Truncated SVD methods for discrete linear ill-posed problems

Peiliang Xu
Research Centre for Earthquake Prediction, Kyoto University at Uji, Kyoto 611-0011, Japan. E-mail: pxu@rcep.dpri.kyoto-u.ac.jp

Accepted 1998 June 15. Received 1998 June 8; in original form 1997 June 5

SUMMARY
Truncated singular value decomposition (TSVD) techniques have been widely used in inversion. The method of truncation determines the quality of a truncated SVD solution, but truncation has often been done arbitrarily. The first workable criterion for truncation was based on $F$-statistical testing, but has only rarely been used in geophysical inversion. Recently, the L-curve approach was proposed for the same purpose, and soon found many applications in interdisciplinary inverse problems. Up to the present, very little has been known of the solution quality using these two workable criteria. We have thus proposed in this paper a new quality-based method for truncation. Six TSVD estimators have been investigated for comparison in regularizing discrete unstably ill-posed problems, based on the statistically frequently used $F$-statistic, the L-curve and our new quality-based mean squared error (MSE) criteria. The three $F$-statistic-based TSVD estimators tested can indeed marginally improve the least squares (LS) solution to the ill-posed downward continuation problem in the sense of long-term averaging, depending on pre-selected significance levels. The simulations have shown that estimators of this type can hardly guarantee the improvement of the condition number of the linearized unstably ill-posed system. In other words, $F$-statistic criterion can frequently lead to incorrect discardings of components. The TSVD estimator by means of an L-curve is the best in (over)stabilizing ill-posed problems, but results in an over-discarding of components. It is extremely poor in terms of biases and mean squared error (MSE) roots of the solution. The quality-based TSVD estimator with the basic ridge estimate of $x$ as its initial value has outperformed the other two TSVD techniques tested in terms of solution stability, bias and MSE. The simulations have clearly shown significant quality advantages of the new method. The simulations have also indicated that the new method is able to achieve a mean accuracy of 5 mgal for gravity anomalies from satellite gradiometry, if the few largest biases are left out of the computation.

Key words: gravity anomalies, ill-posed problem, singular value decomposition.

1 INTRODUCTION
Unstably ill-posed problems have been topical for more than three decades. Examples can be found in mathematical physics (Tikhonov & Arsenin 1977), geodesy (Schwarz 1973; Rummel, Schwarz & Gerstl 1979; Sanso 1986; Marsh et al. 1988, 1990; Schwintzer et al. 1992, 1997), geomagnetism (Gubbins 1983; Gubbins & Bloxham 1985; Gire, LeMouel & Madden 1986; Langel 1987; Backus 1988; Bloxham & Jackson 1989), and geophysics and seismology (Aki & Richards 1980; Tarantola 1984; Menke 1989; Parker 1994). A typical ill-posed problem is the linear Fredholm integral equation of the first kind. The extent of ill posedness mainly depends on the property of the integral kernel, which, together with the accuracy of input measurement data and smoothing techniques, will determine how much information about the field of interest can be extracted from the measurements. The success of solving a downward continuation problem in physical geodesy and geomagnetism, for instance, relies on the size of grid cells, the ratio of the altitude at which measurements are taken to the altitude at which the quantities of interest are to be determined, the accuracy of the measurements, and proper techniques to smooth out noise.

In practical computations, an ill-posed problem can be written as the following linear (discrete) model:

$$y = Ax + \epsilon,$$

where $y$ is an observation vector, $A$ is an $(n \times t)$ design matrix of rank $t$, $x$ is an unknown vector with $t$ elements, $\epsilon$ is the error vector of the observables $y$, with mean zero and...
covariance matrix $P^{-1} \sigma^2$, where $P$ is a given positive-definite matrix. The generalized least squares (LS) estimator of $x$ is $\hat{x}_Q = (A^T PA)^{-1} A^T P y$ with the cofactor matrix $Q_{ij} = (A^T PA)^{-1}$. It is well known that if the linear model (1) is unstable ill-posed, one or more of the eigenvalues of the normal matrix $(A^T PA)$ are close to zero, which is responsible for the large uncertainty of the estimated $\hat{x}_Q$.

Alternatively to the LS method have been proposed to obtain an improved estimate of $x$: regularization (Tikhonov & Arsenin 1977), ridge- and shrinkage-type estimation (Hoerl & Kennard 1970; Vinod & Ullah 1981), Bayesian and stochastic inferences (Backus 1988), constrained estimators (Backus 1988) and truncated singular value decomposition (TSVD) or principal components (Lawless & Wang 1976; Dempster, Schatzoff & Wermuth 1977; Sekii 1991; Hansen 1990, 1992; Gough 1996).

In this paper, we will confine ourselves to estimators of TSVD type, because they have been attracting more and more attention in interdisciplinary inverse problems (Hansen 1990, 1992; Sekii 1991; Deal & Nolet 1996; Eriksson & Wedin 1996; Gough 1996; Kunitake 1996).

A key difficulty in applying TSVD methods has been how to set up proper criteria to truncate the eigenvalues. Due to the lack of a theoretically solid foundation, small eigenvalues were more or less arbitrarily discarded (see e.g. Sekii 1991). The simplest criterion is to use only the largest eigenvalues to derive the solution so that the data are well fitted (Scales 1996). The first statistically sound approach is to discard the eigenvalues on the basis of the $F$-statistic with different significance levels (Lawless & Wang 1976; Dempster et al. 1977) or trial-and-error by monitoring the stability of solutions. The TSVD or principal component method of this type was shown by simulations of the stability of solutions if only a small component method of this type was shown by simulations of the stability of solutions if only a small number of measurements are given. Thus, in this work we investigate in the final part of this work theoretically and by simulation, statistical aspects of the L-curve-based TSVD.

This quality stand will also serve as a basis to judge the two non-quality-based criteria, namely the $F$-statistic-based and $L$-curve-based techniques, through large-scale simulations. Other motivations are summarized in the following.

First, the TSVD estimators constructed using the $L$-statistic with different significance levels will be re-examined to investigate (incorrect) component cutting, which does not seem to have been studied previously. Second, the use of the $L$-curve for TSVD estimation has been topical in some of recent interdisciplinary studies in inverse problems. However, little is known about the statistical quality of the TSVD estimator by means of an L-curve in terms of solution bias and MSE, which have to be clearly clarified if the L-curve method is to be applied successfully in practice. Finally, the ill-posed problems of downward continuation type without a priori information are of urgent importance for future satellite gravimetric missions. Some of the results on downward continuation have been based on Tikhonov regularization. Too large a regularization parameter is very often selected, and a misleading variance-covariance matrix produced in the framework of stochastic inference. Theoretical and practical implications of interpreting an inverse problem from Bayesian and frequentist points of view have been investigated by Xu (1992) and Xu & Rummel (1994, 1995). Therefore, as an example, we have decided to use downward satellite gravimetric problems to demonstrate our quality-based TSVD method and to compare it with other TSVD techniques in terms of solution bias and MSE. In this paper, we assume no a priori (subjective or objective) information, and thus follow a frequentist treatment of inverse problems.

This paper is organized as follows. In Section 2, we will briefly analyse TSVD estimators by significance tests and by means of an L-curve. Section 3 proposes a new quality-based criterion to find TSVD or principal component estimators by minimizing the MSE. The convergence of the new method will be proven. It will then be compared numerically in Section 4 with estimators using the $F$-statistic with different significance levels and Hansen’s modified TSVD estimator by means of an L-curve. Vogel (1996) gave some non-convergence results of the L-curve when the number of measurement data tends to infinity. In the case of noisy data, the applicability of the $L$-curve for unstably ill-posed problems remains unknown in terms of the statistical quality of the solutions if only a small or intermediate number of measurements are given. Thus, we investigate in the final part of this work theoretically and by simulation, statistical aspects of the L-curve-based TSVD.

## 2 TSVD ESTIMATORS USING THE SIGNIFICANCE TEST AND THE L-CURVE

Applying singular value decomposition to the matrix $A$, we have

$$A = U \Sigma V^T,$$

$U$ and $V$ are orthonormal matrices, where $u_i^T u_i = \delta_{ij}$ and $v_i^T v_i = \delta_{ij}$; $u_i$ and $v_i$ are the $i$th column vectors of the matrices $U$ and $V$, respectively. $\Sigma$ is a diagonal matrix of singular values of the matrix $A$, whose non-zero diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_l$ are, without loss of generality, assumed to be positive and arranged in decreasing order, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_l > 0$. 

© 1998 RAS, GJI 135, 505–514
Since the weight matrix $P$ does not present us with any difficulties, we shall assume that $P=1$ in the following. The pseudo-inverse solution of (1) is

$$\hat{x}_P = V A^+ U^T y,$$

which can be easily proven to be equivalent to the LS solution, $\hat{x}_L$:

$$\hat{x}_L = (A^T A)^{-1} A^T y = V (A^T A)^{-1} A^T U^T y,$$

where $A^+$ is the Moore–Penrose inverse of $A$. Rewriting the LS solution $\hat{x}_L$ in spectral form yields

$$\hat{x}_L = \sum_{i=1}^{r} \frac{u_i^T y}{\lambda_i} v_i.$$

(3)

Dropping the smallest singular values from the LS estimator, we obtain a TSVD estimator,

$$\hat{x}_k = \sum_{i=1}^{k} \frac{u_i^T y}{\lambda_i} v_i,$$

(4)

where $k$ is the truncation parameter.

Denoting

$$\hat{z}_i = \frac{u_i^T y}{\lambda_i},$$

the LS solution (3) becomes

$$\hat{x}_k = \sum_{i=1}^{r} \hat{z}_i v_i = V \hat{a},$$

(6)

where $\hat{a} = (\hat{z}_1, \hat{z}_2, \ldots, \hat{z}_r)^T$. It is obvious that the TSVD estimator (4) can actually be derived by discarding those $\hat{z}_i$ corresponding to the smallest singular values from $\hat{x}_L$. It is nothing but the principal components estimator.

The success of applying the TSVD technique in order to improve the LS estimator depends to a great extent on a proper choice of the truncation parameter $k$. The determination of $k$ has been rather arbitrary in inversion, and a theoretical foundation has long been awaited. A first workable approach to determining the parameter $k$ was to use the $F$-statistic and decide which $\hat{z}_i$ ($i=1, 2, \ldots, r$) should be discarded (see, e.g. Lawless & Wang (1976); Dempster et al. (1977)), but this received little attention in inverse problems. A recent approach by Hansen (1990, 1992) was to use the L-curve and truncate $k$ at the corner point with maximum curvature (if it exists at all). The L-curve technique has, surprisingly, found more and more applications in a variety of inversions, although its quality aspect is not yet clear.

2.2 Hansen's TSVD estimators using the L-curve

Consider Tikhonov regularization applied to the discrete linear ill-posed problem (1). The minimization problem to be solved is

$$\min \|Ax - y\|^2 + \gamma^2 \|Lx\|^2,$$

(10)

where $\|Ax - y\|^2$ is the norm of the residuals $(Ax - y)$, $\|Lx\|^2$ is a semi-norm of the linear functional $Lx$. $L$ is a properly pre-selected matrix (or operator) and $\gamma^2$ is frequently called the Tikhonov regularization parameter. The solution of the minimization problem (10) is straightforwardly given by

$$\hat{x} = (A^T A + \gamma^2 L^T L)^{-1} A^T y.$$

(11)

The determination of the regularization parameter $\gamma^2$ is of great importance. If too small a $\gamma^2$ is chosen, the regularization solution (11) can remain unstable; conversely, if $\gamma^2$ is too large, the solution (11) will be dominated by any pre-determined solution. In the case of the ill-posed problem (1), $\hat{x}$ is zero for any invertible $L$ when $\gamma^2$ tends to infinity. No information from the observations $y$ is taken into account to improve our knowledge about $x$. More information about Tikhonov-type regularization can be found in Tikhonov & Arsenin (1977) and Morozov (1984).

Hansen (1990, 1992) recently suggested that the parameter $\gamma^2$ be chosen by inspecting plots of the solution (11) with respect to $\gamma^2$. One such plot is the so-called L-curve, which is defined to be the curve of $(\delta, \eta)$; that is, $\delta = \|L\hat{x}\|$ and $\eta = \|Ax - y\|$. A basic important property of the L-curve is that the curve $(\delta, \eta)$ was reported to have an L-shape, consisting of a relatively vertical segment and a relatively horizontal line in an ideal setting. The parameter $\gamma^2$ is then defined to correspond to the corner point of the L-curve, which is the point with maximum
curvature (if it exists at all). Since Hansen observed that the log–log scale can better emphasize the corner point of the curve, the L-curve actually recommended is the plot of (log δ, log η). In particular, if L = I, then (11) becomes

\[ \hat{x}_k = \sum_{i=1}^{I} \frac{\lambda_i}{\lambda_i^2 + \gamma^2} u_i^T y_i. \]  

(12)

The coordinates of the L-curve are given by

\[ \delta_i^2 = \|\hat{x}_i\|^2 = \sum_{i=1}^{I} \left[ \frac{\lambda_i^2}{\lambda_i^2 + \gamma^2} \right]^2 \delta_i. \]  

(13a)

and

\[ \eta_i^2 = \|A\hat{x}_i - y\|^2 = \sum_{i=1}^{I} y_i^2 - \sum_{i=1}^{I} \left[ \frac{\lambda_i^2}{\lambda_i^2 + \gamma^2} \right]^2 \delta_i. \]  

(13b)

When the idea of an L-curve is applied to derive estimators of the TSVD type, we only have I discrete points. Denoting the TSVD estimator as \( \hat{x}_{ik} \) in Hansen’s sense, the coordinates of these points are explicitly given by

\[ \delta_i^2 = \|\hat{x}_{ik}\|^2 = \sum_{i=1}^{k} \delta_i. \]  

(14a)

and

\[ \eta_i^2 = \|A\hat{x}_{ik} - y\|^2 = \sum_{i=1}^{k} y_i^2 - \sum_{i=1}^{k} \delta_i^2. \]  

(14b)

Although Hansen recommended fitting the discrete points with some curves and then finding the truncation parameter k that is closest to the maximum curvature point, we will simply select the parameter k that minimizes the distance between a discrete point and the origin. The reason is clearly understandable, if we keep in mind that an ideal L-curve consists of a vertical segment and a horizontal line.

3 QUALITY-BASED TSVD METHOD BY MINIMIZING THE MEAN SQUARED ERROR

Smoothing ill-posed problems is highly desirable for obtaining maximum information from measurements, since the LS solution is too uncertain and may even result in incorrect signs of x. Regularization by TSVD could offer improvement over the LS solution if the truncated parameter k is properly chosen. The conventional principal component method fails to establish a connection between the goodness of an estimator and the pre-selected significance level. The L-curve method is intuitively attractive (Hansen 1990, 1992; Regincka 1996; Vogel 1996) if one is satisfied with a best balance between the norms of the model parameters and the residuals. A disadvantage is the lack of theoretically supported relations between the corner point with maximum curvature and the optimality or statistical quality of the estimator \( \hat{x}_{ik} \). Thus we propose the first quality-based criterion for selecting the parameter k so that the quality issue of TSVD estimators is automatically taken into account.

Since the principal component method and Hansen’s L-curve approach are different only in the manner of discarding some \( \hat{x}_i \), they actually result in the same form of estimator as given in (7). Denote a TSVD estimator by \( \hat{x}_k = V_k a_k \), without loss of generality. It has been proven that \( \hat{x}_k \) is no longer unbiased. The bias vector of \( \hat{x}_k \) is

\[ \text{bias}(\hat{x}_k) = -V_H V_I^T x \]  

(15)

(see e.g. Xu & Rummel 1994). Applying the covariance propagation law to \( \hat{x}_k \) yields

\[ D(\hat{x}_k) = V_k D(a_k) V^T_k = V_k A_k^{-2} V_k^T \delta. \]  

(16)

Thus, the mean squared error of the TSVD estimator \( \hat{x}_k \) is defined by

\[ \text{MSE}(\hat{x}_k) = \text{tr}[D(\hat{x}_k)] + \text{bias}(\hat{x}_k)^T \text{bias}(\hat{x}_k) \]

\[ = \delta_k^2 \text{tr} \left[ V_k A_k^{-2} V_k^T \right] + 1 \|V_H V_k^T x \|^2, \]

(17)

where \( z_i = V_i^T x \), \( v_i \) is the i-th column vector of \( V_H \).

The mean squared error check, \( \text{MSE}(\hat{x}_k) \), consists of two terms. The first part depends on the accuracy of the statistically significant principal components \( \hat{z}_i \). An optimal MSE(\( \hat{x}_k \)) should result in a satisfactory solution stability, as is obvious from (17). The second term is determined from the normalized eigenvectors corresponding to the smallest eigenvalues (or statistically insignificant \( \hat{z}_i \)) and the model parameters x. It actually results from the bias in \( \hat{x}_k \). An optimal MSE(\( \hat{x}_k \)) should thus automatically control the biases of the estimated model parameters. With decreasing k, the first term of MSE(\( \hat{x}_k \)) decreases and the second term increases. In other words, the error variance of \( \hat{x}_k \) is a monotonically increasing function of k, while the bias-dependent term is a monotonically decreasing function.

We are now in a position to define our new TSVD estimator by mean squared error, which is the solution of the following minimization problem:

\[ \min \delta^2 \sum_{i=1}^{k} 1/\lambda_i^2 + \sum_{i=k+1}^{l} \delta_i^2, \]

(18)

where \( \delta^2 \) is the LS estimate of \( \delta_i \), \( \hat{z}_i = V_i^T \hat{x}_p \) and \( \hat{x}_p \) is some proper estimator of x.

Corollary 1: The estimated mean squared errors

\[ \left[ \delta^2 \sum_{i=1}^{k} 1/\lambda_i^2 + \sum_{i=k+1}^{l} \delta_i^2 \right] (k = 1, 2, \ldots, t) \]

are almost certainly distinct if \( \hat{x}_{ik} \) is substituted for \( \hat{x}_p \) in (18).

Proof: Let there be two non-distinct estimated mean squared errors, say

\[ \left[ \delta^2 \sum_{i=1}^{k} 1/\lambda_i^2 + \sum_{i=k+1}^{l} \delta_i^2 \right] \]

and

\[ \left[ \delta^2 \sum_{i=1}^{k} 1/\lambda_i^2 + \sum_{i=k+1}^{l} \delta_i^2 \right], \]

© 1998 RAS, GJI 135, 505–514
and $k > l$. We then have
\[
P \left\{ \left[ \delta^2 \sum_{i=1}^{k} 1/\delta_i^2 + \sum_{i=k+1}^{l} \delta_i^2 \right] - \left[ \delta^2 \sum_{i=1}^{k} 1/\delta_i^2 + \sum_{i=k+1}^{l} \delta_i^2 \right] \right\} < c = 1, \tag{19}
\]
where $c$ is a positive constant. (19) can be rewritten as
\[
P \left\{ \delta^2 \sum_{i=1}^{k} 1/\delta_i^2 - \sum_{i=k+1}^{l} \delta_i^2 \right\} < c = 1, \tag{20}
\]
which clearly cannot be true, since the random variable inside the absolute operation is a linear combination of independent random variables $\delta^2$ and $\delta_i$.

**Corollary 2:** The minimization problem (18) almost certainly has a unique solution if $\hat{x}_k$ or the basic ridge estimator is substituted into (18) for $\hat{x}_p$.

Corollary 2 is a simple consequence of the finite discrete and distinct estimated mean squared errors (Corollary 1). It has been recognized, however, that the LS solution $\hat{x}_k$ is not a good estimator of $x$ if the problem is unstably ill posed. Thus one may apply some improved estimators to (18) in deriving MSE-based TSVD estimators. Different estimators $\hat{x}_p$ in (18) will obviously result in different MSE-based TSVD estimators. If the estimators of ridge type are used, the results stated in Corollary 1 and Corollary 2 still hold true. In this paper, we will use the LS and basic ridge estimators to substitute for $\hat{x}_p$, whose corresponding MSE-based TSVD estimators are denoted by $\hat{x}_{\text{lin}}$ and $\hat{x}_{\text{ent}}$, respectively.

### 4 NUMERICAL EXPERIMENTS

In this section, we shall numerically compare our new quality-based method of determining the truncation parameter $k$ with the $F$-statistic-based and L-curve techniques, in terms of solution stability, bias and mean squared error. Six TSVD estimators, namely three $F$-statistic-based TSVDs, $\hat{x}_{\text{lin}}$, $\hat{x}_{\text{ent}}$, $\hat{x}_{\text{lin}}$, and $\hat{x}_{\text{ent}}$, the L-curve-based Hansen’s TSVD, $\hat{x}_{\text{hh}}$, and two MSE-based TSVDs, $\hat{x}_{\text{lin}}$ and $\hat{x}_{\text{ent}}$, have been designed and used in the following numerical experiments. The example taken is a typical downward continuation problem in physical geodesy. More specifically, we shall determine the gravity anomalies on the surface of the Earth using gravimetric observables collected at the altitude of satellites (about 200 km above the Earth). Since the primary emphasis of this paper is to propose a theoretically solid TSVD method for unstably ill-posed problems and compare it with other TSVD criteria, and since, at the present, data on satellite gravimetric missions, such as mission hardware, orbit altitude and inclination and/or date of launch, are not yet finally fixed, we will not go into the technical details of such a mission. For some recent information on satellite gravimetric mission, however, the reader is referred to Sneeuw & Ilk (1996) and Mueller, Sneeuw & Rummel (1996). For other potentially applicable techniques to resolve a downward continuation problem, the reader is referred to Xu & Rummel (1994), where only the iterative ridge estimation method could provide a mean accuracy better than 5 mgal for gravity anomalies among all the estimation techniques compared.

The starting integral equation of the vertical component is
\[
T_{rr} = \frac{R}{4\pi} \int \Delta g S'(r, \psi) d\sigma, \tag{21a}
\]
Here $T_{rr}$ is the second derivative of the anomalous potential $T$ with respect to $r$, which is observable in satellite gravimetry, $R$ is the mean radius of the Earth, $\Delta g$ are gravity anomalies and $\sigma$ is the unit sphere. The kernel function $S'(r, \psi)$ in the integral (21a) is the second derivative of the generalized Stokes function with respect to $r$. Contributions to geopotential fields from other gravity tensor components and their combinations can be found in Rummel & van Gelderen (1992). For an excellent introduction to satellite gravimetry, the reader is referred to Rummel (1986).

After taking the observation error of $T_{rr}$ into account, the integral equation (21a) is then discretized into a standard linear ill-posed problem:
\[
4\pi R/10T_{rr} = \sum_k S'(r, \psi) \Delta g_k + \epsilon \tag{21b}
\]
(see e.g. Xu & Rummel 1994), where $\Delta g_k$ are the block-mean gravity anomalies, the constant $4\pi R/10$ comes from the generalized Stokes’ formula and unit regulation, the random error $\epsilon$ is due to the measurement error of $T_{rr}$, and $\Delta g_k$ and $S'(r, \psi)$ are given by
\[
\Delta g_k = \int d\sigma
\]
and
\[
S'(r, \psi) = t \left\{ \left[ 1 - t \cos \psi \right] \left[ \frac{3(1 - r^2)}{D^5} - 4/D^3 \right] - \frac{1 + r^2}{D^3} - 10/D - 18D + 2 - 3t \cos \psi \left[ 15 + 6n \frac{1 - t \cos \psi + D}{2} \right] \right\},
\]
\[
t = R/r, \quad D = (1 - 2t \cos \psi + r^2)^{1/2}.
\]
Assume that we have $n$ measurements of $T_{rr}$. The matrix form of the observation equation (21b) becomes
\[
y = Ax + \epsilon, \tag{21c}
\]
where $y$ is the observational vector with elements $4\pi R/10T_{rr}$, $A$ is the design matrix, $x$ is the vector of unknown surface gravity anomalies to be determined and $\epsilon$ is the observational error vector of $y$. Since the major purpose of this paper is to establish a new quality-motivated TSVD method by minimizing the MSE and to compare it with the other TSVD techniques, we shall confine ourselves to an area of $10^\circ \times 10^\circ$ centred at longitude $115^\circ$ and latitude $35^\circ$. The unknowns are $100\ 1^\circ \times 1^\circ$ block-mean anomalous gravity values. The number of satellite gravimetric observables $T_{rr}$ is 400 with a sampling interval of 30’, located directly above the area of interest. The accuracy of $T_{rr}$ is set to 0.01 $E(1E = 10^{-9} \text{ s}^{-2})$, which was used in the simulations of Koop (1993) and Xu & Rummel (1994), and may also be comparable with a more recent simulation for
GOCE (Mueller et al. 1996). The eigenvalue spectrum is shown in Fig. 1. The condition number of the example is about $1.06 \times 10^4$.

In these simulations, the 100 mean gravity anomalies were computed from an earth model and taken as true values. Thus the corresponding true observations can be simulated. The errors $y_i$ of $y$ in (21c) are generated by $4\pi R/10 N(T_{rr})$, where $N(T_{rr})$ is a Gaussian random number generator with mean zero and variance $10^{-4}E^2$. A scientifically objective way to assess the performance of an estimation method is through large-scale simulations (Xu & Rummel 1994). In this paper, we shall take the same simulation approach as in Xu & Rummel (1994). More specifically, we repeat the solutions of the mean gravity anomalies 1000 times, each using 400 different Gaussian random numbers generated by $N(T_{rr})$.

### 4.1 Criteria for comparing the six TSVD estimators

The condition number of the normal matrix in the discrete linear problem (1) is an important indicator of how ill-posed (or unstable) the inverse problem (1) is. The more ill-posed the linear problem (1), the bigger the condition number. On the other hand, it can be easily proven that the condition number of the cofactor matrix of the LS estimate of $\mathbf{x}$ is equal to that of the normal matrix. Thus a reasonable measure of instability of a TSVD estimator can be naturally defined by the condition number of its error variance–covariance matrix (16), i.e.

$$\kappa = \frac{T_1^2}{T_k^2}. \quad (22)$$

The concept of error variance plays an important role in judging the goodness of an estimator, if it is unbiased.

![Figure 1. The eigenvalues of the example.](image1.png)

Unfortunately, all TSVD estimators are almost always biased. Thus the bias extent of a TSVD estimator should serve as a second criterion in the comparison of different estimators. By combining the first two criteria, we are led to an intuitively appealing, quality-motivated criterion—mean squared error, which has been defined by (17).

### 4.2 Numerical comparison in solution stability

The solution stability of ill-posed problems has been one of the important quality measures in judging alternatives to the LS method. It will be investigated in terms of the number of components cut by different truncation criteria and the condition numbers of the remaining systems in deriving the TSVD estimators. The number of components cut should indicate roughly how much information cannot be derived from observations using the $F$-statistic-based TSVD estimators and the smoothness extent in the case of TSVD estimators by means of the L-curve and quality-motivated MSE. The numbers of cut components corresponding to the six TSVD estimators over the 1000 repetitions are plotted in Fig. 2, and their statistics are summarized in Table 1. The numbers of cut components are very stable in the simulations (see Fig. 2), which is confirmed by the relatively small standard deviations in the third row of Table 1. It is fairly clear that for the $F$-statistic-based TSVD estimators, the numbers of components cut are approximately 37, 29 and 21 for $x_{pc5}$, $x_{pc16}$ and $x_{pc35}$, respectively. They increase with decreasing significance level, which is expected theoretically. The TSVD estimator by means of an L-curve discards about 90 per cent of the independent model parameter combinations [compare with the value of 67.2 per cent in

![Figure 2. The numbers of components cut by the TSVD estimators over 1000 repetitions.](image2.png)

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$x_{pc5}$</th>
<th>$x_{pc16}$</th>
<th>$x_{pc35}$</th>
<th>$x_{hk}$</th>
<th>$x_{km}$</th>
<th>$x_{ls}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of components cut</td>
<td>36.67</td>
<td>29.11</td>
<td>20.73</td>
<td>90.46</td>
<td>4.16</td>
<td>23.38</td>
</tr>
<tr>
<td>Std of numbers of components cut</td>
<td>2.61</td>
<td>3.00</td>
<td>3.19</td>
<td>0.67</td>
<td>4.04</td>
<td>4.87</td>
</tr>
<tr>
<td>Mean condition number</td>
<td>2811.07</td>
<td>4885.65</td>
<td>7016.18</td>
<td>1.99</td>
<td>5049.56</td>
<td>466.98</td>
</tr>
<tr>
<td>Std of condition numbers</td>
<td>2393.81</td>
<td>2932.42</td>
<td>2881.49</td>
<td>0.23</td>
<td>3183.70</td>
<td>189.53</td>
</tr>
</tbody>
</table>
Hansen et al. (1992)], which is extremely high compared to the other TSVD estimators. This indicates that the TSVD estimator by means of an L-curve has oversmoothed the ill-posed downward continuation problem. The TSVD estimator by minimizing the MSE with the LS estimate of the parameters results in a significantly small number of components cut, 4.16. The reason for this is that when the LS estimate of \( \mathbf{x} \) is used to estimate the biases in (15), \( \text{bias}(\hat{\mathbf{x}}) \) becomes

\[
\text{bias}(\hat{\mathbf{x}}) = -\mathbf{V}_H \hat{\mathbf{a}}_H,
\]

which will be very poorly determined, because \( \hat{\mathbf{a}}_H \) corresponds directly to the smallest eigenvalues and is most poorly resolved by the LS method. Thus, a best trade-off between bias and stability cannot be reached. The MSE-based TSVD estimator with the basic ridge estimator as its initial value, on average (over 1000 repetitions), 23 combinations corresponding to smallest eigenvalues.

Although the number of components cut is instructive, it does not provide us with exact information about the condition numbers in the TSVD estimators. The condition numbers over the 1000 repetitions for the six TSVD estimators are plotted in Fig. 3. We also see from Table 1 that for the \( F \)-statistic-based TSVD estimators, the condition numbers increase with decreasing significance level. In particular, the maximum condition number of the original ill-posed problem has unfortunately been reached frequently by these three estimators and the MSE-based TSVD estimator with the LS estimate as their initial values (see Fig. 3), which probably also explains the large standard deviations (see the last row of Table 1) of their condition numbers in the simulations. This clearly indicates that \( F \)-statistic-based TSVD estimators, as often investigated in statistical literature (see e.g. Dempster et al. 1977; Lawless & Wang 1976), cannot really help to improve the solution stability. In other words, the \( F \)-statistic criterion can frequently lead to incorrect cuttings of components. For a particular data realization, \( F \)-statistic-based TSVD estimators can perform very poorly. The LS estimate of the model parameters is not suitable for the derivation of an MSE-based TSVD estimator. The MSE-based TSVD estimator with the basic ridge estimate of \( \mathbf{x} \) significantly improves the condition number of the ill-posed problem. The TSVD estimator by means of an L-curve has an extremely small condition number (a mean of 1.99), which, together with a large number of cut components, reaffirms that the ill-posed problem has been oversmoothed.

### 4.3 Numerical comparisons of solution bias and mean squared error

In these simulations, we repeated the computations 1000 times for each TSVD estimator using different sets of random numbers. The results can be represented by a data matrix, i.e. \( \mathbf{Y}_i = (\hat{\mathbf{x}}_{i,1}, \hat{\mathbf{x}}_{i,2}, \ldots, \hat{\mathbf{x}}_{i,1000}^i) \), where the subscript \( i \) stands for one of the six TSVD estimators and the superscripts (1, 2, \ldots, 1000) are the sequential numbers of the repeated computations. Therefore, the average value of \( \hat{\mathbf{x}}_{i}^j \) over the 1000 experiments for the \( i \)th TSVD estimator can be obtained from \( \mathbf{Y}_i \), and is given by

\[
\hat{\mathbf{x}}_i^j = \frac{1}{1000} \sum_{j=1}^{1000} \hat{\mathbf{x}}_{i,j}^j,
\]

which indicates the long term performance of the \( i \)th TSVD estimator, since the experimental set-ups are identical.

The average biases of \( \hat{\mathbf{x}}_i^j \) are computed by

\[
\text{bias}(\hat{x}_i^j) = \hat{x}_i^j - \mathbf{x}.
\]

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are computed by

\[
\text{bias}(\hat{x}_i^j) = \hat{x}_i^j - \mathbf{x}.
\]

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.

The six bias vectors \( \text{bias}(\hat{x}_i^j) \) are given in Fig. 4. For the \( F \)-statistic-based TSVD estimators, the larger a significance level, the smaller the biases of the estimated parameters, since a larger significance level always prevents more linear combinations of \( \mathbf{x} \) from being discarded. The absolute values of the mean biases for all the model parameters, listed in Table 2, are 1.94, 0.99 and 0.39 mgal for \( \hat{x}_{16,5}, \hat{x}_{16,16} \) and \( \hat{x}_{35,5} \), respectively.
oversmoothed, to an extreme extent, the solution of the ill-posed problem. The MSE-based TSVD estimator using the basic ridge estimate gives a reasonable bias of 3.42 mgal, which is still somewhat too large from the point of view of a space gradiometry mission. If the largest seven biases are left out of the computation, the average absolute bias drops to about 2.5 mgal. Further research is expected to remove or reduce a few of the largest biased points, and hopefully also biases of 5 mgal or more. We then compute the mean MSE root for each TSVD estimator by using only the 95 per cent of the estimated model parameters with the smallest MSE roots. \( \hat{\mathbf{x}}_{p,16}, \hat{\mathbf{x}}_{p,35}, \hat{\mathbf{x}}_{kl,5} \) and \( \hat{\mathbf{x}}_{ls,5} \) result in mean MSE roots of 6.84, 7.23, 6.97 and 6.25 mgal, respectively, which certainly cannot satisfy the 5 mgal requirement or target for a space gradiometry mission. Only about 10 per cent of the estimated parameters have an MSE root equal to or smaller than 5 mgal, as seen in Table 2.

Of the six TSVD estimators, that using the L-curve has the poorest performance in terms of mean squared error, with a mean MSE root of 15.25 mgal. Although 18 per cent of the estimated parameters have an accuracy better than 5 mgal, an overwhelming majority (70 per cent) of the estimated parameters have an accuracy worse than 8 mgal. The MSE-based TSVD estimator with the basic ridge estimate as its initial value produces the best results, with a mean MSE root of 4.90 mgal. 61 per cent of the estimated model parameters have an accuracy better than 5 mgal, and 92 per cent have an MSE root smaller than 8 mgal, as seen in Table 2. Of the six TSVD estimators, that using the L-curve has the poorest performance in terms of mean squared error, with a mean MSE root of 15.25 mgal. Although 18 per cent of the estimated parameters have an accuracy better than 5 mgal, an overwhelming majority (70 per cent) of the estimated parameters have an accuracy worse than 8 mgal. The MSE-based TSVD estimator with the basic ridge estimate as its initial value produces the best results, with a mean MSE root of 4.90 mgal. 61 per cent of the estimated model parameters have an MSE root smaller than 8 mgal, and 92 per cent have an MSE root smaller than 8 mgal. If the eight largest biases (see Table 2) are not used in the computation, the mean accuracy of 4.4 mgal, which is not achievable by any of the other TSVD estimators tested or the LS method.

5 CONCLUSIONS

TSVD techniques have been widely applied in inversion, but the truncation parameter \( k \) was more or less arbitrarily decided. No solid foundation has been available to help find the truncation parameter from the point of view of the quality of a TSVD solution. We have thus proposed a quality-based TSVD method for the first time in this paper. In order to compare the performance of this quality-based TSVD method with the other two TSVD truncation criteria, we have investigated six TSVD estimators for the regularization of discrete ill-posed problems, based on the statistically frequently used \( F \)-statistic, the L-curve of much recent interest in interdisciplinary inversion, and our new quality-motivated MSE criteria. The three \( F \)-statistic-based TSVD estimators are shown to improve the LS solution to the ill-posed downward continuation problem marginally in the sense of long-term averaging, depending on (arbitrarily) pre-selected significance levels. At present, there exist no practical guidelines for selecting a reasonable significance level in order to obtain a quality solution to the ill-posed problem. The simulations have also shown that estimators of this type cannot guarantee an
improvement in the condition number of an ill-posed linear system. Very often, the maximum condition number inherited from the original ill-posed problem has been reached. In other words, the F-statistic criterion can frequently lead to incorrect cuttings of components, which leads to the failure of trying to improve solution stability, and makes it almost impossible to achieve a good balance among the stability, the bias and the MSE of the estimated solution. For any particular data realization, the effect of this incorrect cutting of components on the solution to an ill-posed problem is very negative.

The TSVD technique by means of an L-curve was recently proposed to minimize simultaneously the norm of the residuals of observations and a (semi-)norm of the model parameters. It has since been widely applied to a number of inverse problem areas, even though its quality aspect has not yet been clarified. Our quality investigations (simulations) have shown that the TSVD estimator by means of an L-curve is the best in (over-)stabilizing the ill-posed problem but results in an over-cutting of components. The condition number of around 3 has been kept over the 1000 repetitions and about 90 per cent of the model parameters have been cut, which clearly indicates that the ill-posed problem under study has been oversmoothed by the TSVD estimator using an L-curve. In fact, the L-curve-based TSVD estimator has been emphasized for stabilizing solutions to ill-posed problems (see e.g. Gough 1996; Hansen et al. 1992). It is extremely poor in terms of biases and MSE roots of the solutions, which are too large and practically meaningless. The quality-based TSVD method by minimizing the MSE of a solution with the basic ridge estimate of x as its initial value has been shown clearly to outperform the F-statistic and L-curve criteria. It provides the best solutions in terms of stability, bias and mean squared error. The simulations have also shown that the mean accuracy of the quality-based TSVD can be better than 5 mgal, if a few of the largest biased points are left out of the computation. Eliminating and/or reducing a few of the largest biases is expected to be an achievable target for future theoretical improvements.

ACKNOWLEDGMENTS

Mr Nico Sneeuw at the Technical University of Munich is very much appreciated for his information on the recent progress in and mission possibilities of space satellite gravimetry. The comments of an anonymous reviewer helped to clarify the experiment and correct some misprints, for which I am grateful.

REFERENCES


© 1998 RAS, GJI 135, 505–514


Downloaded from https://academic.oup.com/gji/article-abstract/135/2/513/779202 by guest on 20 February 2019