + 2 \sum_{i=p-r+1}^p (f_i - 1) = \sum_{i=p-r+1}^p \tilde{s}_i^2 - 2 \sum_{i=p-r+1}^p \frac{1}{\gamma_i^2\sigma_L^2 + 1}. \quad (\text{A.6})$$

**GCV:** Minimize

$$GCV(\sigma_L) = \frac{\|\tilde{G}\mathbf{y}(\sigma_L) - \tilde{\mathbf{r}}\|_2^2}{\text{trace}(I_m - G(\sigma_L))^2} = \frac{N(\sigma_L)}{T^2(\sigma_L)} \quad (\text{A.7})$$

**$\chi^2$ -principle** The iteration to find  $\sigma_L$  requires

$$\|\mathbf{k}(\sigma_L)\|_2^2 = \sum_{i=q+1}^p \frac{s_i^2}{\gamma_i^2\sigma_L^2 + 1}, \quad \frac{\partial \|\mathbf{k}(\sigma_L)\|_2^2}{\partial \sigma_L} = -2\sigma_L \sum_{i=q+1}^p \frac{\gamma_i^2 s_i^2}{(\gamma_i^2\sigma_L^2 + 1)^2} = -\frac{2}{\sigma_L^3} \|L\mathbf{y}(\sigma_L)\|_2^2, \quad (\text{A.8})$$

and with a search parameter  $\beta^{(j)}$  uses the Newton iteration

$$\sigma^{(j+1)} = \sigma^{(j)} \left( 1 + \beta^{(j)} \frac{1}{2} \left( \frac{\sigma^{(j)}}{\|L\mathbf{y}(\sigma^{(j)})\|_2} \right)^2 (\|\mathbf{k}(\sigma^{(j)})\|_2^2 - (m + p - n)) \right). \quad (\text{A.9})$$

This iteration holds for the filtered case by defining  $\gamma_i = 0$  for  $q + 1 \leq i \leq p - r$ , removing the constant terms in (15) and using  $r$  degrees of freedom in place of  $m + p - n$ , [24].

**MDP** : For  $0 < \rho \leq 1$  and  $\delta = m$ , solve

$$\|(I_m - G(\sigma_L))\tilde{\mathbf{r}}\|_2^2 = N(\sigma_L) = \rho\delta. \quad (\text{A.10})$$

**L-curve**: Determine the corner of the log-log plot of  $\|L\mathbf{y}\|_2$  against  $\|\tilde{G}\mathbf{y}(\sigma_L) - \tilde{\mathbf{r}}\|_2$ , namely the corner of the curve parameterized by

$$\left( \sqrt{N(\sigma_L)}, \sigma_L^2 \sqrt{\sum_{i=p-r+1}^p \frac{\gamma_i^2 s_i^2}{(\gamma_i^2 \sigma_L^2 + 1)^2}} \right).$$

## References

- [1] Aster R C, Borchers B and Thurber C H 2013 *Parameter Estimation and Inverse Problems* second edition ( Amsterdam The Netherlands: Elsevier Inc. )
- [2] Donatelli M, Hanke M 2013 Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring *Inverse Problems* **29** 9 095008
- [3] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (Dordrecht The Netherlands: Kluwer)
- [4] Golub G H, Heath M and Wahba G 1979 Generalized Cross Validation as a method for choosing a good ridge parameter *Technometrics* **21** 2 215-223
- [5] Golub G H and van Loan C 1996 *Matrix Computations* (Baltimore USA: John Hopkins Press) 3rd ed.
- [6] Hanke M and Groetsch CW 1998 Nonstationary iterated Tikhonov regularization *J. Optim. Theor. Appl.* **98** 37-53.
- [7] Hansen P C 1992 Analysis of discrete ill-posed problems by means of the L-curve *SIAM Review* **34** 561-580.
- [8] Hansen P C 1998 *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion* SIAM Monographs on Mathematical Modeling and Computation **4** (Philadelphia USA: SIAM)
- [9] Hansen, P. C., 2007, *Regularization Tools Version 4.0 for Matlab 7.3*, Numerical Algorithms, **46**, 189-194, and <http://www2.imm.dtu.dk/~pcha/Regutools/>.
- [10] Hansen P C, Kilmer M E and Kjeldsen R H 2006 Exploiting residual information in the parameter choice for discrete ill-posed problems *BIT* **46** 41-59.
- [11] Last B J and Kubik K 1983 Compact gravity inversion *Geophysics* **48** 713-721
- [12] Li Y and Oldenburg D W 1999 3D Inversion of DC resistivity data using an L-curve criterion *69th Ann. Internat. Mtg., Soc. Expl. Geophys. Expanded Abstracts* 251-254, <http://dx.doi.org/10.1190/1.1820992>
- [13] Marquardt D W 1970 Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation *Technometrics* **12** (3) 591-612.
- [14] Mead J L 2008 Parameter estimation: A new approach to weighting a priori information *Journal of Inverse and Ill-Posed Problems* **16** 2 175-194.
- [15] Mead J L 2013 Discontinuous parameter estimates with least squares estimators *Applied Mathematics and Computation* **219** 5210-5223.
- [16] Mead J L and Hammerquist C C 2013  $\chi^2$  tests for choice of regularization parameter in nonlinear inverse problems *SIAM Journal on Matrix Analysis and Applications* **34** 3 1213-1230.
- [17] Mead J L and Renaut R A 2009 A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems *Inverse Problems* **25** 025002 doi: 10.1088/0266-5611/25/2/025002.
- [18] Mead J L and Renaut R A 2010 Least Squares problems with inequality constraints as quadratic constraints *Linear Algebra and its Applications* **432** 8 1936-1949 doi:10.1016/j.laa.2009.04.017.

- [19] Morozov V A 1966 On the solution of functional equations by the method of regularization *Sov. Math. Doklady* **7** 414-417.
- [20] Paige C C and Saunders M A 1981 Towards a generalized singular value decomposition *SIAM Journal on Numerical Analysis* **18** 3 398-405.
- [21] Paige C C and Saunders M A 1982 LSQR: An algorithm for sparse linear equations and sparse least squares *ACM Trans. Math. Software* **8** 43-71.
- [22] Paige C C and Saunders M A 1982 ALGORITHM 583 LSQR: Sparse linear equations and least squares problems *ACM Trans. Math. Software* **8** 195-209.
- [23] Portniaguine O and Zhdanov M S 1999 Focusing geophysical inversion images *Geophysics* **64** 874-887.
- [24] Renaut R A, Hnetyňková I and Mead J L 2010 Regularization parameter estimation for large scale Tikhonov regularization using a priori information *Computational Statistics and Data Analysis* **54** 12 3430-3445 doi:10.1016/j.csda.2009.05.026.
- [25] Ring W 1999 *Structural Properties of solutions of Total Variation regularization problems Preprint* <http://www.uni-graz.at/imawww/ring/publist.html>
- [26] Rudin L I Osher S and Fatemi E 1992 *Nonlinear total variation based noise removal algorithms, Physica D* **60** 259-268.
- [27] Rust B W and O'Leary D.P 2008 Residual periodograms for choosing regularization parameters for ill-posed problems *Inverse Problems* **24** 034005.
- [28] Stefan W Garnero E and Renaut R A 2006 Signal restoration through deconvolution applied to deep mantle seismic probes *Geophysical Journal International* **167** 1353-1362.
- [29] Tarantola A 2005 *Inverse Problem Theory and Methods for Model Parameter Estimation* SIAM Series: Other Titles in Applied Mathematics (Philadelphia USA: SIAM)
- [30] Thompson A M, Kay J W and Titterton D M 1991 Noise Estimation in Signal Restoration using Regularization *Biometrika* **78** 3 475-488.
- [31] Vatankhah S Ardestani V E and Renaut R A 2014 Automatic estimation of the regularization parameter in 2-D focusing gravity inversion: an application to the Safo manganese mine in northwest of Iran, in press, *J Geophysics and Engineering* <http://arxiv.org/abs/0813162>
- [32] Vogel C R 2002 *Computational Methods for Inverse Problems* SIAM Frontiers in Applied Mathematics (Philadelphia USA: SIAM)
- [33] Wohlberg B and Rodriguez P 2007 An Iteratively Reweighted Norm Algorithm for Minimization of Total Variation Functionals *IEEE Signal Processing Letters* **14** 948-951.
- [34] Zhdanov M S 2002 *Geophysical Inverse Theory and Regularization Problems* (Amsterdam The Netherlands: Elsevier Inc.)