Published in final edited form as:

Mech Syst Signal Process. 2009 August ; 23(6): 1954–1964. doi:10.1016/j.ymssp.2008.09.008.

Robust Satisficing Linear Regression: performance/robustness trade-off and consistency criterion

Miriam Zacksenhouse^{1,*}, Simona Nemets¹, Mikhail A. Lebedev², and Miguel A.L. Nicolelis²

¹ Faculty of Mechanical Engineering, Technion – Israel Institute of Technology, Haifa, Israel

² Dept of Neurobiology and Center for Neuro-engineering, Duke Univ., Durham, NC, USA

Abstract

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-satisficing regression provides superior performance for new observations under both similar and different conditions.

Keywords

Linear regression; Robust regression; Regularization; Information-gap; Uncertainties; Brain machine interface

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

*Corresponding Author: mermz@tx.technion.ac.il, Phone: +972-4-8292092, Fax: +972-4-8295711.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

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 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 the 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 V). 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 II, 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 H. Simon [9]) refers to satisfying a sufficient level of performance [10].

The info-gap approach emphasizes the robustness/performance trade-off detailed in section III: high robustness to uncertainties can be achieved only by relinquishing performance. To determine the appropriate trade-off, Section IV 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 V 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 $\underline{a} \in \mathbb{R}^K$, computed from a set of N available observations $\{\tilde{\underline{a}}_i, \tilde{b}_i\}_{i=1}^N$ of the independent and dependent variables, respectively. The estimation of the regression is complicated by potential deviations between the available observations $\{\tilde{\underline{a}}_i, \tilde{b}_i\}_{i=1}^N$ 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 $\tilde{A} = [\tilde{\underline{a}}_1, \tilde{\underline{a}}_2, \dots, \tilde{\underline{a}}_N]^T$, and an observation vector $\tilde{\underline{b}} = [\tilde{b}_1, \tilde{b}_2, \dots, \tilde{b}_N]^T$. Given a regression vector $\underline{x} \in \mathbb{R}^K$, the common criterion for its performance is the 2-norm of the residual error $r(\underline{x}; \tilde{A}, \tilde{\underline{b}}) \equiv \|\tilde{A}\underline{x} - \tilde{\underline{b}}\|_2$. 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 $A = [\underline{a}_1, \underline{a}_2, \dots, \underline{a}_N]^T$ and vector $\underline{b} = [b_1, b_2, b_N]^T$ 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:

$$U(\alpha; \tilde{A}, \tilde{\underline{b}}) = \{\underline{b} \in \mathbb{R}^N, A \in \mathbb{R}^{N \times K} : \|A - \tilde{A}\|_2 \leq \alpha, \|\underline{b} - \tilde{\underline{b}}\|_2 \leq \nu\alpha\} \quad (1)$$

Where $\alpha \geq 0$ is the level of uncertainty in the observation matrix A and $\nu \geq 0$ 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.

We note that Eq. (1) defines a nested and unbounded set of ideal observations that might have resulted in the available observations, under increasing levels of uncertainty. Nesting implies that ideal observations associated with larger levels of uncertainty include ideal observations associated with smaller levels of uncertainty, while only the available observations are possible when there is no uncertainty. Unboundedness implies that the norm of the deviation from the available observations may increase without bound as the level of uncertainty increases. Since the worst case depends on the level of uncertainty, unboundedness implies that the worst case is not known. Such a structure is referred to as an info-gap (short for information-gap) model of uncertainty [10].

The model in (1) has been used in [7] and a similar model with the Frobenius norm was used in [6]. Another model, which bounds the uncertainty in each row of the observation matrix A and vector $\underline{b}$, is considered in [11]. Herein we assume that the relative weight ν is known (but not necessarily one) and capture the uncertainty with the single parameter α .

2.2 Robustness and Robust-Satisficing Regression

Minmax regression minimizes the maximum residual norm under a presumed level of uncertainty and thus depends on knowing that level. Furthermore, concentrating on the worst case, which may be very unlikely, may result in poor performance under more likely cases. Instead we propose the robust-satisficing approach, which focuses on the required level of performance without making any assumptions about the level of uncertainty. This approach is outlined in the next two definitions.

Definition 1 – Robustness and robustness curve—Given a critical performance ρ , the robustness of the regression $\underline{x}$ to uncertainties in the observations is the largest level of uncertainty α up to which all possible ideal observations result in residual norms that are smaller than ρ :

$$\widehat{\alpha}(\rho; \underline{x}) \equiv \max_{\alpha} \left(\alpha \mid \max_{A, \underline{b} \in U(\alpha; \tilde{A}, \tilde{\underline{b}})} \|A\underline{x} - \underline{b}\|_2 \leq \rho \right) \quad (2)$$

The robustness is zero if the critical performance cannot be achieved even under zero uncertainty. For a given regression $\underline{x}$, the robustness as a function of the critical performance ρ is termed the *robustness curve*.

Definition 2 – Robust-satisficing regression—Given a sub-optimal critical performance ρ , the robust satisficing (RS) regression is the regression that maximizes the robustness for satisficing ρ :

$$\underline{x}_{rs}(\rho) \equiv \arg \max_{\underline{x}} \widehat{\alpha}(\rho; \underline{x}) \quad (3)$$

The next Proposition evaluates the robustness when the uncertainties are described by the info-gap model of Eq. (1):

Proposition 1 – Robustness—Given the available observations $\tilde{A}, \tilde{\underline{b}}$, the robustness of the regression $\underline{x}$ for achieving the critical performance ρ under uncertainties in the observations described by the info-gap model of Eq. (1) is:

$$\widehat{\alpha}(\rho; \underline{x}) = \max \left(0, \frac{\rho - \|\tilde{A}\underline{x} - \tilde{\underline{b}}\|}{\|\underline{x}\| + \nu} \right) \quad (4)$$

Proof: The internal maximization in Eq. (2) defines the worst residual norm

$\widehat{r}(\alpha; \underline{x}) \equiv \max_{A, \underline{b} \in U(\alpha; \tilde{A}, \tilde{\underline{b}})} \|A\underline{x} - \underline{b}\|_2$ that may occur with the regression $\underline{x}$ when the level of uncertainty is α . It can be shown ([7], see also Lemma B for detailed analogous derivation of the best residual norm) that the worst residual norm is given by:

$$\widehat{r}(\alpha; \underline{x}) = \|\tilde{A}\underline{x} - \tilde{\underline{b}}\| + \alpha(\|\underline{x}\| + \nu) \quad (5)$$

Substituting (5) in (2), completes the proof.

The first term in Eq. (5) describes the nominal residual error $r_{\text{nom}}(\underline{x}) = \|\tilde{A}\underline{x} - \tilde{\underline{b}}\|_2$ when there is no uncertainty, whereas the second term describes the effect of uncertainty. Thus, a critical performance that equals the nominal residual error $\rho = r_{\text{nom}}(\underline{x})$, can be achieved only with zero robustness. Positive robustness can be attained only for $\rho > r_{\text{nom}}(\underline{x})$, i.e., by relinquishing performance, and in that range the robustness curve increases linearly.

Examples of robustness curves for the application described in Section V are shown in Figure 1. Different robustness curves correspond to different Tikhonov regressions $\underline{x}_\mu(\tilde{A}, \tilde{\underline{b}}) \equiv (\tilde{A}^T \tilde{A} + \mu I)^{-1} \tilde{A}^T \tilde{\underline{b}}$ with different values of the Tikhonov parameter μ . The Tikhonov regression with $\mu = 0$ corresponds to the optimal least-squares (LS) regression $\underline{x}_{LS} = \underline{x}_{\mu=0}(\tilde{A}, \tilde{\underline{b}}) \equiv (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T \tilde{\underline{b}}$, and achieves the optimal performance $r_{\text{nom}}(\underline{x}_{LS}) \equiv \|\tilde{A}\underline{x}_{LS} - \tilde{\underline{b}}\|$. The LS regression minimizes the nominal residual norm, but achieves this critical optimal performance, i.e., $\rho = r_{\text{nom}}(\underline{x}_{LS})$, with zero robustness.

Curve crossing: The above discussion implies that the robustness curve of a sub-optimal regressions $\underline{x} \neq \underline{x}_{LS}$ becomes positive at $\rho > r_{\text{nom}}(\underline{x}) > r_{\text{nom}}(\underline{x}_{LS})$. Furthermore, when $\|\underline{x}\| < \|\underline{x}_{LS}\|$, the slope of the robustness curve of $\underline{x}$ is larger than the slope of the robustness curve of the LS regression, so the two robustness curves cross each other (see Figure 1). Hence, a sub-optimal regression with $\|\underline{x}\| < \|\underline{x}_{LS}\|$ (e.g., Tikhonov regression with positive Tikhonov parameter) provides a higher robustness for some sub-optimal performance $\rho > r_{\text{nom}}(\underline{x})$, as detailed in the next proposition.

Proposition 2 – Robust satisficing regression—For critical sub-optimal performance $\rho > \|\tilde{A}\underline{x}_{LS} - \tilde{\underline{b}}\|$, the robust-satisficing regression $\underline{x}_{RS}(\rho)$ has the form of a Tikhonov regression $\underline{x}_\mu(\tilde{A}, \tilde{\underline{b}}) \equiv (\tilde{A}^T \tilde{A} + \mu I)^{-1} \tilde{A}^T \tilde{\underline{b}}$ with the parameter $\mu = \hat{\mu}$, which satisfies:

$$\hat{\mu} = \frac{\|\tilde{A}\underline{x}_{\hat{\mu}} - \tilde{\underline{b}}\|}{\|\underline{x}_{\hat{\mu}}\|} \left(\frac{\rho - \|\tilde{A}\underline{x}_{\hat{\mu}} - \tilde{\underline{b}}\|}{\|\underline{x}_{\hat{\mu}}\| + \nu} \right) \quad (6)$$

Proof: At a fixed critical performance ρ , the gradient of the robustness given by (4), with respect to the regression $\underline{x}$, can be expressed as:

$$\nabla_{\underline{x}} \hat{\alpha}(\underline{x}; \rho) = \frac{-\tilde{A}^T (\tilde{A}\underline{x} - \tilde{\underline{b}})(\|\underline{x}\| + \nu)\|\underline{x}\| - \underline{x} \left(\rho - \|\tilde{A}\underline{x} - \tilde{\underline{b}}\| \right) \|\tilde{A}\underline{x} - \tilde{\underline{b}}\|}{\|\tilde{A}\underline{x} - \tilde{\underline{b}}\| \|\underline{x}\| (\|\underline{x}\| + \nu)^2}. \quad (7)$$

The gradient vanishes when $\underline{x}$ has the form of a Tikhonov regression with the parameter specified by Eq. (6).

Assuming a non singular case, (i.e., $0 < \|\tilde{A}\underline{x} - \tilde{\underline{b}}\|$ for all $\underline{x}$), the robustness reaches a **global** maximum at the point where the gradient vanishes (Lemma A1 in Appendix A). Hence, there is a unique $\hat{\mu}$ that satisfies Eq. (6), and the RS regression defined by Eq. (3) has the form of a Tikhonov regression with the Tikhonov parameter $\hat{\mu}$.

Proposition 2 indicates that for each critical performance ρ within the specified range, the RS regression has the form of a Tikhonov parameter with a parameter given implicitly by (6). The opposite is also true, as stated in Proposition 3.

Proposition 3 – Tikhonov regression and robust satisficing—A Tikhonov regression $\underline{x}_\mu (\tilde{A}, \tilde{b}) \equiv (\tilde{A}^T \tilde{A} + \mu I)^{-1} \tilde{A}^T \tilde{b}$ with parameter μ is the RS regression for the critical performance:

$$\rho_{RS}(\mu) = \left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\| + \frac{\mu \left\| \underline{x}_\mu \right\| (\left\| \underline{x}_\mu \right\| + \nu)}{\left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\|} \quad (8)$$

Proof: We note that $\rho_{RS}(\mu)$ specified in Eq. (8) is always larger than the optimal residual norm, i.e., $\rho_{RS}(\mu) \geq \left\| \tilde{A} \underline{x}_{LS} - \tilde{b} \right\|$, since the optimal performance is the minimum of the first term and the second term is always positive. According to Proposition 2, the RS regression for a critical performance $\rho > \left\| \tilde{A} \underline{x}_{LS} - \tilde{b} \right\|$ is a Tikhonov regression with the parameter $\mu = \hat{\mu}$ that satisfies Eq. (6). Substituting $\rho = \rho_{RS}(\mu)$ given by Eq. (8), it is easily verified that Eq. (6) is satisfied for $\hat{\mu} = \mu$. Hence the corresponding RS regression for $\rho = \rho_{RS}(\mu)$ is the Tikhonov regression with the parameter μ .

III. Robustness/Performance trade-off

Eq. (8) specifies explicitly the critical performance for which a Tikhonov regression is the RS regression. We next specify the resulting robustness:

Proposition 4 – Robustness of a Tikhonov regression—Given a Tikhonov regression $\underline{x}_\mu$, its robustness for obtaining the critical performance $\rho = \rho_{RS}(\mu)$ (Eq. (8)), for which it is the RS regression, is:

$$\hat{\alpha}_{RS}(\mu) \equiv \hat{\alpha}_{RS}(\rho_{RS}(\mu); \underline{x}_\mu) = \frac{\mu \left\| \underline{x}_\mu \right\|_2}{\left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\|_2} \quad (9)$$

Furthermore, the RS robustness $\hat{\alpha}_{RS}(\rho_{RS}(\mu); \underline{x}_\mu)$ is a non-decreasing function of the critical performance $\rho = \rho_{RS}(\mu)$.

Proof: Substituting Eq. (8) in Eq. (4), results in Eq. (9) for the robustness. Equation (4) indicates that the robustness of a fixed regression is a non-decreasing function of the critical performance. The robustness of the RS regression is the maximum robustness that can be achieved for a given critical performance, so it is also a nondecreasing function of the critical performance ρ .

Equations (8) and (9) specify the envelope of the robustness curves as shown in Figure 1, which depicts the RS robustness $\hat{\alpha}_{RS}$ as a function of the critical performance ρ . The circles on each of the robustness curves indicate the points at which the corresponding Tikhonov regressions are the RS regressions, with the critical performance given by Eq. (8), and the resulting robustness by Eq. (9).

In agreement with Proposition 4, the RS robustness $\hat{\alpha}_{RS}$ is a non-decreasing function of the critical performance. Hence, the family of RS regressions establishes a trade-off between the critical performance and the robustness: Higher robustness can be achieved only by relinquishing performance (increasing the critical performance), while better performance (lower critical performance) can be obtained only with lower robustness.

The performance/robustness trade-off defined by (8) and (9) describes how the Tikhonov parameter affects both the performance and the robustness, and thus provides important insight for the choice of the proper Tikhonov parameter. However, the trade-off indicates that neither the robustness nor the performance criterion provides a unique solution. An additional criterion, which assesses the consistency between the RS regression model and the level of uncertainty in the observations, is developed in the next section. It is shown that the model-consistency criterion defines a unique regression.

IV. Model Consistency

The level of uncertainty in the observations affects not only the worst case performance (as described in Section II and III) but also the best case performance. Given the regression $\underline{x}$, the smallest residual norm that can be achieved under the level of uncertainty α , is defined by:

$\widehat{s}(\alpha; \underline{x}) \equiv \max_{A, \underline{b} \in U(\alpha; A, \underline{b})} \|A\underline{x} - \underline{b}\|_2$. The smallest residual norm assesses how well an ideal observation within the level of uncertainty α , may match the linear model with the regression $\underline{x}$. In particular, $\widehat{s}(\alpha; \underline{x}) = 0$ implies that there is an ideal observation within the info-gap model with uncertainty α that matches the linear model defined by the regression $\underline{x}$ exactly. The strategy proposed in this section is based on evaluating the smallest residual norm of the RS regressions when the level of uncertainty is the corresponding robustness.

The best performance that can be obtained by the regression $\underline{x}$ when there is no uncertainty is the nominal performance $r_{nom}(\underline{x}) = \|\tilde{A}\underline{x} - \tilde{b}\|_2$. Better than nominal performance $\rho < r_{nom}(\underline{x})$ is referred to as windfall performance. Windfall performance cannot be achieved robustly, but as the level of uncertainty increases, the opportunity for its occurrence increases, leading to the notion of opportuneness [10] defined next:

Definition 3 – Opportuneness—The opportuneness for obtaining a windfall performance ϕ , with the regression $\underline{x}$ is the smallest level of uncertainty for which the info-gap model includes an ideal observation with a smaller or equal residual norm:

$$\widehat{\beta}(\rho; \underline{x}) \equiv \min_{\alpha} \left(\alpha \mid \min_{A, \underline{b} \in U(\alpha; A, \underline{b})} \|A\underline{x} - \underline{b}\|_2 \leq \phi \right) \quad (10)$$

For a given regression $\underline{x}$, the opportuneness as a function of the windfall performance ϕ is termed the *opportuneness curve*.

The next Proposition evaluates the opportuneness when the uncertainties are described by the info-gap model of Eq. (1):

Proposition 5 – Opportuneness—Given the available observations $\tilde{A}$, $\tilde{b}$, the opportuneness of the regression $\underline{x}$ for achieving the windfall performance ϕ under uncertainties in the observations described by the info-gap model of Eq. (1) is:

$$\widehat{\beta}(\phi; \underline{x}) = \max \left(0, \frac{\|\tilde{A}\underline{x} - \tilde{b}\|_2 - \phi}{\|\underline{x}\|_2 + \nu} \right) \quad (11)$$

Proof: The internal minimization in Eq. (10) defines the smallest residual norm

$\widehat{s}(\alpha; \underline{x}) \equiv \max_{\tilde{A}, \tilde{b} \in U(\alpha; \underline{A}, \underline{b})} \|\tilde{A}\underline{x} - \tilde{b}\|_2$ that may occur with regression $\underline{x}$ when the level of uncertainty is α . Lemma B (Appendix B) shows that the smallest residual error is given by:

$$\widehat{s}(\alpha, \underline{x}) = \max \left(0, \left\| \tilde{A}\underline{x} - \tilde{b} \right\|_2 - \alpha(\|\underline{x}\| - \nu) \right) \quad (12)$$

Substituting (12) in (10) completes the proof.

Eq. (11) indicates that the opportuneness decreases linearly as the windfall performance increases until it reaches zero for the nominal residual norm. Examples of opportunity curves for the three Tikhonov regressions analyzed in Figure 1 are shown in Figure 2, along with the corresponding robustness curves. Each pair of robustness and opportunity curves, for the same Tikhonov regression, emanates from the x -axis at nominal performance $\rho = \phi = r_{nom}(\underline{x}_\mu)$.

Definition 4 – Consistent windfall performance of a RS regression—A windfall performance that is consistent with the RS regression $\underline{x}_{RS}(\rho)$, is a windfall performance ϕ_{RS} whose opportuneness $\beta(\phi_{RS}, \underline{x}_{RS}(\rho))$ is the same level of uncertainty as the robustness $\hat{\alpha}(\rho, \underline{x}_{RS}(\rho))$.

The consistent windfall performance of a RS regression assesses how best it may fit an ideal observation within the level of uncertainty that equals the corresponding robustness. It therefore provides a criterion for choosing the model-consistent robust-satisficing (MCRS) regression, as defined next:

Definition 5 – Model consistent robust-satisficing (MCRS) regression—The model-consistent robust satisficing (MCRS) regression is the RS regression that minimizes the consistent windfall performance: $\underline{x}_{MCRS} \equiv \underline{x}_{RS}(\rho_{MCRS})$ where

$$\rho_{MCRS} = \begin{cases} \arg \min_{\rho} \phi_{RS}(\rho, \underline{x}_{RS}(\rho)) & \text{if } \forall \rho \quad \phi_{RS}(\rho, \underline{x}_{RS}(\rho)) > 0 \\ \min(\rho | \phi_{RS}(\rho, \underline{x}_{RS}(\rho)) = 0) & \text{if } \exists \rho \quad \phi_{RS}(\rho, \underline{x}_{RS}(\rho)) = 0 \end{cases} \quad (13)$$

That is, an MCRS regression is a robust-satisficing regression that under the level of uncertainty that corresponds to its robust-satisficing robustness provides the best performance. Note that determining $\underline{x}_{MCRS}$ does not require any assumption about the level of uncertainty nor the critical performance, and is therefore parameter free.

For the info-gap model of Eq. (1), the RS regression is a Tikhonov regression $\underline{x}_\mu$. The following Propositions develop: (i) the corresponding consistent windfall performance (Proposition 6), and (ii) the MCRS regression (Proposition 7).

Proposition 6 – Consistent windfall performance of a Tikhonov regression—Given the info-gap model of Eq. (1), the consistent windfall performance of the RS regression $\underline{x}_{RS}(\rho)$ is:

$$\phi_{RS}(\mu) = \max(0, \frac{\left\| \tilde{A} \tilde{x}_\mu - \tilde{b} \right\|_2^2 - \mu \left\| \tilde{x}_\mu \right\|_2 (\left\| \tilde{x}_\mu \right\|_2 + \nu)}{\left\| \tilde{A} \tilde{x}_\mu - \tilde{b} \right\|_2}) \quad (14)$$

Where $\tilde{x}_\mu$ is a Tikhonov regression with parameter μ satisfying Eq. (6).

Proof: According to Proposition 4, the robustness of a Tikhonov regression $\hat{\alpha}_{RS}(\mu) \equiv \hat{\alpha}_{RS}(\rho_{RS}(\mu); \tilde{x}_\mu)$ is given by Eq. (9). Assuming that $\hat{\alpha}_{RS}(\mu) < \beta(\phi=0, \tilde{x}_\mu)$, the condition $\beta(\phi_{RS}; \tilde{x}_\mu) =$

$\hat{\alpha}_{RS}(\mu)$, results in: $\frac{\mu \left\| \tilde{x}_\mu \right\|_2}{\left\| \tilde{A} \tilde{x}_\mu - \tilde{b} \right\|_2} = \frac{\left\| \tilde{A} \tilde{x}_\mu - \tilde{b} \right\|_2 - \phi_{RS}}{\left\| \tilde{x}_\mu \right\|_2 + \nu}$, which can be inverted to express ϕ_{RS} as the second term in Eq. (14). If $\hat{\alpha}_{RS}(\mu) \geq \beta(\phi=0, \tilde{x}_\mu)$, the robustness is higher than the opportuneness for zero windfall performance so $\phi_{RS}(\mu) = 0$.

The consistent windfall performance is determined graphically in Figure 2 for each of the Tikhonov regressions analyzed there. The Figure depicts the robustness and opportuneness curves for three Tikhonov regressions. The circle on each robustness curve marks the critical performance for which that regression is robust-satisficing and the resulting robustness. Assuming that the level of uncertainty equals the robustness, the corresponding opportuneness curve is crossed at $\beta(\phi_{RS}; \tilde{x}_\mu) = \hat{\alpha}_{RS}(\mu)$, as marked by the triangle on the matching opportuneness curve, thereby determining the consistent windfall performance for that RS regression. The three pairs of robustness and opportunity curves in Figure 2 suggests that the consistent windfall performance reaches a minimum for some μ , as proved in the next proposition.

Proposition 7 - Model Consistent Tikhonov Regression—Given the info-gap model of Eq. (1), the MCRS regression is a Tikhonov regression with $\mu = \mu_c$ given by:

$$\mu_c = \begin{cases} \arg \min_{\mu} (\phi_{RS}(\mu)) & \text{if } \forall \mu \phi_{RS}(\tilde{x}_\mu) > 0 \\ \min(\mu | \phi_{RS}(\mu) = 0) & \text{if } \exists \mu \phi_{RS}(\tilde{x}_\mu) = 0 \end{cases} \quad (15)$$

Proof: According to Proposition 2, the RS regressions under uncertainties described by the info-gap model of Eq. (1) have the form of Tikhonov regressions. Lemma C in Appendix C states the consistent windfall performance either becomes zero at some μ , or reaches a single minimum. Hence, the MCRS is the Tikhonov regression with the parameter μ_c given by Eq. (15).

Note that when $\phi_{RS}(\mu_0) = 0$ for some μ_0 , the unique MCRS regression defined with $\mu_c = \min(\mu | \phi_{RS}(\mu) = 0)$ is satisfied exactly by a possible ideal observation within the specified level of uncertainty $\hat{\alpha}_{RS}(\mu_c)$.

V. Application to brain machine interface

Brain-machine interfaces (BMIs) are based on extracting movement-related signals from the neural activity of a large number of cortical neurons with the goal of restoring motor functions in severely paralyzed patients [12]–[15]. The data used in this paper was recorded during the BMI experiments reported in [16][17]. The experiments were conducted with monkeys that controlled the position of a cursor on a computer screen using either a hand-held pole (*pole control*) or a BMI (*brain control*). The last ten minutes of pole control were used to determine

the LS regression between the measured velocity of hand movements and the neural activity. The resulting regression was subsequently used in brain control to generate real time predictions of the velocity from the recorded neural activity and direct the cursor accordingly.

The neural activity was represented by the spike counts in 100msec bins, and the regression included ten lags of binned spike counts from each of the recorded neurons [16]. The typical session analyzed here included 183 neurons so the regression was based on 1831 inputs. Non-linear methods, including Kalman filter, and multi-layer feedforward artificial neural networks were investigated off line and it was concluded that they could not consistently outperform the linear filter [16][18].

The L-curve shown in Figure 3 depicts the solution norm versus the residual norm for different Tikhonov regressions, based on the last ten minutes of pole control (training data). The L-curve is usually characterized by an “L” shape and the Tikhonov parameter is selected at its corner, by locating, for example, the point of maximum curvature [4][8]. However, the L-curve in Figure 3 does not depict the characteristic L-shape, reflecting the low signal-to-noise ratio in the neural activity. Thus, the L-curve method is inadequate for choosing the proper Tikhonov parameter in BMI applications.

As expected from proposition 7, the consistent windfall performance reaches a unique minimum as a function of the Tikhonov parameter, as demonstrated in Figure 4 (for the same 10 minutes of training data used in Figure 3). Based on a signal-to-noise ratio analysis, the neural activity is expected to be 2–3 orders of magnitude more noisy than the velocity measurements, so the relative weight ν was set to $\nu = 0.001$. The resulting minimum is reached at $\mu = 3600$. Sensitivity analysis indicates that increasing or decreasing the relative weight ν by an order of magnitude has only a small effect on the chosen parameter which remains in the range [3300,3900].

Figure 5 compares the performance of different Tikhonov regressions for ten minutes of testing data during pole control and brain control. Whereas the LS algorithm minimizes the residual error for the training data (last ten minutes of pole control), alternative Tikhonov regressions outperform on testing data from pole control (collected during the preceding ten minutes of pole control) and brain control (collected during the first ten minutes of brain control). Consider, in particular, the performance on testing data from brain control depicted in the lower panel of Figure 5. The LS regression results in large residual norm above 180, which is outside the scale. All the Tikhonov regressions with the Tikhonov parameter in the depicted range outperform the LS regression. However, while the Tikhonov regression selected using the L-curve reduces the residual norm to 78.5, the MCRC regression reduces the residual norm to 67.5, close to the minimum level that can be achieved by any Tikhonov regression.

The MCRC provides significant improvement in performance for both pole and, most importantly, brain control. For testing data from pole control, additional improvement in performance could be achieved with a Tikhonov parameter that is smaller than the one chosen by the model-consistency criterion. Thus, the level of uncertainty implied by the model consistency criterion is higher than the level of perturbation in adjacent ten minutes records of the neural activity. However, when testing the performance on brain control, which is the critical application of the linear regression, it is evident that the method indeed captures the adequate level of uncertainty in the data, and provides close to optimal performance.

VI. Discussion

In this paper we considered the inverse problem of estimating the linear regression $\underline{x}$ from available observations of the dependent and independent variables. We developed an info-gap robust satisficing approach to regression, which maximizes the robustness for obtaining a

critical sub-optimal performance. For the particular info-gap considered here, the resulting regression was shown to have the form of a Tikhonov regularized solution. This is the same regression that solves the dual min-max problem of maximizing the performance for a given level of uncertainty [6]. We note that different info-gap models would result in different RS regressions (and min-max regressions) that are not necessarily the same as Tikhonov regressions, as demonstrated in [11].

The notion of robustness was applied in [19] to estimate a deterministic regression vector $\underline{x}$ from observations $\underline{y} = A\underline{x} + \underline{w}$ where A is known and $\underline{w}$ is additive noise. The robust regression for a given level of performance was shown to be the min-max regression for the corresponding level of uncertainty. It was concluded there that the resulting robustness/performance trade off establishes an important design tool for choosing the regression based on both criteria. In our case the matrix A is not known and only uncertain observations of its values are available. Furthermore, here we use also the notion of opportuneness to resolve the resulting trade-off and select a unique model-consistent robust-satisficing regression.

The info-gap approach emphasizes the trade-off between performance and robustness: increasing robustness is possible only by relinquishing performance. The proper trade-off is usually determined by presuming the knowledge of an associated design parameter: The min-max approach presumes a level of uncertainty; the robust-satisficing approach developed here presumes a level of critical performance; Tikhonov regression relies on the weight of the regression norm relative to the residual norm in the performance criterion. Thus, all these methods rely on proper selection of a design parameter: the level of uncertainty, the critical performance or the relative weight of the regression norm.

Here we introduced a new criterion, based on the consistency between the observations and the linear model, to resolve the robustness/performance trade-off and determine a unique parameter-free regression. The consistency of a robust satisfying (RS) regression is assessed by the consistent windfall performance – the minimum residual error that the regression can achieve with an ideal observation that is consistent with the corresponding level of robustness. The consistent windfall performance has a unique minimum as a function of the Tikhonov parameter and thus can be used to determine a unique regression, the model consistent robust satisfying (MCRS) regression, and assess the underlying level of uncertainty in the observations.

We demonstrated the model-consistent algorithm for choosing the Tikhonov parameter for the challenging application of neural decoding for brain-machine interfaces (BMIs). The MCRS provides significant improvement in performance for both pole and, most importantly, brain control.

Acknowledgments

The authors are pleased to acknowledge valuable comments by Yakov Ben-Haim and Yonina Eldar. This research was supported by the Abramsom Center for the Future of Health and by the fund for the promotion of research at the Technion and by grants from DARPA and NIH to MALN.

References

1. Tikonov, AN.; Arsenin, VY. Solution to Ill-Posed Problems. Washington, DC: V.H. Winston; 1977.
2. Groetsch, CW. Inverse Problems in the Mathematical Sciences. Vieweg; Wiesbaden: 1993.
3. Golub, GH.; Van-Loan, CF. Matrix Computations. The Johns Hopkins University Press; 1996.
4. Hansen, PC. Rank-Deficient and Discrete Ill-Posed Problems. SIAM; 1997.
5. Fierro RD, Golub GH, Hansen PC, O'Leary DP. Regularization by truncated total least square. SIAM J Sci Comput 1997;18(4):1223–1241.

6. Ghaoui LE, Le Bret H. Robust solutions to least square problems with uncertain data. *SIAM J Matrix Anal Appl* October;1997 18(4):1035–1064.
7. Chandrasekaran S, Golub GH, Gu M, Sayed AH. Parameter Estimation in the presence of bounded data uncertainties. *SIMAX* 1998;19(1):235–252.
8. Hansen PC. Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review* 1992;34(4):561–580.
9. Simon, HA. Models of bounded rationality. MIT Press; Cambridge, MA: 1982.
10. Ben-Haim, Y. Info-gap Decision Theory: Decisions under severe uncertainty. 2. Academic Press; London: 2006.
11. Zacksenhouse M, Yaffe A, Nemets S, Ben-Haim Y, Lebedev MA, Nicolelis MAL. An Info-gap approach to linear regression. *Int Conf on Acoustic, Speech and Signal Processing, Proc ICASSP2006*;3:800–803.
12. Wessberg J, Stambaugh CR, Kralik JD, Beck PD, Laubach M, Chapin JK, Kim J, Biggs SJ, Srinivasan MA, Nicolelis MAL. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature* 2000;408:361–365. [PubMed: 11099043]
13. Nicolelis MAL. Actions from thoughts. *Nature* 2001;409(18):403–407. [PubMed: 11201755]
14. Taylor DM, Tillery SI, Schwartz AB. Direct cortical control of 3D neuroprosthetic devices. *Science* 2002;296:1829–1832. [PubMed: 12052948]
15. Nicolelis MAL. Brain-machine interfaces to restore motor functions and probe neural circuits. *Nature Rev* 2003;4:417–422.
16. Carmena JM, Lebedev MA, Crist RE, O'Doherty JE, Santucci DM, Dimitrov D, Patil PG, Henriquez CS, Nicolelis MAL. Learning to control a brain-machine interface for reaching and grasping by primates. *PLoS Biol* 2003;1:193–208.
17. Lebedev MA, Carmena JM, O'Doherty JE, Zacksenhouse M, Henriquez CS, Principe JC, Nicolelis MAL. Cortical ensemble adaptation to represent velocity of an artificial actuator controlled by a brain machine interface. *J Neurosci* 2005;25(19):4681–4693. [PubMed: 15888644]
18. Zacksenhouse, M. Strategies for neural ensemble data analysis, to appear. In: Nicolelis, MAL., editor. *Methods for Neural Ensemble Recordings*. 2. CRC Press; 2006.
19. Ben-Haim Z, Eldar YC. Maximum Set Estimators with Bounded Estimation Error. *IEEE Trans Signal Processing* 2005;53(8):3172–3182.

Appendix A: Maximum robustness

Lemma A1

For non singular observations, i.e., $0 < \|\tilde{A}\underline{x} - \tilde{b}\|$ for all $\underline{x}$, the robustness

$$\hat{\alpha}(\rho; \underline{x}) = \max \left(0, \frac{\rho - \|\tilde{A}\underline{x} - \tilde{b}\|}{\|\underline{x}\| + \nu} \right) \text{ has a unique global maximum for desired performance } \rho > \|\tilde{A}\underline{x}_{LS} - \tilde{b}\| \text{ where } \underline{x}_{LS} \text{ is the least-squares regression.}$$

Proof

The gradient of the robustness can be expressed as:

$$\nabla_{\underline{x}} \hat{\alpha}(\rho; \underline{x}) = - \frac{1}{\|\underline{x}\| + \nu} \left(\frac{\tilde{A}^T (\tilde{A}\underline{x} - \tilde{b})}{\|\tilde{A}\underline{x} - \tilde{b}\|} + \hat{\alpha}(\rho; \underline{x}) \frac{\underline{x}}{\|\underline{x}\|} \right) \quad (\text{A.1})$$

At a point where the gradient vanishes, the Hessian of the robustness is given by:

$$\begin{aligned} \Delta_{\underline{x}} \widehat{\alpha}(\rho; \underline{x}) |_{\nabla_{\underline{x}} \widehat{\alpha}(\rho; \underline{x})=0} = & - \frac{1}{\|\underline{x}\| + \nu} \left\{ \left(\frac{\tilde{A}^T \tilde{A}}{\|\tilde{A}\underline{x} - \tilde{b}\|} - \frac{\tilde{A}^T (\tilde{A}\underline{x} - \tilde{b})(\tilde{A}\underline{x} - \tilde{b})^T \tilde{A}}{\|\tilde{A}\underline{x} - \tilde{b}\|^3} \right) + \right. \\ & \left. + \widehat{\alpha}(\rho; \underline{x}) \left(\frac{I}{\|\underline{x}\|} - \frac{\underline{x}\underline{x}^T}{\|\underline{x}\|^3} \right) + \nabla_{\underline{x}} \widehat{\alpha}(\rho; \underline{x}) \frac{\underline{x}}{\|\underline{x}\|} \right\} \end{aligned} \quad (\text{A.2})$$

Furthermore, at the point where the gradient of the robustness vanishes, the last term in (A.2) is zero. We will show that the sum of the remaining two matrices in the external parenthesis is positive definite, and thus that after multiplication by minus, the resulting Hessian is negative definite.

The first matrix can be expressed as:

$$H_1 = \frac{1}{\|\underline{x}\| + \nu} \tilde{A}^T \left(\frac{I}{\|\tilde{A}\underline{x} - \tilde{b}\|} - \frac{(\tilde{A}\underline{x} - \tilde{b})(\tilde{A}\underline{x} - \tilde{b})^T}{\|\tilde{A}\underline{x} - \tilde{b}\|^3} \right) \tilde{A}$$

The inner matrix in H_1 has the form of Lemma-A2 (see below) and hence all its eigen-values are positive except for a single eigen-value that is zero. The eigen-vector that corresponds to the zero eigen-value is $\tilde{A}\underline{x} - \tilde{b}$. Considering the non-singular problem, $\tilde{A}\underline{y}$ is never parallel to $\tilde{A}\underline{x} - \tilde{b}$ for any $\underline{x}$ and $\underline{y}$ (since otherwise there exist χ such that $\tilde{A}\chi\underline{y} = \tilde{A}\underline{x} - \tilde{b}$ and hence $\tilde{A}(\underline{x} - \chi\underline{y}) = \tilde{b}$). Since no vector in $\underline{y} \in \mathbb{R}^n$ is mapped into the single eigen-vector with zero eigen-value, H_1 is positive definite. The second term in the Hessian has of the form of the Lemma-A2 and hence is positive semi-definite. The sum of a positive definite matrix and a positive semi-definite matrix is a positive definite matrix. Finally, the multiplication by minus results in a negative definite Hessian matrix.

A negative definite Hessian implies that any point, at which the gradient of the robustness vanishes, is a local maximum. However, since there are no local minima there can be only one local maximum, which is therefore the global unique maximum.

Lemma A2

For any vector $\underline{y} \in \mathbb{R}^n$ the matrix $\left(\frac{I}{\|\underline{y}\|} - \frac{\underline{y}\underline{y}^T}{\|\underline{y}\|^3} \right)$ is positive semi-definite, with a single zero eigen-value whose corresponding eigen-vector is $\underline{y}$.

Proof of Lemma-A2

The first matrix is full rank with n eigen-values $\|\underline{y}\|^{-1}$. The second matrix has a single eigen-value equal of $\|\underline{y}\|^{-1}$ with eigen-vector $\underline{y}$. Hence their difference is a matrix with rank $n-1$, with $n-1$ positive eigen-values of $\|\underline{y}\|^{-1}$ and one zero eigen-value, forming a semi-positive matrix. The eigen-vector with a zero eigen-value is $\underline{y}$.

Appendix B: Smallest residual norm

Lemma B

Within the Info-gap $U(\alpha)$ of Eq. (1), the smallest residual norm that can be achieved by the regression $\underline{x}$ is given by (see also [7] for the largest residual norm):

$$\widehat{s}(\alpha, \underline{x}) = \max(0, \|\tilde{A}\underline{x} - \tilde{b}\|_2 - \alpha \|\underline{x}\| - \nu\alpha) \quad (16)$$

Proof

Triangle inequality implies that given $A, b \in U(\alpha)$:

$$\|A\underline{x} - \underline{b}\| \geq \|\tilde{A}\underline{x} - \tilde{b}\| - \|(A - \tilde{A})\underline{x} - (b - \tilde{b})\| \geq \|\tilde{A}\underline{x} - \tilde{b}\| - \alpha \|\underline{x}\| - \nu\alpha \quad (17)$$

and equality is achieved for: $A_{best} = \tilde{A} - \frac{(\tilde{A}\underline{x} - \tilde{b})\underline{x}^T}{\|\tilde{A}\underline{x} - \tilde{b}\| \|\underline{x}\|} \alpha$ and $b_{best} = \tilde{b} + \frac{(\tilde{A}\underline{x} - \tilde{b})}{\|\tilde{A}\underline{x} - \tilde{b}\|} \nu\alpha$. Since $A_{best}, b_{best} \in U(\alpha)$, the minimum of $\|A\underline{x} - \underline{b}\|$ with respect to all $A, b \in U(\alpha)$ is $\|\tilde{A}\underline{x} - \tilde{b}\| - \alpha \|\underline{x}\| - \nu\alpha$, or zero if this is negative.

Appendix C: Consistent windfall performance

Lemma C

For the Info-gap of Eq. (1), the consistent windfall performance either becomes zero or reaches a single minimum as a function of the Tikhonov parameter μ , i.e., when $\phi_{RS}(\mu) > 0$ for all μ , then $\mu_c = \arg \min_{\mu} (\phi_{RS}(\mu))$ is uniquely defined.

Proof

The consistent windfall performance is given by:

$$\phi_{RS}(\mu) = \max(0, \frac{\|\tilde{A}\underline{x}_{\mu} - \tilde{b}\|^2 - \mu \|\underline{x}_{\mu}\|(\|\underline{x}_{\mu}\| + \nu)}{\|\tilde{A}\underline{x}_{\mu} - \tilde{b}\|}). \quad \text{Consider the case when } \phi_{RS}(\mu) > 0 \text{ for all } \mu, \text{ then the derivative of } \phi_{RS}(\mu) \text{ with respect to } \mu \text{ is given by:}$$

$$\frac{d\phi_{RS}(\mu)}{d\mu} = \left(\frac{d\underline{x}_{\mu}}{d\mu} \right)^T \nabla_{\underline{x}} \phi_{RS}(\mu) + \frac{\partial \phi_{RS}(\mu)}{\partial \mu} \quad (B.1)$$

Noting that for a Tikhonov regression $\underline{x}_{\mu} = (A^T A + \mu I)^{-1} A^T \underline{b}$, implies $A^T (A\underline{x}_{\mu} - \underline{b}) = -\mu \underline{x}_{\mu}$, the gradient of $\phi_{RS}(\mu)$ with respect to $\underline{x}_{\mu}$ is:

$$\begin{aligned}\nabla_{\underline{x}} \phi_{RS} &= \frac{(-4\mu \underline{x}_\mu - \mu \nu \underline{x}_\mu / \|\underline{x}_\mu\|) \|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2 + \mu \underline{x}_\mu \left(\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2 - \mu \|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \right)}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^3} \\ &= \frac{-\left(\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2 (3 + \nu / \|\underline{x}_\mu\|) + \mu \|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \right) \mu \underline{x}_\mu}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^3}\end{aligned}\quad (\text{B.2})$$

The derivative of $\underline{x}_\mu$ with respect to μ is given by:

$$\left(\frac{d\underline{x}_\mu}{d\mu} \right) = \frac{d(A^T A + \mu I)^{-1} A^T \underline{b}}{d\mu} = -(A^T A + \mu I)^{-2} A^T \underline{b} = -(A^T A + \mu I)^{-1} \underline{x}_\mu \quad (\text{B.3})$$

Inserting (B.3) and (B.2) in (B.1) yields:

$$\begin{aligned}\frac{d\phi_{RS}(\mu)}{d\mu} &= \left(\frac{d\underline{x}_\mu}{d\mu} \right)^T \nabla_{\underline{x}} \phi_{RS}(\mu) + \frac{\partial \phi_{RS}(\mu)}{\partial \mu} = \\ &= \frac{\mu \left(\underline{x}_\mu^T (A^T A + \mu I)^{-1} \underline{x}_\mu \right) \left(\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2 (3 + \nu / \|\underline{x}_\mu\|) + \mu \|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \right)}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^3} \\ &\quad - \frac{\|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^3}\end{aligned}\quad (\text{B.4})$$

Since $\left. \frac{d\phi_{RS}(\mu)}{d\mu} \right|_{\mu=0} = -\frac{\|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu)}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2} < 0$ while $\frac{d\phi_{RS}(\mu)}{d\mu} \xrightarrow{\mu \rightarrow \infty} \frac{\mu^{-2} \|\tilde{A}^T \tilde{b}\|^3}{\|\tilde{b}\|^3} \nu \rightarrow 0^+ > 0$, it follows that $\phi_{RS}(\mu)$ reaches a local minimum in at least one point.

At the point where the first derivative vanishes, the second derivative of $\phi_{RS}(\mu)$ can be evaluated as:

$$\begin{aligned}\left. \frac{d^2 \phi_{RS}(\mu)}{d\mu^2} \right|_{\frac{d\phi_{RS}}{d\mu}=0} &= \frac{1}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^3} \frac{d}{d\mu} (g(\mu)) \\ &= \frac{1}{\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^3} \left(\left(\frac{d\underline{x}_\mu}{d\mu} \right)^T \nabla_{\underline{x}} g(\mu) + \frac{\partial}{\partial \mu} (g(\mu)) \right)\end{aligned}\quad (\text{B.5})$$

where $g(\mu)$ is the nominator of $\frac{d\phi_{RS}(\mu)}{d\mu}$ in (B.4), i.e.,

$$\begin{aligned}g(\mu) &= \mu \left(\underline{x}_\mu^T (\tilde{A}^T \tilde{A} + \mu I)^{-1} \underline{x}_\mu \right) \left(\|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2 (3 + \nu / \|\underline{x}_\mu\|) + \mu \|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \right) \\ &\quad - \|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \|\tilde{A} \underline{x}_\mu - \tilde{b}\|_2^2\end{aligned}\quad (\text{B.6})$$

The second derivative can be expressed as:

$$\begin{aligned}
& \left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\|^3 \left(\frac{d^2 \phi_{RS}(\mu)}{d\mu^2} \Big|_{\frac{d\phi_{RS}}{d\mu}=0} \right) = \\
& \left(\mu^2(4+\nu/\|\underline{x}_\mu\|) + \left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\|^2 \mu \nu / \|\underline{x}_\mu\|^3 \right) \left(\underline{x}_\mu^T (\tilde{A}^T \tilde{A} + \mu I)^{-1} \underline{x}_\mu \right)^2 \\
& + \mu \left(\left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\|^2 (3+\nu/\|\underline{x}_\mu\|) + \mu \|\underline{x}_\mu\| (\|\underline{x}_\mu\| + \nu) \right) \left(\underline{x}_\mu^T (\tilde{A}^T \tilde{A} + \mu I)^{-2} \underline{x}_\mu \right) \\
& + \left(\left\| \tilde{A} \underline{x}_\mu - \tilde{b} \right\|^2 (5+2\nu/\|\underline{x}_\mu\|) \right) \left(\underline{x}_\mu^T (\tilde{A}^T \tilde{A} + \mu I)^{-1} \underline{x}_\mu \right) > 0
\end{aligned} \tag{B.7}$$

The positive second derivative of (B.7) indicates that the extreme points of $\phi_{RS}(\mu)$ are local minima. Since no extreme point is a local maximum there can be only one local minimum and

hence $\mu_c = \arg \min_{\mu} \min_{A, \underline{b} \in U(\alpha(\mu); \tilde{A}, \tilde{b})} \left\| A \underline{x}_\mu - \underline{b} \right\|$ is unique.

Figure 1.

Robustness curves for three Tikhonov regressions ($\nu = 0.001$). The envelope of the robustness curves defines the maximum robustness that can be achieved at any level of the critical performance and the corresponding robust-satisficing regression.

Figure 2. Robustness and opportuneness curves for three Tikhonov regressions ($\nu = 0.001$). The grey circle on each robustness curve marks the critical performance and robustness at which the regression is robust satisfying. At the same level of uncertainty, the opportuneness curve determines the consistent windfall performance that could be achieved with that regression, as marked by the corresponding grey triangles.

Figure 3.

L-curve depicting the solution norm versus the residual norm of Tikhonov regressions with increasing Tikhonov parameter for the training data.

Figure 4. Consistent windfall performance reaches a unique minimum as a function of the Tikhonov parameter ($\nu = 0.001$).

Figure 5. Residual norm for testing data from pole control (top panel) and brain control (bottom panel) for Tikhonov regressions with increasing parameter μ .