Abstract

In this paper, we consider the problem of identifying a linear map from measurements which are subject to intermittent and potentially unbounded gross errors. This is a fundamental problem in many estimation-related applications such as fault detection, state estimation in lossy networks, hybrid system identification, robust estimation, etc. The problem is hard because it exhibits some intrinsic combinatorial features. Therefore, obtaining an effective solution necessitates relaxations that are both solvable at a reasonable cost and effective in the sense that they can return the true parameter vector. The current paper discusses a nonsmooth convex optimization approach. In particular, it is shown that under appropriate conditions on the data, an exact estimate can be recovered from data corrupted by a large (even infinite) number of gross errors.

Keywords. robust estimation, outliers, system identification, nonsmooth optimization.

1 Introduction

1.1 Problem and motivations

We consider a linear measurement model of the form

\[ y_t = x_t^\top \theta^o + f_t + e_t \]  

(1)

where \( y_t \in \mathbb{R} \) is the measured signal, \( x_t \in \mathbb{R}^n \) the regression vector, \( \{e_t\} \) a sequence of zero-mean and bounded errors (e.g., measurement noise, model mismatch, uncertainties, etc.) and \( \{f_t\} \) a sequence of intermittent and possibly unbounded errors. Assume that we observe the sequences \( \{x_t\}_{t=1}^N \) and \( \{y_t\}_{t=1}^N \) and would like to compute the parameter vector \( \theta^o \) from these observations. We are interested in doing so without knowing any of the sequences \( \{f_t\} \) and \( \{e_t\} \). We do however make the following assumptions:

L. Bako is with Laboratoire Ampère – Ecole Centrale de Lyon – Université de Lyon, France. E-mail: laurent.bako@ec-lyon.fr, Tel.: +33 472 186 452.

H. Ohlsson is with (a) – Department of Electrical Engineering, Linköping University, SE-581 83 Linköping, Sweden and (b) – Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, CA, USA. E-mail: ohlsson@berkeley.edu
• \{e_t\} is a bounded sequence.
• \{f_t\} is a, possibly, unbounded sequence with intermittent zeros.

This is an important estimation problem arising in many situations such as fault detection, hybrid system identification, subspace clustering, error correction in communication networks. The case when \(\{f_t\}\) is zero and \(\{e_t\}\) is a Gaussian process has been well-studied in linear system identification theory (see, e.g., the text book [17]). A less studied, but very relevant scenario is when the additional perturbation \(\{f_t\}\) in (1) is nonzero, possibly unbounded and contains intermittent zeros. It is worth noticing the difference to the problems studied in the field of compressive sensing [5, 11, 8]. In compressive sensing, the sought parameter vector is assumed sparse and the measurement noise \(\{e_t\}\), often Gaussian or bounded. Here, no assumptions are made concerning sparsity of \(\theta^o\). We will, in this contribution, study essentially the case when the data is noise-free (i.e., \(e_t = 0\) for all \(t\)) and \(\{f_t\}\) is a time sequence with intermittent zeros. We will derive conditions for perfect recovery and provide effective algorithms for computing \(\theta^o\).

In the second part of the paper, the model assumption is relaxed to allow both \(e_t\) and \(f_t\) to be simultaneously nonzero. Note that this might be a more realistic scenario since most applications have measurement noise.

For illustrative purposes, let us discuss briefly some applications where a model of the form (1) is of interest.

**Switched linear system identification.** A discrete-time Multi-Input Single-Output (MISO) Switched Linear System (SLS) can be written on the form

\[
y_t = x_t^\top \theta_{\sigma_t}^o + e_t, \tag{2}
\]

where \(x_t \in \mathbb{R}^n\) is the regressor at time \(t \in \mathbb{Z}_+\) defined by

\[
x_t = [y_{t-1} \cdots y_{t-n_a} u_{t-1}^\top \cdots u_{t-n_b}^\top]^\top. \tag{3}
\]

where \(u_t \in \mathbb{R}^{n_u}\) and \(y_t \in \mathbb{R}\) denote the input and the output of the system. The integers \(n_a\) and \(n_b\) in (3) are the maximum output and input lags (also called the orders of the system). \(\sigma_t \in \{1, \ldots, s\}\) is the discrete mode (or discrete state) indexing the active subsystem at time \(t\). \(\sigma_t\) is in general assumed unobserved. \(\theta_{\sigma_t}^o \in \mathbb{R}^n\), \(n = n_a + n_b n_u\), is the parameter vector (PV) associated with the mode \(\sigma_t\). For \(\theta^o \in \{\theta_1^o, \ldots, \theta_s^o\}\), the Switched Auto-Regressive eXogenous (SARX) model (2) can be written in the form (1), with unknown \(f_t\) of the following structure

\[f_t = x_t^\top (\theta_{\sigma_t}^o - \theta^o).\]

For a background on hybrid system identification, we refer to the references [23, 13, 26, 11, 15, 18, 22, 20].
Identification from faulty data. A model of the form (1) also arises when one has to identify a linear dynamic system which is subject to intermittent sensor faults. This is the case in general when the data are transmitted over a communication network [5]. Model (1) is suitable for such situations and the sequence \{f_t\} then models occasional data packets losses or potential outliers.

State estimation from measurements corrupted by intermittent errors. Considering a MISO dynamic system with state dynamics described by
\[
\begin{align*}
\hat{y}_t &= C^T z_t + f_t, \\
&= (A, B, C) \text{ being known matrices of appropriate dimensions, and } \{f_t\}
\end{align*}
\]
a sparse sequence of possibly unbounded errors, the finite horizon state estimation problem reduces to the estimation of the initial state \(z_0 = \theta\). We get a model of the form (1) by setting
\[
\begin{align*}
y_t &= \hat{y}_t - C^T \Delta_t \bar{u}_t \\
x_t &= (A^t)^T C, \text{ with } \Delta_t = [A^{t-1} B \cdots AB B], \\
&\bar{u}_t = [u_0^T \cdots u_{t-1}^T]^T.
\end{align*}
\]
In this latter case, it can however be noted that the dataset \{x_t\} may not be generic enough.

Contributions. One promising method for estimating model (1) is by nonsmooth convex optimization as suggested in [5, 1, 3]. More precisely, inspired by the recent theory of compressed sensing [5, 11, 8], the idea is to minimize a nonsmooth (and non differentiable) sum-of-norms objective function involving the fitting errors. Under noise-free assumption, such a cost function has the nice property that it is able to provide the true parameter vector in the presence of arbitrarily large errors \{f_t\} provided that the number of nonzero errors is small in some sense. Of course, when the data are corrupted simultaneously by the zero mean noise \{e_t\} and the gross errors \{f_t\}, the recovery cannot be exact any more. It is however expected (as simulations tend to suggest) that the estimate will still be closed to the true one.

The current paper intends to present a new analysis of the nonsmooth optimization approach and provide some elements for further understanding its behavior. For this purpose, we start by showing in Section 2.1 that the nonsmooth optimization can be viewed as the convex relaxation of a (ideal) combinatorial \(\ell_0\)-"norm" formulation. We then derive in Section 2.2 necessary and sufficient conditions for optimality. Based on those conditions we later (Section 2.3) establish new sufficient conditions for exact recovery of the true parameter vector in (1). The noisy case is treated in Section 3.2. Finally, numerical experiments are described in Section 5 and concluding remarks are given in Section 6.

1.2 Notations

Let \(I = \{1, \ldots, N\}\) be the index set of the measurements. For any \(\theta \in \mathbb{R}^n\), define a partition of the set of indices \(I\) by
\[
I^-(\theta) = \{t \in I : \theta^T x_t - y_t < 0\}, \quad I^+(\theta) = \{t \in I : \theta^T x_t - y_t > 0\},
\]
1 In this paper, the term genericity for a dataset characterizes a notion of linear independence. For example, a set of \(N > n\) data points in general linear position in \(\mathbb{R}^n\) is more generic than a set of data points contained in one subspace. We will introduce different quantitative measures of data genericity in the sequel (see Definition 1 and Theorem 4).
\[ \mathbb{I}^0(\theta) = \{ t \in I : \theta^\top x_t - y_t = 0 \} \]. Let \( X = [x_1 \ x_2 \ \cdots \ x_N] \in \mathbb{R}^{n \times N} \) be the matrix formed with the available regressors \( \{ x_t \}_{t=1}^N \). If \( I \subset \mathbb{I} \), the notation \( X_I \) denotes a matrix formed with the columns of \( X \) indexed by \( I \). Likewise, with \( y = [y_1 \ y_2 \ \cdots \ y_N]^\top \in \mathbb{R}^N \), \( y_I \) consists of the vector formed with the entries if \( y \) indexed by \( I \). We will use the convention that \( X_I = 0 \) (resp. \( y_I = 0 \)) when the index set \( I \) is empty.

**Cardinality of a finite set.** Throughout the paper, whenever \( S \) is a finite set, the notation \( |S| \) will refer to the cardinality of \( S \). However, for a real number \( x \), \( |x| \) will denote the absolute value of \( x \).

**Vector norms.** For any vector \( z = [z_1 \ \cdots \ z_N]^\top \in \mathbb{R}^N \), by \( \|z\|_p = (|z_1|^p + \cdots + |z_N|^p)^{1/p} \). Note that \( \|x\|_\infty = \max_{i=1,...,N} |x_i| \). The \( \ell_0 \) "norm" of \( z \) is defined to be the number of nonzero entries in \( z \), i.e., \( \|z\|_0 = |\{ i : z_i \neq 0 \}| \).

**Matrix norms.** The following matrix norms will be used: \( \|\cdot\|_2 \), \( \|\cdot\|_{2,\text{col}} \), \( \|\cdot\|_{2,\infty} \). They are respectively defined by: for a matrix \( A = [a_1 \ \cdots \ a_N]^\top \in \mathbb{R}^{n \times N} \), \( \|A\|_2 = \sigma_{\max}(A) \) (i.e., the maximum singular value of \( A \)), \( \|A\|_{2,\text{col}} = \sum_{i=1}^N \|a_i\|_2 \), and \( \|A\|_{2,\infty} = \max_{i=1,...,N} \|a_i\|_2 \).

### 2 Nonsmooth optimization for the estimation problem

#### 2.1 Sparse optimization

The main idea for identifying the parameter vector \( \theta^o \) is by solving a sparse optimization problem, that is, a problem which involves the minimization of the number of nonzeros entries in the error vector. To be more specific, assume for the time being that the error sequence \( \{e_t\} \) is identically equal to zero. Consider a candidate parameter vector \( \theta \in \mathbb{R}^n \) and let

\[ \phi(\theta) = y - X^\top \theta, \]

where \( y = [y_1 \ \cdots \ y_N]^\top \), \( X = [x_1 \ \cdots \ x_N] \), be the fitting error vector induced by \( \theta \) on the experimental data. Then the identification of \( \theta^o \) can be written on the form of an \( \ell_0 \) objective minimization problem as

\[ \min_{\theta \in \mathbb{R}^n} \|\phi(\theta)\|_0 \quad (4) \]

where \( \|\cdot\|_0 \) denotes the \( \ell_0 \) (pseudo) norm which counts the number of nonzero entries. Because problem (4) aims at making the error \( \phi(\theta) \) sparse by minimizing the number of nonzero elements (or maximizing the number of zeros), it is sometimes called a sparse optimization problem (I).

As can be intuitively guessed, the recoverability of the true parameter vector \( \theta^o \) from (4) depends naturally on some properties of the available data. This is outlined by the following lemma.
Lemma 1 (A first sufficient condition for $\ell_0$ recovery). Assume that $\{e_t\}$ is equal to zero and let $f = [f_1 \cdots f_N]^\top$. Assume that for any $I \subset \mathbb{I}$ with $|I| > n$, $f_I \notin \text{im}(X_I^\top)$ whenever $f_I \neq 0$, with im$(\cdot)$ referring here to range space. Then

$$\theta^o \in \arg \min_\theta \|\phi(\theta)\|_0.$$  \hspace{1cm} (5)

Proof. We proceed by contradiction. Assume that (5) does not hold, i.e., $\min_\theta \|\phi(\theta)\|_0 < \|\phi(\theta^o)\|_0$. Then, by letting $\theta^m$ be any vector in $\arg \min_\theta \|\phi(\theta)\|_0$, the inequality translates into $\|I(\theta^m)\| > \|I(\theta^o)\|$. It follows that $f_{I(\theta^m)} \neq 0$ because $\|I(\theta^o)\| = |\{t \in \mathbb{I} : f_t = 0\}|$ is the exact (largest) number of zero elements in the sequence $\{f_t\}_{t=1}^N$. On the other hand, with $y_{I(\theta^m)} = X_{I(\theta^m)}\theta^m = X_{I(\theta^m)}\theta^o + f_{I(\theta^m)}$, it can be seen that $f_{I(\theta^m)} = X_{I(\theta^m)}(\theta^m - \theta^o) \in \text{im}(X_{I(\theta^m)}^\top)$. This, together with $f_{I(\theta^m)} \neq 0$, constitutes a contradiction to the assumption of the Lemma. Hence, (5) holds as claimed. \hfill \Box

Lemma 1 specifies a condition involving both $X$ and $f$ and under which $\theta^o$ lies in the solution set but it does not ensure that $\theta^o$ will be recovered uniquely from data. Before proceeding further, we recall from [1] a sufficient condition under which $\theta^o$ is the unique solution to (5).

Definition 1 (An integer measure of genericity). Let $X \in \mathbb{R}^{n \times N}$ be a data matrix satisfying $\text{rank}(X) = n$. The $n$-genericity index of $X$ denoted $\nu_n(X)$, is defined as the minimum integer $m$ such that any $n \times m$ submatrix of $X$ has rank $n$,

$$\nu_n(X) = \min \left\{ m : \forall S \subset I \text{ with } |S| = m, \text{rank}(X_S) = n \right\}.$$ \hspace{1cm} (6)

Theorem 1 (Sufficient condition for $\ell_0$ recovery). Assume that the sequence $\{e_t\}$ in (1) is identically equal to zero. If the sequence $\{f_t\}$ in (1) contains enough zero values in the sense that

$$|I^0(\theta^o)| = |\{ t \in \mathbb{I} : f_t = 0 \}| \geq \frac{N + \nu_n(X)}{2},$$ \hspace{1cm} (7)

then $\theta^o$ is the unique solution to the $\ell_0$-norm minimization problem (1).

Proof. A proof of this result can be found in [1]. \hfill \Box

In other words, if the number of nonzero gross errors $\{f_t\}$ affecting the data generated by (1) does not exceed the threshold $(N - \nu_n(X))/2$, then $\theta^o$ can be exactly recovered by solving (1). Unfortunately, this problem is a hard combinatorial optimization problem. A more tractable solution can be obtained by relaxing the $\ell_0$-norm into its best convex approximant, the $\ell_1$-norm.
Doing this substitution in (4) gives

$$\minimize_{\theta \in \mathbb{R}^n} \parallel \phi(\theta) \parallel_1 = \sum_{t=1}^{N} \left| y_t - \theta^\top x_t \right|.$$  \hspace{1cm} (8)

The latter problem is termed a nonsmooth convex optimization problem \cite{19, Chap. 3} because the objective function is convex but non-differentiable. Compared to (4), problem (8) has the advantage of being convex and can hence be efficiently solved by many existing numerical solvers, \textit{e.g.}, \cite{14}. Note further that it can be written as a linear programming problem. The $\ell_1$ relaxation process has been intensively used in the compressed sensing literature \cite{12} for approaching the sparsest solution of an underdetermined set of linear equations. As will be shown next, the underlying reason why problem (8) can obtain the true parameter vector despite the presence of gross perturbations $\{f_t\}$ is related to its nonsmoothness.

2.2 Characterization of the solution to the $\ell_1$ problem

A number of sufficient conditions for the equivalence of problems similar to (4) and (8) have been derived in the literature of compressed sensing using the concepts of mutual coherence \cite{12} and the Restricted Isometry Property \cite{6}. Here, we shall propose a parallel but different analysis for the robust estimation problem. We start by characterizing the solution to the $\ell_1$-norm problem (8).

\textbf{Theorem 2} (Characterization of solution to the $\ell_1$ problem). $\theta^* \in \mathbb{R}^n$ solves the $\ell_1$-norm problem (8) iff any of the following equivalent statements hold:

\textit{S1.} There exist some numbers $\lambda_t \in [-1, 1]$, $t \in \mathbb{I}_0(\theta^*)$, such that

$$\sum_{t \in \mathbb{I}^+(\theta^*)} x_t - \sum_{t \in \mathbb{I}^-(\theta^*)} x_t = \sum_{t \in \mathbb{I}_0(\theta^*)} \lambda_t x_t.$$  \hspace{1cm} (9)

\textit{S2.} For any $\eta \in \mathbb{R}^n$,

$$\left| \sum_{t \in \mathbb{I}^+(\theta^*)} \eta^\top x_t - \sum_{t \in \mathbb{I}^-(\theta^*)} \eta^\top x_t \right| \leq \sum_{t \in \mathbb{I}_0(\theta^*)} |\eta^\top x_t|.$$  \hspace{1cm} (10)

\textit{S3.} The optimal value of the optimization problem

$$\min_\alpha \|\alpha\|_\infty \text{ subject to } z = X_{\mathbb{I}_0(\theta^*)} \alpha,$$

where $z = \sum_{t \in \mathbb{I}^+(\theta^*)} x_t - \sum_{t \in \mathbb{I}^-(\theta^*)} x_t$, $\alpha \in \mathbb{R}^{\mathbb{I}_0(\theta^*)}$, is less than or equal to 1.

\footnote{Eq. (9) should be understood here with the implicit convention that any of the three terms is equal to zero whenever the corresponding index set is empty.}
Moreover, the solution $\theta^*$ is unique iff

\( S1' \). \( [9] \) holds and rank(\( X_S \)) = \( n \) where \( S = \{ t \in \mathbb{I}^0(\theta^*) : |\lambda_t| < 1 \} \).

\( S2' \). \( [10] \) holds with strict inequality symbol for all \( \eta \in \mathbb{R}^n, \eta \neq 0 \).

**Proof. Proof of S1.** Since \( \|\phi(\theta)\|_1 \) is a proper convex function, it has a non empty subdifferential \cite{24}. The necessary and sufficient condition for \( \theta^* \) to be a solution of \( [8] \) is then

\[
0 \in \partial \|\phi(\theta^*)\|_1,
\]

where the notation \( \partial \) refers to subdifferential with respect to \( \theta \). Indeed, using additivity of subdifferentials, it is straightforward to write

\[
\partial \|\phi(\theta^*)\|_1 = \sum_{t \in \mathbb{I}^+(\theta^*)} x_t - \sum_{t \in \mathbb{I}^-(\theta^*)} x_t + \sum_{t \in \mathbb{I}^0(\theta^*)} \text{conv} \{-x_t, x_t\} \tag{12}
\]

where \( \text{conv} \) refers to the convex hull. Here, the addition symbol is meant in the Minkowski sum sense. It follows that \( 0 \in \partial \|\phi(\theta^*)\|_1 \) is equivalent to the existence of a set of numbers \( \lambda_t \) in \([-1, 1], t \in \mathbb{I}^0(\theta^*) \), such that \( [9] \) holds.

**Proof of S2.** Define two functions \( q, h : \mathbb{R}^n \to \mathbb{R}_{\geq 0} \) by \( q(\theta) = \sum_{t \notin \mathbb{I}^0(\theta^*)} |y_t - \theta^\top x_t| \) and \( h(\theta) = \sum_{t \in \mathbb{I}^0(\theta^*)} |y_t - \theta^\top x_t| \). Then \( \|\phi(\theta)\|_1 = q(\theta) + h(\theta) \) and \( f \) is differentiable at \( \theta^* \). It follows that \( \partial \|\phi(\theta^*)\|_1 = \nabla q(\theta^*) + \partial h(\theta^*) \), where \( \nabla q(\theta^*) \) is the gradient of \( f \) at \( \theta^* \). We can hence write

\[
\theta^* \text{ minimizes } \|\phi(\theta)\|_1 \iff 0 \in \partial \|\phi(\theta^*)\|_1 \iff -\nabla q(\theta^*) \in \partial h(\theta^*).
\]

Note from \( [12] \) that \( \partial h(\theta^*) = \sum_{t \in \mathbb{I}^0(\theta^*)} \text{conv} \{-x_t, x_t\} \) so that \( -\nabla q(\theta^*) \in \partial h(\theta^*) \) iff \( \pm \nabla q(\theta^*) \in \partial h(\theta^*) \) and this in turn is equivalent to \( g^\top (\theta - \theta^*) \leq h(\theta) - h(\theta^*) \forall \theta, \) for \( g \in \{-\nabla q(\theta^*), +\nabla q(\theta^*)\} \).

It follows that \( \theta^* \) minimizes \( \|\phi(\theta)\|_1 \) iff

\[
|\nabla q(\theta^*)^\top (\theta - \theta^*)| \leq h(\theta) - h(\theta^*) = \sum_{t \in \mathbb{I}^+(\theta^*)} |(\theta - \theta^*)^\top x_t| \forall \theta \tag{13}
\]

The last equality is obtained by using the fact that \( y_t - x_t^\top \theta^* = 0 \) for all \( t \in \mathbb{I}^0(\theta^*) \). Finally the result follows by setting \( \eta = \theta - \theta^* \) and noting that \( \nabla q(\theta^*) = \sum_{t \in \mathbb{I}^+(\theta^*)} x_t - \sum_{t \in \mathbb{I}^-(\theta^*)} x_t \).

**Uniqueness.** For convenience, we first prove \( S2' \). Along the lines of the proof of \( S2 \) (see Eq. \( [13] \) and preceding arguments), we can see that strict inequality in \( [10] \) is equivalent to the following strict inequality \( -\nabla q(\theta^*)^\top (\theta - \theta^*) < h(\theta) - h(\theta^*) \) \( \forall \theta \neq \theta^* \). On the other hand,
\[ \nabla q(\theta^*)^\top (\theta - \theta^*) \leq q(\theta) - q(\theta^*) \quad \forall \theta. \]  
Summing the two yields

\[ \|\phi(\theta^*)\|_1 = q(\theta^*) + h(\theta^*) < q(\theta) + h(\theta) = \|\phi(\theta)\|_1 \quad \forall \theta \neq \theta^*. \]

Hence S2' is proved.

For the proof of S1', we proceed in two steps.

**Sufficiency.** Assume \( \text{rank}(X_S) = n \). Then for any nonzero vector \( \eta \in \mathbb{R}^n \) there is at least one \( t_0 \in S \) such that \( \eta^\top x_{t_0} \neq 0 \). Recall that \( |\lambda_{t_0}| < 1 \) by definition of \( S \). It follows that by multiplying (9) on the left by \( \eta \) with \( \eta \in \mathbb{R}^n \) an arbitrary nonzero vector, and taking the absolute value yields (10) with strict inequality symbol. We can therefore apply the proof of S2' to conclude that \( \theta^* \) is unique.

**Necessity.** Assume \( \text{rank}(X_S) < n \). Then pick any nonzero vector \( \eta \in \ker(X_S^\top) \). Set \( \eta_1 = \nu \eta \) with \( \nu \neq 0 \). Indeed \( \nu \) can be chosen sufficiently small such that \( x_1^\top (\theta^* + \eta_1) - y_t \) has the same sign as \( x_1^\top \theta^* - y_t \). This implies that \( \mathbb{I}^+(\theta^*) \subset \mathbb{I}^+(\theta^* + \eta_1) \) and \( \mathbb{I}^-(\theta^*) \subset \mathbb{I}^-(\theta^* + \eta_1) \). Moreover, since \( \eta_1 \in \ker(X_S^\top) \), \( x_1^\top (\theta^* + \eta_1) - y_t = \eta_1^\top x_t = 0 \quad \forall \ t \in S \), so that \( S \subset \mathbb{I}^0(\theta^* + \eta_1) \). Finally, it remains to re-assign the indices \( t \) contained in \( \mathbb{I}^0(\theta^*) \setminus S \) for which \( \lambda_t = 1 \). We get the following partition \( \mathbb{I}^+(\theta^* + \eta_1) = \mathbb{I}^+(\theta^*) \cup \{ t \in \mathbb{I}^0(\theta^*): \eta_1^\top x_t > 0 \} \), \( \mathbb{I}^-(\theta^* + \eta_1) = \mathbb{I}^-(\theta^*) \cup \{ t \in \mathbb{I}^0(\theta^*): \eta_1^\top x_t < 0 \} \), \( \mathbb{I}^0(\theta^* + \eta_1) = S \cup \{ t \in \mathbb{I}^0(\theta^*): \eta_1^\top x_t = 0 \} \). It follows that \( \theta^* + \eta_1 \neq \theta^* \) also satisfies (9) with the sequence \( \{\lambda_t\}_{t \in S} \) and is therefore a minimizer. In conclusion, if \( \text{rank}(X_S) < n \), the minimizer cannot be unique. \( \square \)

A number of important comments follow from Theorem 2:

- One first consequence of the theorem is that \( \theta^\circ \) can be computed exactly from a finite set of erroneous data provided it satisfies the conditions of the theorem. Note that there is no explicit boundedness condition imposed on the error sequence \( \{f_t\} \). Hence the nonzero errors in this sequence can have arbitrarily large magnitudes as long as the optimization problem makes sense, i.e., provided \( \|\phi(\theta^*)\|_1 \) remains finite.

- Second, \( \theta^\circ \) can be exactly recovered in the presence of, say, infinitely many nonzero errors \( f_t \) (see Proposition 1). For example, if the regressors \( \{x_t\} \) satisfy

\[ \sum_{t \in \mathbb{I}^+(\theta^\circ)} x_t - \sum_{t \in \mathbb{I}^-(\theta^\circ)} x_t = 0, \]

and \( \text{rank}(X_{\|y\|_2}) = n \), then \( \theta^\circ \) is the unique solution to problem (5) regardless of the number of errors affecting the data.
Third, if problem (8) admits a solution \( \theta^* \) that satisfies \( y_t - x_t^\top \theta^* \neq 0 \) for all \( t = 1, \ldots, N \), then \( \theta^* \) is non-unique. In effect, \( I^0(\theta^*) = \emptyset \) in this case and so, \( \operatorname{rank}(X_{I^0(\theta^*)}) = 0 < n \) which, by Theorem 2 implies non-uniqueness. Indeed this is typically the case whenever the noise \( \{e_t\} \) is nonzero.

Another immediate consequence of Theorem 2 can be stated as follows.

**Corollary 1** (On the special case of affine model). If the model (1) is affine in the sense that the regressor \( x_t \) has the form \( x_t = [\tilde{x}_t^\top 1]^\top \), with \( \tilde{x}_t \in \mathbb{R}^{n-1} \), then a necessary condition for \( \theta^* \) to solve problem (8) is that

\[
\left| |I^+(\theta^*)| - |I^-(\theta^*)| \right| \leq |I^0(\theta^*)|. \tag{14}
\]

Here, the outer bars \( |\cdot| \) refer to absolute value while the inner ones which apply to sets refer to cardinality.

**Proof.** The proof is immediate by considering the condition (10) and taking \( \eta = [0^\top 1]^\top \in \mathbb{R}^n \).

Eq. (14) implies that if the measurement model is affine and all the \( f_t \)'s have the same sign, i.e., if one of the cardinalities \( |I^+(\theta^*)| \) or \( |I^-(\theta^*)| \) is equal to zero, then problem (8) cannot find the true \( \theta^o \) whenever more than 50% of the gross errors \( f_t \) are nonzero.

Next, we discuss a special case in which the true parameter vector \( \theta^o \) in (1) can, in principle, be obtained asymptotically in the presence of an infinite number of errors \( f_t \)'s.

**Proposition 1** (Infinite number of outliers). Assume that the error sequence \( \{e_t\} \) in (1) is identically equal to zero. Assume further that the data are generated such that

- There is a set \( I^0 \) with \( |I^0| \geq n \), such that for any \( t \in I^0 \), \( f_t = 0 \) and \( \operatorname{rank}(X_{I^0}) = n \),
- For any \( t \notin I^0 \), \( f_t \) is sampled from a uniform distribution.
- The regression vector sequence \( \{x_t\} \subset \mathbb{R}^n \) is drawn from a probability distribution having a finite first moment.

Then

\[
\lim_{N \to \infty} \arg \min_{\theta \in \mathbb{R}^n} \frac{1}{N} \sum_{t=1}^{N} \left| y_t - x_t^\top \theta \right| = \{ \theta^o \}. \tag{15}
\]

**Proof.** Under the conditions of the proposition, we have \( \operatorname{Prob}(y_t - x_t^\top \theta^o < 0) = \operatorname{Prob}(y_t - x_t^\top \theta^o > 0) = 1/2 \), where \( \operatorname{Prob} \) denotes the uniform probability measure. It follows that \( |I^+(\theta^o)| \) and \( |I^-(\theta^o)| \) go jointly to infinity as the total number of samples \( N \) tends to infinity. Hence,
the expressions \( \frac{1}{\|\|^{(\theta^o)} \|} \sum_{t \in I^{+}(\theta^o)} x_t \) and \( \frac{1}{\|\|^{(\theta^o)} \|} \sum_{t \in I^{-}(\theta^o)} x_t \) are both sample estimates for the expectation of the process \( \{x_t\} \). As \( N \to \infty \), the two quantities converge to the true expectation of the process \( \{x_t\} \), so that

\[
\lim_{N \to \infty} \left[ \frac{1}{\|\|^{(\theta^o)} \|} \sum_{t \in I^{+}(\theta^o)} x_t - \frac{1}{\|\|^{(\theta^o)} \|} \sum_{t \in I^{-}(\theta^o)} x_t \right] = 0.
\]

As a consequence, \( \theta^o \) satisfies condition S1’ of Theorem 2 asymptotically with \( \lambda_t = 0 \) for any \( t \in \mathbb{I}^0(\theta^o) = \mathbb{I}^0 \). Hence the solution of the \( \ell_1 \) minimization problem tends to \( \theta^o \) as the number of samples approaches infinity.

### 2.3 Sufficient conditions for recoverability by convex optimization

In this section we derive sufficient conditions under which \( \theta^o \) in model (1) belongs to the solution set of problem (8). The conditions say essentially that if the number of zero entries in vector \( f \) is larger than certain thresholds depending on the data matrix \( X \), then \( \theta^o \) solves (8). To some extent, the thresholds can be interpreted as quantitative measures of the richness (genericity) of \( X \). Our derivations rely on Theorem 2.

**Theorem 3** (Sufficient conditions for exact recovery). Suppose that the sequence \( \{e_t\} \) in model (1) is zero and \( \text{rank}(X) = n \). Then the following hold.

- There exists a number \( k_1(X) \), \( k_1(X) \geq \nu_n(X) \), such that for any partition \( (I, I^c) \) of \( \mathbb{I} \),

\[
\left\| X^T_I (X_I X_I^T)^{-1} X_{I^c} \right\|_\infty \leq 1 \quad (16)
\]

whenever \( |I| \geq k_1(X) \).

- There exists a number \( k_2(X) \) such that for any partition \( (I, I^c) \) of \( \mathbb{I} \),

\[
\left\| X^T_{I^c} (X X^T)^{-1} X \right\|_1 \leq 1/2 \quad (17)
\]

whenever \( |I| \geq k_2(X) \).

Moreover, any vector \( \theta \in \mathbb{R}^n \) obeying \( \|\|^{0(\theta)} \| \geq \min(k_1(X), k_2(X)) \), solves the \( \ell_1 \) minimization problem (8).

**Proof.** Part 1: existence of \( k_i(X) \), \( i = 1, 2 \). Since \( \text{rank}(X) = n \), we have \( \nu_n(X) \leq N \) so that \( \nu_n(X) \) is finite. Hence the condition \( k_i(X) \geq \nu_n(X) \) can be satisfied. Moreover, by considering the trivial partition \( (I, I^c) \) with \( I = \mathbb{I} \) and \( I^c = \emptyset \), we see that a possible (the largest
indeed) value for \( k_{i}(X) \) is \( N \). Define 
\[ v_{1}(k) = \arg \max_{(I,I^{c})} \| X_{I}^{\top} (X_{I} X_{I}^{\top})^{-1} X_{I^{c}} \|_{\infty} \]
and 
\[ v_{2}(k) = \arg \max_{(I,I^{c})} \| X_{I^{c}}^{\top} (X X^{\top})^{-1} X_{I} \|_{1} \],
where the maximum is taken over the set of those partitions \((I,I^{c})\) of \( \mathbb{I} \) that satisfy \( |I| = k \). It can be easily shown that \( v_{1}(k) \) and \( v_{2}(k) \) are decreasing functions of \( k \). By setting \( k_{1}(X) = \min_{k \in \mathbb{I}} \{ k : v_{1}(k) \leq 1 \} \) and \( k_{2}(X) = \min_{k \in \mathbb{I}} \{ k : v_{2}(k) \leq 1/2 \} \), it appears that \( v_{1}(k) \leq 1 \) whenever \( k \geq k_{1}(X) \) and \( v_{2}(k) \leq 1/2 \) whenever \( k \geq k_{2}(X) \).

**Part 2.** Assume that \( \| \theta_{0} \| \geq k_{1}(X) \). Then, by definition of \( k_{1}(X) \), (18) holds with \( I = \theta_{0} \) and \( I^{c} = \mathbb{I} \setminus \theta_{0} \). We just need to show that (18) is a sufficient condition for \( \theta \) to solve (8). It is enough that \( \theta \) satisfies Condition S3 of Theorem 2. In turn, a sufficient condition for S3 is that 
\[ \alpha^{*} = X_{I^{c}}^{\top} (X_{I} X_{I}^{\top})^{-1} X_{I^{c}} = \arg \max_{v} \| X_{I^{c}}^{\top} (X X^{\top})^{-1} X_{I} \|_{1} \]

be smaller or equal to one for any \( h \in \{ \pm 1 \}^{|I^{c}|} \). Let \( p^{*} \) be this optimal value for a fixed but arbitrary \( h \) and pose 
\[ \alpha^{*} = \arg \min_{\alpha} \| h - \alpha \|_{2} \text{ s.t. } X_{I^{c}}^{\top} h = X_{I}^{\top} \alpha \].
Since \( \alpha^{*} \) is a feasible point for problem (15), we see that \( \| \alpha^{*} \|_{\infty} \leq 1 \) is a sufficient condition for \( p^{*} \leq 1 \). The so-defined \( \alpha^{*} \) is the well-known least euclidean-norm solution to an underdetermined system of linear equations [4]; \( \alpha^{*} \) can be analytically expressed as 
\[ \alpha^{*} = X_{I}^{\top} (X_{I^{c}} X_{I^{c}}^{\top})^{-1} X_{I^{c}} h \].
As a consequence, we get a more restrictive sufficient condition as 
\[ \| X_{I}^{\top} (X_{I} X_{I}^{\top})^{-1} X_{I} h \|_{\infty} \leq 1 \].
This is the same as condition (16) since \( \| h \|_{\infty} = 1 \).

**Part 3.** Assume that \( \| \theta_{0} \| \geq k_{2}(X) \). Then, by definition of \( k_{2}(X) \), (17) holds with \( I = \theta_{0} \) and \( I^{c} = \mathbb{I} \setminus \theta_{0} \). By following a similar reasoning as in Part 1, we just need to show that (17) is a sufficient condition for \( \theta \) to solve (8). For this purpose, note from the condition S2 of Theorem 2 that for \( \theta \) to be a minimizer of (8), it suffices that 
\[ \sum_{t \in I^{c}} \eta_{t} x_{t} \leq \sum_{t \in I} \eta_{t} x_{t} \]
for any \( \eta \in \mathbb{R}^{n} \). By adding the term \( \sum_{t \in I^{c}} \eta_{t} x_{t} \) on both sides of the inequality symbol, it follows that 
\[ \sum_{t \in I^{c}} \eta_{t} x_{t} \leq \frac{1}{2} \| X \eta \|_{1} \].
It is therefore enough that 
\[ \max_{\| X \eta \|_{1} = 1} \| X_{I}^{\top} \eta \|_{1} \leq \frac{1}{2} \]
where the maximum is taken with respect to \( \eta \in \mathbb{R}^{n} \). By letting \( b = X \eta \) and hence replacing \( \eta \)
with \( \eta = (XX^\top)^{-1}Xb \), the previous condition becomes

\[
\max_{\|b\|_1=1} \left\| X_J^\top (XX^\top)^{-1}Xb \right\|_1 = \left\| X_J^\top (XX^\top)^{-1}X \right\|_1 \leq 1/2,
\]

which corresponds exactly to (17). \( \square \)

The conditions proposed in Theorem 3 are only sufficient and hence necessarily conservative. To get a sense of how conservative those conditions are, it would be useful to be able to compute the thresholds \( k_1(X) \) and \( k_2(X) \). However, these numbers are very hard to evaluate numerically because of the combinatorial optimization involved. Next, we investigate another sufficient condition for the solution of (8) to coincide with \( \theta^o \) [2]. This last condition is more restrictive than those of Theorem 3 in the sense that the corresponding threshold is an upper bound on \( k_1(X) \) and \( k_2(X) \). But the new threshold can be approximated.

**Theorem 4** (Another sufficient condition of recoverability). Assume that the sequence of errors \( \{e_t\} \) is zero in (2) and that \( \text{rank}(X) = n \). Define the number

\[
r(X) = \max_{z \in \mathbb{R}^N, \|z\|_1=1} \left\| (XX^\top)^{-1}Xz \right\|_2.
\]

If

\[
|\mathbb{I}^o(\theta^o)| > N - \frac{1}{2\rho r(X)},
\]

with \( \rho = \max_{t=1,...,N} \|x_t\|_2 \) and \( |\mathbb{I}^o(\theta^o)| \) denoting the cardinality of \( \mathbb{I}^o(\theta^o) \), then \( \theta^o \) is the unique minimizer of (8).

**Proof.** The proof follows by similar arguments as in Part 3 of the proof of Theorem 3 with the slight difference that the uniqueness condition S2' (see Theorem 2) is invoked. For \( \theta^o \) to be the unique minimizer of (8), it is sufficient that (19) holds with strict inequality symbol. By the Cauchy-Schwarz inequality, \( \|X_J^\top \eta\|_1 \leq \rho |F^c| \|\eta\|_2 \). The conclusion now follows from the observation that \( \max_{\|\eta\|_1=1} \|\eta\|_2 = r(X) \). \( \square \)

Theorem 4 says that under noise-free assumption, if the number of zero entries in \( f \) is larger than a certain threshold (depending on the degree of genericity of the data), then the true parameter vector \( \theta^o \) can be computed efficiently by solving the convex problem (8). It follows from (21) that it is desirable to have \( r(X) \) be as small as possible. This is indeed a property of richness on the measured data matrix \( X \). The richer the regressors \( \{x_t\} \), the smaller \( r(X) \).

Because \( r(X) \) might be a bit difficult to compute directly in the form (20), we give some rough estimates of it in the following lemma.
Lemma 2 (Estimation of \( r(X) \)). \textit{Under the assumption that} \( \text{rank}(X) = n \), \( r(X) \) \textit{satisfies the following inequalities:}

1. \[
\frac{1}{\|X\|_{2,\text{col}}} \leq r(X) \leq \left\| X^\top (XX^\top)^{-1} \right\|_2 \tag{22}
\]
\textit{where} \( \|X\|_{2,\text{col}} = \sum_{t=1}^{N} \|x_t\|_2 \).

2. \[
\frac{1}{r(X)} \leq \rho(N - \nu_n(X) + 1). \tag{23}
\]

\textbf{Proof.} \textbf{Proof of (22).} By the Cauchy-Schwarz inequality, we have \( \left\| X^\top \eta \right\|_1 \leq \|X\|_{2,\text{col}} \|\eta\|_2 \) for any \( \eta \in \mathbb{R}^n \). It follows that

\[r(X) = \max_{\eta \neq 0} \frac{\|\eta\|_2}{\|X^\top \eta\|_1} \geq \frac{\|\eta\|_2}{\|X\|_{2,\text{col}}} \geq \frac{1}{\|X\|_{2,\text{col}}}.\]

Hence the first inequality in (22) is proved. The second inequality of (22) is also immediate by noting that

\[
r(X) \leq \left\| X^\top (XX^\top)^{-1} \right\|_2 \max_{z \in \mathbb{R}^N, \|z\|_1 = 1} \|z\|_2
\]
\textit{and} \( \max_{z \in \mathbb{R}^N, \|z\|_1 = 1} \|z\|_2 = 1 \).

\textbf{Proof of (23).} For \( \eta \in \mathbb{R}^n \), let \( T(\eta) = \text{supp} (X^\top \eta) = \{ t \in \mathbb{I} : x_t^\top \eta \neq 0 \} \). Then

\[
r(X) = \max_{\eta \neq 0} \frac{\|\eta\|_2}{\|X^\top_{T(\eta)} \eta\|_1}.
\]

By invoking again the Cauchy-Schwarz inequality, we get

\[
r(X) \geq \max_{\eta \neq 0} \frac{1}{\|X^\top_{T(\eta)} \eta\|_{2,\text{col}}} \geq \max_{\eta \neq 0} \frac{1}{\rho|T(\eta)|}
\]
\textit{with} \( |T(\eta)| \) \textit{standing for the cardinality of} \( T(\eta) \). When \( \eta \neq 0 \), the smallest value \( |T(\eta)| \) \textit{can take is} \( N - \nu_n(X) + 1 \). It can therefore be concluded that \( r(X) \geq 1/(\rho(N - \nu_n(X) + 1)) \) which is equivalent to (23).

\textbf{Theorem 5 (\( \ell_0/\ell_1 \) equivalence).} \textit{If a parameter vector} \( \theta^o \in \mathbb{R}^n \) \textit{satisfies condition (21), then it is the unique solution to both the} \( \ell_0 \) \textit{problem (4) and the} \( \ell_1 \) \textit{problem (8).}

\textbf{Proof.} If \( \theta^o \) \textit{satisfies condition (21), then by applying Eq. (23), it is easy to see that} \( \|\phi(\theta^o)\|_0 < 1/2(N - \nu_n(X) + 1) \). \textit{This last condition holds iff} \( \theta^o \) \textit{satisfies (7). Hence} \( \theta^o \) \textit{satisfies both the conditions of Theorem [1] and Theorem [4]. The result to be proved follows.}
3 Some implementation aspects

3.1 Enforcing recoverability by the $\ell_1$ norm minimization

The parameter vector $\theta^o$ from the model (1) can be uniquely recovered by solving the convex problem (8) if $\theta^o$ satisfies the condition (21) of Theorem 4. In case this condition is not naturally satisfied, an interesting question is how we can process the data in order to promote it. In this section we discuss an algorithmic strategy for enhancing the recoverability of $\theta^o$ by $\ell_1$ minimization. Our discussion is inspired by [7]. The idea is to solve a sequence of problems of the type (8) with different weights computed iteratively [7, 1]. The iterative scheme can be defined for a fixed number $r_{max}$ of iterations as follows. At iteration $r = 0, \ldots, r_{max}$, compute

$$\hat{\theta}(r) = \arg \min_{\theta} \sum_{t=1}^{N} w_t^{(r)} |y_t - \theta^\top x_t|,$$

with weights defined, for all $t$, by $w_t^{(0)} = 1/N$, and

$$w_t^{(r)}(t) = \frac{\xi_t^{(r)}}{\sum_{t=0}^{N-1} \xi_t^{(r)}}, \quad \text{if } r \geq 1,$$

where

$$\xi_t^{(r)} = \frac{1}{|y_t - x_t^\top \hat{\theta}^{(r-1)}| + \delta},$$

with $\delta > 0$ a small number whose role is to prevent division by zero and $r$ is the iteration number. Note that there are many other reweighting strategies which can be used in (24), see e.g., [10, 27, 16]. Since we are dealing here with a sequence of convex optimization problems, they can be numerically implemented using any convex solver. In particular the CVX software package [14] solves efficiently this category of problems in a Matlab environment.

3.2 On the treatment of the noise $\{e_t\}$

The formulations (4) and (8) are convenient when the noise $\{e_t\}$ is equal to zero. Nevertheless, they are expected to work in the presence of a moderate amount of noise. To take explicitly the noise $\{e_t\}$ into account, we propose to compute estimates $\hat{e} \in \mathbb{R}^N$ and $\varphi \in \mathbb{R}^N$ (of $e$ and $f$ respectively) by minimizing a cost of the form $\|\hat{e}\|_2^2 + \lambda \|\varphi\|_0$ under an equality constraint of the form (1). In other words, we consider the problem

$$\min_{(\hat{e}, \varphi) \in \mathbb{R}^n \times \mathbb{R}^N} \left[ \frac{1}{2} \|y - X^\top \theta - \varphi\|_2^2 + \lambda \|\varphi\|_0 \right],$$

(25)
and its convex relaxation,

\[
\begin{align*}
\text{minimize}_{(\theta, \varphi) \in \mathbb{R}^n \times \mathbb{R}^N} & \quad \frac{1}{2} \| y - X^\top \theta - \varphi \|^2_2 + \lambda \| \varphi \|_1, \\
\text{subject to } & \quad \lambda \geq 0. \tag{26}
\end{align*}
\]

where \( \lambda \geq 0 \) is a regularization parameter.

**Lemma 3.** A pair \((\theta^*, \varphi^*) \in \mathbb{R}^n \times \mathbb{R}^N\) solves (26) iff it satisfies

\[
\begin{align*}
XX^\top \theta^* - X(y - \varphi^*) &= 0, \tag{27} \\
X^\top \theta^* - (y - \varphi^*) &= -\lambda s(\varphi^*), \tag{28}
\end{align*}
\]

where \( s(\varphi^*) \) is a vector in \( \mathbb{R}^N \) whose entries \( s_i(\varphi^*) \), \( i = 1, \ldots, N \), are defined by: \( s_i(\varphi^*) = \text{sign}(\varphi^*_i) \) if \( \varphi^*_i \neq 0 \) and \( s_i(\varphi^*) \in [-1, 1] \) if \( \varphi^*_i = 0 \).

**Proof.** Let \( l(\theta, \varphi) = \frac{1}{2} \| y - X^\top \theta - \varphi \|^2_2 + \lambda \| \varphi \|_1 \) be the objective function of the problem (26). Then \( l \) is a proper convex function which is differentiable with respect to variable \( \theta \) on \( \mathbb{R}^n \) and admits a subdifferential at any variable \( \varphi \in \mathbb{R}^N \). \((\theta^*, \varphi^*)\) minimizes \( l(\theta, \varphi) \) iff \( 0 = \nabla_{\theta} l(\theta^*, \varphi^*) \) and \( 0 \in \partial_{\varphi} l(\theta^*, \varphi^*) \). These conditions translate immediately into \( XX^\top \theta^* - X(y - \varphi^*) = 0 \) and \(- (y - \varphi^* - X^\top \theta^*) + \lambda s(\varphi^*) = 0 \), where \( s(\varphi^*) \in \partial \| \varphi^* \|_1 \) is any subgradient of \( \| \varphi \|_1 \) at \( \varphi^* \).

It is interesting to note that (27)-(28) imply \( Xs(\varphi^*) = 0 \), which is very similar to (9). The following lemma characterizes the uniqueness of the solution of (26).

**Lemma 4 (Uniqueness of solution to (26)).** A pair \((\theta^*, \varphi^*)\) is the unique solution to problem (26) iff both of the following statements are true

(i) \((\theta^*, \varphi^*)\) satisfies conditions (27)-(28)

(ii) \( \text{rank}(X) = n \) and \( \text{rank}(\Psi_{1,S^c}) = |S^c| \).

Here, \( \Psi = I_N - X^\top (XX^\top)^{-1}X \), with \( I_N \) being the identity matrix of order \( N \), \( \Psi_{I,S^c} \) is a matrix formed with the columns of \( \Psi \) indexed by \( S^c \) defined by \( S^c = \mathbb{I} \setminus S \), with \( S = \{ t \in \mathbb{I} : |s_t(\varphi^*)| < 1 \} \).

The expression of \((\theta^*, \varphi^*)\) is then given by

\[
\begin{align*}
\theta^* &= (XX^\top)^{-1}X(y - \varphi^*), \tag{29} \\
\varphi^*_{S^c} &= (\Psi_{I,S^c}^\top \Psi_{I,S^c})^{-1} \Psi_{I,S^c}^\top (\Psi y - \lambda s(\varphi^*)) = 0. \tag{30}
\end{align*}
\]

**Proof.** \( l(\theta, \varphi) \) is a quadratic function of \( \theta \). For a fixed \( \varphi^* \), the minimizer \( \theta^* \) of \( l(\theta, \varphi^*) \) is unique iff \( X \) has full row rank, i.e., \( \text{rank}(X) = n \). The unique value of \( \theta^* \) is expressed in function of \( \varphi^* \).
by (29). Plugging the expression (29) of \( \theta^* \) in the objective gives

\[
\tilde{l}(\varphi) \triangleq l(\theta^*, \varphi) = \frac{1}{2} \| \Psi y - \Psi \varphi \|_2^2 + \lambda \| \varphi \|_1.
\]

The rest of the proof then boils down to showing that the minimizer \( \varphi^* \) of \( \tilde{l}(\varphi) \) is unique iff \( \text{rank}(\Psi_{1,S^c}) = |S^c| \). To begin with, let us point out the following (see also [25]). If \( \varphi^* \) and \( \xi^* \) are two minimizers of \( \tilde{l}(\varphi) \), then we have necessarily

\[
\Psi \varphi^* = \Psi \xi^*,
\]

(31)

\[
s(\varphi^*) = s(\xi^*).
\]

(32)

The relation (31) follows from the strict convexity of \( \tilde{l}(\varphi) \) as a function of \( \Psi \varphi \). In effect, by changing the optimization variable into \( \delta = \Psi \varphi \), \( \tilde{l}(\varphi) \) becomes \( \frac{1}{2} \| \Psi y - \delta \|_2^2 + \lambda \| \Psi \delta + v \|_1 \), with \( v \) a vector in \( \ker(\Psi) \) and \( \dagger \) referring to generalized inverse. This last function is strictly convex with respect to \( \delta \). As a consequence, its minimizer is unique and equal to \( \delta^* = \Psi \varphi^* \). To see why the relation (32) holds, plug the expression (29) of \( \theta^* \) in (28). We get \( \lambda s(\varphi^*) = \Psi y - \Psi \varphi^* \). Combining this with (31) (i.e., the uniqueness of \( \Psi \varphi^* \)) yields immediately (32).

**Sufficiency.** Assume that \( \text{rank}(\Psi_{1,S^c}) = |S^c| \). As argued above, any two minimizers \( \varphi^* \) and \( \xi^* \) of \( \tilde{l}(\varphi) \) obey (31)-(32). From (32) we get that \( S \subset \{ t \in I : \xi^*_t = 0 \} \), which implies that \( S^c \supset \text{supp}(\xi^*) \). As a consequence, we can write (31) in the following reduced form \( \Psi_{1,S^c}(\varphi^* - \xi^*) = 0 \). With \( \text{rank}(\Psi_{1,S^c}) = |S^c| \), this implies that \( \varphi^* = \xi^* \) and that the minimizer of \( \tilde{l}(\varphi) \) is unique.

**Necessity.** Assume that \( \text{rank}(\Psi_{1,S^c}) < |S^c| \). Consider a nonzero vector \( \eta \in \mathbb{R}^N \) such that \( \eta_{S^c} = 0 \) and \( \eta_{S^c} \in \ker(\Psi_{1,S^c}) \). Let \( \eta_1 = \nu \eta \), with \( \nu \neq 0 \). It is straightforward to verify that \( \Psi \varphi^* = \Psi(\varphi^* + \eta_1) \). Note that \( \nu \) can be chosen sufficiently small such that \( \varphi^*_t \) and \( \varphi^*_t + \eta_{1,t} \) have the same sign whenever \( \varphi^*_t \neq 0 \). Following a similar path as in the proof of Theorem 2, we can establish that \( s(\varphi^*) = s(\varphi^* + \eta_1) \). Finally, with \( \Psi \varphi^* = \Psi(\varphi^* + \eta_1) \), \( s(\varphi^*) = s(\varphi^* + \eta_1) \) and the fact that \( \varphi^* \) is an optimal solution (hence satisfying (28)), it is easy to check that \( \varphi^* + \eta_1 \) also satisfies (28). By Lemma 3 \( \varphi^* + \eta_1 \) is not unique. Hence, the solution is not unique.

**Derivation of Eqs (29)-(29).** These relations result from simple rearrangements of (27)-(28). \[ \square \]

From Lemma 1 it appears that the true vector \( f \) can be found by problem (26) if there is a vector \( \hat{\theta} \in \mathbb{R}^n \) such that \((\hat{\theta}, f) \) satisfies the conditions (i)-(ii) of Lemma 1. In particular, \((\hat{\theta}, f) \) must satisfy (28). A necessary condition for this is that \( \Psi e = \lambda s(f) \). And this implies that the regularization parameter must verify \( \lambda \geq \| \Psi e \|_{\infty} \) when \( f = 0 \), and \( \lambda = \| \Psi e \|_{\infty} \) when \( f \neq 0 \).

\[ \text{It is to be noted that the analysis in [25] provides only a sufficient condition.} \]
Note further that if \( e = 0 \) and \( f \neq 0 \), then \( \lambda \) must be equal to zero! However, if \( \lambda \) is set to zero in (26), then the solution set is

\[
\left\{ (\theta, \varphi) : \theta = (XX^\top)^{-1}X(y - \varphi), \varphi \in y + \text{im}(X^\top) \right\}.
\]

Since this set contains infinitely many elements, we conclude that it is unlikely to get exactly the true \( f \) by solving (26) irrespective of the value of the regularization parameter \( \lambda \).

4 Extension to multivariable systems

We consider now the multivariable analogue of model (1) written in the form

\[
y_t = A^o x_t + f_t + e_t,
\]

where \( y_t \in \mathbb{R}^m \) is the output vector at time \( t \), \( \{f_t\} \subset \mathbb{R}^m \) is the sequence of errors, \( \{e_t\} \subset \mathbb{R}^m \) is the noise sequence, \( A^o \in \mathbb{R}^{m \times n} \) is the parameter matrix.

The question of interest is how to recover the matrix \( A^o \) from measurements corrupted by a vector sequence of sparse errors \( \{f_t\} \). The sparse optimization approach is still applicable to this case, that is, we can formulate the estimation problem as

\[
\min_{A \in \mathbb{R}^{m \times n}} \left| \left| \left\{ t : y_t - Ax_t \neq 0 \right\} \right| \right|
\]

with \( | \cdot | \) standing for cardinality. It can be easily verified that Theorem 1 applies to (34) as well.

The convex relaxation takes the form of a nonsmooth optimization with a cost functional consisting of a sum-of-norms of errors [21, 9, 3],

\[
\min_{A \in \mathbb{R}^{m \times n}} \sum_{t=1}^{N} \| y_t - Ax_t \|_2
\]

with \( \| \cdot \|_2 \) referring to the Euclidean norm.

**Theorem 6.** A matrix \( A^* \in \mathbb{R}^{m \times n} \) solves the sum-of-norms problem (35), if and only if any of the following equivalent statements holds

**T1.** There exists a sequence of vectors \( \{\beta_t\}_{t \in \mathbb{N}}(A^*) \subset B_2(0, 1) \) such that

\[
\sum_{t \notin \mathbb{N}(A^*)} v_t^* x_t^\top + \sum_{t \in \mathbb{N}(A^*)} \beta_t x_t^\top = 0,
\]

17
where \( v_t^* = (y_t - A^* x_t) / \|y_t - A^* x_t\|_2 \). Here, \( B_2(0, 1) \subset \mathbb{R}^m \) is the Euclidean unit ball of \( \mathbb{R}^m \).

**T2.** For any matrix \( \Lambda \in \mathbb{R}^{m \times n} \),

\[
\left| \sum_{t \in I^0(A^* \{A^* \})} v_t^* \top \Lambda x_t \right| \leq \sum_{t \in I^0(A^* \{A^* \})} \|\Lambda x_t\|_2.
\]  
(37)

**T3.** The optimal value of the problem

\[
\min_{Z \in \mathbb{R}^{m \times p}} \|Z\|_{2, \infty} \quad \text{subject to} \quad V^* X_{I^0(A^* \{A^* \})}^\top = Z X_{I^0(A^* \{A^* \})}^\top
\]

\[
p = |I^0(A^*)| \quad \text{and} \quad V^* \quad \text{being a matrix formed with the unit 2-norm vectors} \quad v_t^*, \quad \text{for} \quad t \in \mathbb{I} \setminus I^0(A^*), \quad \text{is smaller than 1.}
\]

Moreover, the solution \( A^* \) is unique iff

**T1’.** \((36)\) holds and \( \text{rank}(X_T) = n \) where \( T = \{ t \in I^0(A^*) : \|\beta_t\|_2 < 1 \} \).

**T2’.** \((37)\) holds with strict inequality symbol for all \( \Lambda \in \mathbb{R}^{m \times n}, \Lambda \neq 0 \).

**Proof.** The proof is similar to that of Theorem 2. It is therefore omitted here. \( \square \)

It is interesting to note that based on Theorem 6, the analysis carried out in the previous sections can be easily generalized to the multivariable case. In particular, Proposition 1 and Theorems 3-5 can be restated for the multivariable model (33) with only some slight modifications.

## 5 Numerical illustration

### 5.1 Static models subject to intermittent gross errors

In our first experiment we consider static linear and affine models of the form (1) with \( n = 4 \) and \( N = 500 \). The affine model refers to the case where the regressor \( x_t \) has the form \( x_t = [\tilde{x}_t^\top 1]^\top \). The goal is to estimate the probability of exact recovery of the true parameter vector by problem \((8)\) in function of the number of nonzero elements in the sequence \( \{f_t\} \). For this purpose, the noise \( \{e_t\} \) is set to zero. The nonzero elements of \( \{f_t\} \) are drawn from a Gaussian distribution with mean 100 and variance 1000. For each level of sparsity (i.e., proportion on nonzeros), a Monte Carlo simulation of size 100 is carried out with randomly generated static/affine models and 500 data samples at each run. Repeating this for four situations (linear/affine and linear/affine with positive \( f_t \)'s), we obtain the results depicted in Figure 1. We observe that in the linear...
case, problem (8) provides the true parameter vector when the output is affected by up to 80% of nonzero gross errors. This is because the data \( \{x_t\} \) which were sampled from a Gaussian distribution are very generic. In the affine case, the performance is a little less good. If we set all \( f_t \)'s to have the same sign, then as suggested by condition (14), the percentage of outliers that can be corrected by the optimization problem (8) cannot exceed 50%.

Figure 1: Estimates of probabilities of exact recovery when noise \( \{e_t\} \) is equal to zero.

5.2 Static models with both noise and gross errors

Consider now the case of static models of the form (1) in the presence of both \( \{e_t\} \) and \( \{f_t\} \) sampled from Gaussian distributions \( \mathcal{N}(0, \sigma_e^2) \) and \( \mathcal{N}(100, 1000) \) respectively. The variance \( \sigma_e^2 \) is selected so as to achieve a certain signal to noise ratio before the gross error sequence is added to the output. Again, by carrying out a Monte-Carlo simulation of size 100 with different sparsity levels and randomly generated models at each run, we obtain the average errors plotted in Figure 2. It turns out that the results returned by problems (8) and (26) with \( \lambda = 0.10 \) are almost the same for an SNR in \( \{10 \text{ dB}, 20 \text{ dB}\} \). The performance can be assessed by comparing with an "oracle" estimate \( i.e., \) the least squares estimate one would obtain if the locations of zeros in the sequence \( \{f_t\} \) were known. The results in Figure 2 tend to suggest that the proposed approach performs very well. For the current numerical experiment, our results are very close to the ideal estimate when the proportion of nonzeros is less than 70%.

5.3 Dynamic linear models subject to sensor intermittent faults

In the case when (1) represents a dynamic ARX model subject to gross errors, it can be observed (see Fig. 3) that the probabilities of exact recovery are much smaller than in the static case studied in Section 5.1. This difference is related to the richness (or genericity) of the regression
vectors (columns of $X$) involved in each case. In the static example above, the vectors $\{x_t\}$ are freely sampled in any direction of $\mathbb{R}^n$ by following a Gaussian distribution. In the dynamic system case however, the data vectors $\{x_t\}$ constructed as in (3) are constrained to lie on a manifold. As a result, the data matrix $X$ generated by the dynamic system is less generic. According to conditions of the paper, and (21) in particular, there is a threshold depending on the richness of the data such that exact recovery is guaranteed whenever the number of zero entries in $f$ is larger than this threshold. So, the more generic the data contained in $X$ are, the more outliers can be removed by problem (8). Note that the lack of sufficient genericity can be compensated (to some extent) by implementing the iterative sparsity enhancing technique ($\ell_1$ reweighted algorithm) described in Section 3.1. This leads, for only two iterations, to significantly improved results as represented in Figure 4.
Figure 3: Estimates of probabilities of exact recovery when noise \( \{e_t\} \) is equal to zero. Results of a Monte-Carlo simulation of size 100 with randomly generated linear ARX systems of order \( n_a = n_b = 2 \).

Figure 4: Estimates of probabilities of exact recovery by reweighted \( \ell_1 \) minimization when noise \( \{e_t\} \) is equal to zero. Results of a Monte-Carlo simulation of size 100 with randomly generated linear ARX systems with orders \( n_a = n_b = 2 \).
6 Conclusion

In this paper we have discussed the potential of nonsmooth convex optimization for addressing the problem of robust estimation. Considering in particular the problem of inferring an unknown parameter vector from measurements which are subject to possibly unbounded gross errors, we showed that an exact recovery is possible regardless of the number of gross errors provided certain conditions of genericity hold. The conditions for exact recovery are essentially sufficient conditions. Simulations results reveal that the proposed conditions are conservative. Concerning the identification problem, future work will consider the problem of designing the excitation of a dynamic system so as to achieve such genericity conditions on the regressor matrix.

References


