Robust satisficing linear regression: Performance/robustness trade-off and consistency criterion

Miriam Zacksenhouse [a,*], Simona Nemets [a], Mikhail A. Lebedev [b], Miguel A.L. Nicolelis [b]

[a] *Faculty of Mechanical Engineering, Technion-Israel Institute of Technology, Haifa, Israel*
[b] *Department of Neurobiology and Center for Neuro-engineering, Duke University, Durham, NC, USA*

A B S T R A C T

Linear regression quantifies the linear relationship between paired sets of input and output observations. The well known least-squares regression optimizes the performance criterion defined by the residual error, but is highly sensitive to uncertainties or perturbations in the observations. Robust least-squares algorithms have been developed to optimize the worst case performance for a given limit on the level of uncertainty, but they are applicable only when that limit is known. Herein, we present a *robust-satisficing* approach that maximizes the robustness to uncertainties in the observations, while satisficing a critical sub-optimal level of performance. The method emphasizes the trade-off between performance and robustness, which are inversely correlated. To resolve the resulting trade-off we introduce a new criterion, which assesses the consistency between the observations and the linear model. The proposed criterion determines a unique robust-satisficing regression and reveals the underlying level of uncertainty in the observations with only weak assumptions. These algorithms are demonstrated for the challenging application of linear regression to neural decoding for brain-machine interfaces. The model-consistent robust-satisfying regression provides superior performance for new observations under both similar and different conditions.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Linear regression is a classical inverse problem where the parameters of a linear model that relate the dependent variables to the independent variables need to be estimated from a set of observations [1,2]. The problem is complicated by three major sources of uncertainties: (i) measurement uncertainty, which accounts for inaccuracies in the observations, (ii) model uncertainty, which accounts for possible non-linear effects, and (iii) temporal uncertainty, which accounts for potential changes that are not present in the available observations.

A major goal of linear regression is to use the estimated parameters to predict the dependent variable from new observations of the independent variables. The performance criterion of interest in this case is the norm of the residual error, which is estimated from the available set of observations. The well known least-squares regression is derived by minimizing the residual norm. However, given the above uncertainties, the resulting least-squares regression may fail to provide high, or even acceptable, performance for new observations [3]. An alternative criterion is based on the combination of the residual norm and the weighted regression norm [1–5]. The weight of the regression norm is referred to as the regularization parameter: increasing the regularization parameter reduces the sensitivity to uncertainties in the

---

* Corresponding author. Tel.: +972 4 8292092; fax: +972 4 8295711.
 E-mail address: mermz@tx.technion.ac.il (M. Zacksenhouse).

observations at the expense of increasing the residual norm. Thus regularized regression depends on properly choice of the regularization parameter to balance this trade-off [4].

The choice of the regularization parameter depends on the distribution of the singular values of the observation matrix A, whose columns describe the independent variables [4]. In general, when the condition number of the observation matrix A is large, the least squares solution is greatly affected by uncertainties in the observations and may differ considerably from the underlying linear model [3]. If there is a clear gap in the spectrum of the singular values, i.e., A is rank deficient, the regularization may be based on truncated singular value decomposition (tSVD) [5], and the rank of A can be used to determine the regularization parameter. However, if the singular values of A decay gradually, the problem is ill-posed and the choice of the regularization parameter is more complicated. For ill-posed problems there are two classes of parameter choice methods:

(a) Methods based on knowledge or good estimate of the uncertainties, including for example Morozov's discrepancy principle [4], and the robust min-max approach [6,7].
(b) Methods which assess the level of uncertainty given the measurements, including for example, the generalized cross validation, and the L-curve [4].

When the uncertainty in the observations is bounded by a known limit, a robust least-squares regression can be determined using the min–max approach, which minimizes the worst case residual norm [6,7] Under a specific set of assumptions, the robust least-squares regression has been shown to have the form of a Tikhonov regularized regression [1] with a regularization parameter that depends on the presumed bound on the level of uncertainty in the observations. Hence, the method substitutes the choice of the Tikhonov regularization parameter with the assessment of the bound on the level of uncertainty in the observations.

The L-curve has been suggested as a method for choosing the regularization parameter when bound on the level of uncertainties or perturbations is not known [8]. However, the L-curve may loose its characteristic L-shape in the presences of large uncertainties and thus the method may fail to determine the appropriate regularization parameter (see Section 5). Consequently, the choice of the Tikhonov regularization parameter under uncertainties remains a challenge.

Here we address the case where the bound on the uncertainty in the observations is unknown. In Section 2, we describe the uncertainty in the observations by an information-gap (info-gap) model, which is parameterized by the level of uncertainty. Instead of optimizing performance, we focus on maximizing the level of uncertainty under which a critical level of performance can be guaranteed. This approach has been termed robust-satisficing, where satisficing (coined by Simon [9]) refers to satisfying a sufficient level of performance [10].

The info-gap approach emphasizes the robustness/performance trade-off detailed in Section 3: high robustness to uncertainties can be achieved only by relinquishing performance. To determine the appropriate trade-off, Section 4 introduces a new criterion, which assesses the consistency between the observations and the robust-satisficing regression. The model-consistency criterion determines a unique regression and reveals the appropriate performance and robustness trade-off. The effectiveness of the model-consistent robust-satisficing method is demonstrated in Section 5 for the challenging application of linear regression to neural decoding in brain-machine interfaces (BMI), where the L-curve method is inadequate due to the large level of uncertainty.

2. Robust satisficing regression

2.1. Information-gap models of uncertainty

We consider a linear regression between the dependent variable b and the (column) vector of independent variables $\underline{a} \in \Re^K$, computed from a set of N available observations $\{\underline{\tilde{a}}_i, \tilde{b}_i\}_{i=1}^N$, of the independent and dependent variables, respectively. The estimation of the regression is complicated by potential deviations between the available observations $\{\underline{\tilde{a}}_i, \tilde{b}_i\}_{i=1}^N$ and the actual values of the underlying variables $\{\underline{a}_i, b_i\}_{i=1}^N$ that are expected to be related linearly. The latter are referred to as ideal observations—observations that would have been obtained under no uncertainties.

For convenience, the available observations are arranged into a $N \times K$ observation matrix $\tilde{A} = [\underline{\tilde{a}}_1, \underline{\tilde{a}}_2, \ldots, \underline{\tilde{a}}_N]^T$, and an observation vector $\underline{\tilde{b}} = [\tilde{b}_1, \tilde{b}_2, \ldots, \tilde{b}_N]^T$. Given a regression vector $\underline{x} \in \Re^K$, the common criterion for its performance is the 2-norm of the residual error $r(\underline{x}; \tilde{A}, \underline{\tilde{b}}) \equiv \|\tilde{A}\underline{x} - \underline{\tilde{b}}\|_2$. However, when there are uncertainties in the observations, the criterion of interest is $r(\underline{x}; A, \underline{b}) \equiv \|A\underline{x} - \underline{b}\|_2$ based on the ideal observation matrix $A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_N]^T$ and vector $\underline{b} = [b_1, b_2, \ldots, b_N]^T$ of the linearly related variables. In order to assess the relation between these two criteria, we need to specify the set of ideal observations that might have resulted in the available observations, given the level of uncertainty. We consider models of the form

$$U(\alpha : \tilde{A}, \underline{\tilde{b}}) = \{\underline{b} \in \Re^N, A \in \Re^{N \times K} : \|A - \tilde{A}\|_2 \leqslant \alpha, \| \underline{b} - \underline{\tilde{b}} \|_2 \leqslant \nu\alpha\} \tag{1}$$

where $\alpha \geqslant 0$ is the level of uncertainty in the observation matrix A and $\nu \geqslant 0$ is the relative level of uncertainty in $\underline{b}$. The 2-norm is used as the vector-norm and the induced 2-norm as the matrix norm. The index-2 that specifies the nature of the norm would be omitted except for the definitions.