Sensitivity Analysis of Constrained Linear $L_1$ Regression: 
Perturbations to Response and Predictor Variables

Mingren Shi$^a$, Mark A. Lukas$^b$

$^a$ Mathematics and Computing, University of Southern Queensland, Toowoomba, Qld. 4350, Australia
Fax: 61746315550 E-mail: shi@usq.edu.au

$^b$ Mathematics and Statistics, Murdoch University, Murdoch, W.A. 6150, Australia
Fax: 6189360332 E-mail: M.Lukas@murdoch.edu.au

Abstract. The active set framework of the reduced gradient algorithm is used to develop a direct sensitivity analysis of linear $L_1$ (least absolute deviations) regression with linear equality and inequality constraints on the parameters. We investigate the effect on the $L_1$ regression estimate of a perturbation to the values of the response or predictor variables. For observations with non-zero residuals, we find intervals for the values of the variables for which the estimate is unchanged. For observations with zero residuals, we find the change in the estimate due to a small perturbation to the variable value. The results provide practical diagnostic formulae. They quantify some robustness properties of constrained $L_1$ regression and show that it is stable, but not uniformly stable. The level of sensitivity to perturbations depends on the degree of collinearity in the model and, for predictor variables, also on how close the estimate is to being non-unique. The results are illustrated with numerical simulations on examples including curve fitting and derivative estimation using trigonometric series.

Key words: sensitivity analysis, stability, $L_1$ regression, least absolute deviations, robustness, diagnostics, active set, reduced gradient algorithm.

AMS (MOS) subject classifications: 62J05, 90C31, 65K05

Abbreviated title: Sensitivity of $L_1$ regression

1. Introduction

Consider a linear model $y = X\beta + \epsilon$, where $y$ is an $n \times 1$ response vector corresponding to the $n \times p$ design matrix $X$ of predictor variable values, $\beta$ is an unknown $p \times 1$ vector of parameters and $\epsilon$ is an $n \times 1$ vector of random errors. For our purposes it will be convenient to write the model as a system of linear equations $y_i = x_i^T \beta + \epsilon_i$, $i = 1, \ldots, n$, where $x_i^T$ is the $i$th row of $X$.

In many applications there are additional linear constraints that must be satisfied by some or all of the parameters, for example, positivity. In particular, biometric and econometric models of the form $\tilde{y}_i = \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \tilde{\beta}_3 e_i$, with positive $\tilde{\beta}_1$, $\tilde{\beta}_2$ and $\tilde{\beta}_3$, are of this type after a logarithmic transformation (see p 444 in Judge et al. (1985)). Constrained regression problems also arise naturally in the important areas of parametric (and nonparametric) curve and surface fitting, and in the estimation of solutions of ill-posed and inverse problems from noisy data (see Wahba (1990)). Here, extra information such as the value of the solution at some point leads to a linear equality constraint on the parameters. Extra information such as positivity, monotonicity, concavity or convexity of the solution leads to a set of linear inequality constraints on the parameters (see Wahba (1982) and O’Leary and Rust (1986)). For notational simplicity we will write the linear equality constraints as $x_i^T \beta - y_i = 0$, $i \in \mathcal{E} = \{n + 1, \ldots, n + n_\epsilon\}$, and the inequality constraints as $x_i^T \beta - y_i \leq 0$, $i \in \mathcal{I} = \{n + n_\epsilon + 1, \ldots, n + n_\epsilon + n_\tau\}$.

In the unconstrained case, it is usual to estimate $\beta$ using least squares ($L_2$) regression. For the constrained problem, restricted or constrained least squares regression (as well as other approaches) have been used (see Knautz (2000)). However, as is well known, the least squares method is not robust; it is not optimal for error distributions with long tails and the estimates are overly sensitive to outliers.
Over the past 25 years there has been growing interest in the method of least absolute deviations or \(L_1\) regression as an alternative to least squares regression. For the linear model with linear constraints above, the \(L_1\) regression estimate of \(\beta\) is the solution to the problem (denoted \(LL_1\))

\[
\begin{align*}
\text{minimize} & \quad S(\beta) = \sum_{i=1}^{n} |x_i^T \beta - y_i|, \quad \beta \in \mathbb{R}^p, \\
\text{subject to} & \quad x_i^T \beta - y_i = 0, \quad i \in \mathcal{E} = \{n + 1, \ldots, n + n_\mathcal{E}\}, \\
& \quad x_i^T \beta - y_i \leq 0, \quad i \in \mathcal{I} = \{n + n_\mathcal{E} + 1, \ldots, n + n_\mathcal{E} + n_\mathcal{I}\},
\end{align*}
\]

where we assume that \(n_\mathcal{E} < p < n + n_\mathcal{E} + n_\mathcal{I}\).

An important advantage of \(L_1\) regression over \(L_2\) regression is its robustness. For the unconstrained problem (denoted \(UL_1\), it is well known that the \(L_1\) regression estimator can resist a few large errors in the data \(y\). In fact (see Lemma 3.1 or Bloomfield and Steiger (1983)), the optimal solution (regression estimate) to \(UL_1\) is completely unaffected by a perturbation of \(y\) which maintains the same signs of the residuals. Bloomfield and Steiger (1983, Sec. 2.3) also derived the generalized influence function for \(L_1\) regression, which shows its robustness with respect to \(y_i\), but lack of robustness with respect to \(x_i\). See Huber (1987) for further discussion of the \(L_1\) approach in robust estimation.

In this paper we analyse the sensitivity of the constrained \(L_1\) regression estimate to general perturbations in the data, both in the response \(y\) and in the row vectors \(x_i^T\). The results show that the constrained estimate is also robust with respect to some large perturbations in \(y\). Furthermore they show that if the estimate is unique, then it is stable, in that it depends continuously on the data \(y\) and \(x_i^T\). However, the stability is not uniform but depends on the degree of collinearity in the model and, for \(x_i^T\), also on how close the estimate is to being non-unique. This is consistent with the findings of Ellis (1998) who characterised the singular set in unconstrained \(L_1\) regression.

It is well known that the \(UL_1\) and \(LL_1\) problems can be formulated as linear programming (LP) problems (see e.g. Arthanari and Dodge (1981)). Efficient simplex-type methods have been developed to solve these LP problems including the algorithm of Barrodale and Roberts (1978). Other methods including interior point algorithms have also been developed (see Portnoy and Koenker (1997), Shi and Lukas (2002) and the references there).

With the LP formulation of the \(L_1\) regression problem it is possible to derive sensitivity results using the usual LP sensitivity analysis (i.e. post optimality analysis) based on the simplex method. This approach was used by Narula and Wellington (1985) for \(UL_1\) to find an interval for each data value \(y_i\) for which the regression estimate is unchanged and to determine the effect of deleting an observation. In a similar way, Narula and Wellington (2002) find an interval for each predictor variable value for which the regression estimate is unchanged. Corresponding results for the case of simple linear \(L_1\) regression are derived in Narula et al. (1993). Independently, Dupačová (1992) developed a sensitivity analysis for the LP formulation of \(UL_1\) which considers the effect of perturbing the response \(y\), perturbing the row vectors \(x_i^T\), and adding or deleting an observation.

We extend the results of Narula and Wellington (1985, 2002) and Dupačová (1992) by deriving a direct sensitivity analysis for the general \(LL_1\) problem. The analysis is based on the active set framework of the reduced gradient algorithm (RGA), as developed in Shi and Lukas (2002). Using the corresponding terminology, we will call a model equation \(y_i = x_i^T \beta + \varepsilon\) an active equation at some point \(\beta\) if the residual \(r_i(\beta) = x_i^T \beta - y_i\) equals 0 and inactive otherwise, and similarly for the constraints. Note that the residual here is the opposite of the usual definition. In this framework, many of the results can be easily visualized geometrically in terms of the movement of hyperplanes defined by the equations and constraints. Note that the results themselves are independent of the algorithm used to find the solution. Also the results are quantitative with computable formulae, making them useful in practice for diagnostic purposes. We do not
assume the design matrix has full rank, but we assume throughout that the optimal solution is nondegenerate (see Definition 2.1).

In Section 2 we describe the RGA framework including appropriate optimality conditions. An optimal solution to (1.1) occurs at a special kind of point determined by a basis of the vectors \( x_i \) (possibly augmented). We call such a point and its associated basis matrix a base point and base matrix, respectively (see Definition 2.1). In Section 3 we investigate the effect on the optimal solution of a perturbation to the responses \( y_i, \ i = 1, 2, \ldots, n \). In Lemma 3.1, for each inactive equation at the solution, we find an interval for \( y_i \) for which the solution remains unchanged. The intervals agree with those of Narula and Wellington (1985, 2002) and the sufficiency condition of Dupačová (1992, eq. (13)). We also show by counterexample that these interval conditions are not necessary for the solution to remain unchanged. In Lemma 3.3, for each active equation, we derive the change (error) in the solution due to a sufficiently small perturbation in \( y_i \). The result allows one to decide which of the responses for the active equations has the greatest marginal influence on the \( L_1 \) regression estimate. Theorem 3.1 considers the effect of an arbitrary (sufficiently small) perturbation \( \Delta y \) to the data vector \( y \) and gives a bound on the \( L_1 \) norm of the error in the solution, showing that the solution is stable.

The results in Section 3 are illustrated in Section 5 using numerical simulations with the well known stack loss data from Brownlee (1965, p. 454) and the three problems of curve fitting, and the estimation of first and second derivatives using the parametric form of a trigonometric series. The estimation of derivatives is important in many areas, in particular in the analysis of human growth curves (Gasser et al. (1984) and Eubank (1988)) and pharmacokinetic data (Song et al. (1995)). Numerical differentiation is an ill-posed problem which leads to a collinear (ill-conditioned) design matrix (see Anderssen and Bloomfield (1974)). Such problems were in fact a major motivation for the sensitivity analysis in this paper. The numerical simulations show that the bound derived in Theorem 3.1 is quite useful in assessing the accuracy of the \( L_1 \) estimate under small perturbations in the data.

In Section 4 we consider perturbations to the row vectors \( x_i^T \) in the model equations. In Theorem 4.1 and Corollary 4.1 we consider the inactive and active equations and find an interval for each element of \( x_i^T \) such that the optimal solution to (1.1) is unchanged. In Theorem 4.2, for each active equation, we find the change (error) due to a perturbation in the row vector, showing that the solution is stable. These results again provide useful diagnostic information in deciding which elements have the greatest influence on the solution. In Section 5, numerical simulations with the stack loss model are used to illustrate the results.

Most of the sensitivity results of this paper are contained in the thesis by Shi (1997) where some more examples can be found. Results about perturbations to the constraints and the addition or deletion of observations are derived in Lukas and Shi (2004). This includes the calculation of an \( L_1 \) version of the Cook distance to determine the influence of each observation on the \( L_1 \) regression estimate.

2. Active set framework

First, we give some notation, definitions and theoretical results from Shi and Lukas (2002) (see also Bloomfield and Steiger (1983) and Osborne (1985)). Let the active set at a given point \( \beta \) and its associated sets be

\[
\mathcal{A} = \mathcal{A}(\beta) = \{ i \mid r_i(\beta) \equiv x_i^T \beta - y_i = 0, \ 1 \leq i \leq n + n_z + n_x \}, \tag{2.1A}
\]

\[
\mathcal{A}_S = \mathcal{A} \cap \{1, \ldots, n\}, \quad \mathcal{A}_I = \mathcal{A} \cap \mathcal{I}, \quad \mathcal{A}_S^c = \{1, \ldots, n\} \setminus \mathcal{A}_S. \tag{2.1B}
\]

**Definition 2.1** (1) Denoting \( \text{rank} \{ x_i \}_{i=1}^{n+n_z+n_x} = r \leq p \), we say \( \beta \) is a base point of problem (1.1) if the rank of the active gradient vectors \( \{ x_i \mid i \in \mathcal{A} \} \) is \( r \). (2) A feasible base point \( \beta \) is nondegenerate if the active gradient vectors \( \{ x_i \mid i \in \mathcal{A} \} \) are linearly independent. If every feasible base point of problem (1.1) is nondegenerate, the problem (1.1) is said to be nondegenerate and otherwise it is degenerate.
We will be using the following known optimality and uniqueness conditions.

**Theorem 2.1** Necessary and sufficient conditions for $\mathbf{\beta}$ to be a solution to (1.1) are that $\mathbf{\beta}$ is feasible, i.e.

$$ x_i^T \mathbf{\beta} - y_i = 0, \quad i \in \mathcal{E}, \quad \text{and} \quad x_i^T \mathbf{\beta} - y_i \leq 0, \quad i \in \mathcal{I}, $$

and there exist multipliers $\lambda_i, \ i \in \mathcal{A}$, such that

$$ c = \sum_{i \in \mathcal{A}_S} \sigma_i x_i = \sum_{i \in \mathcal{A}_S} \lambda_i x_i + \sum_{i \in \mathcal{E}} \lambda_i x_i + \sum_{i \in \mathcal{I}} \lambda_i x_i, $$

$$ |\lambda_i| \leq 1, \quad i \in \mathcal{A}_S, \quad \text{and} \quad \lambda_i \leq 0, \quad i \in \mathcal{A}_I, $$

where $\sigma_i = \sigma_i(\mathbf{\beta}) = \text{sign}(r_i(\mathbf{\beta})).$ Furthermore, (1.1) has a unique solution if and only if the inequalities (2.3B) are strict.

This result can be proved indirectly by using an LP formulation of (1.1) and the duality theorem. It can also be proved directly by convex analysis (see Osborne (1985)).

The following is a well known, fundamental result.

**Theorem 2.2** The minimum value of problem (1.1) is achieved and can be found at a (feasible) base point of (1.1). (It is assumed that the feasible region is not empty.)

Theorems 2.1 and 2.2 provide a characterization of the optimal solution to (1.1). In order to analyse the sensitivity, we need to be able to derive the effect on the solution of a change in the problem. This will be done using the active set framework of the reduced gradient algorithm (RGA). Now we briefly describe the steps of the RGA to solve (1.1), as developed in Shi and Lukas (2002) (see also Bloomfield and Steiger (1983) and Osborne (1985,1987)). We only consider the nondegenerate case here.

We require a feasible starting point $\mathbf{\beta}^{(0)}$ at which there are $n$ linearly independent active gradient vectors. To achieve this, we may need to introduce some artificial constraints (e.g. $\beta_i \leq 0, \ i = 1, \ldots, p,$ for unconstrained problems starting at 0). Then $\mathbf{\beta}^{(0)}$ is a base point of the augmented problem, with active set $\mathcal{A} = \mathcal{A}_S \cup \mathcal{E} \cup \mathcal{A}_I \cup \mathcal{A}_0,$ where $\mathcal{A}_0$ denotes a set of indices of artificial constraints. The nonsingular $p \times p$ matrix $A$ with column vectors $x_i, \ i \in \mathcal{A},$ is called the base matrix corresponding to $\mathbf{\beta}^{(0)}$. Note that the use of artificial constraints is only a computational device; the algorithm does not enforce them.

Suppose that $\mathbf{\hat{\beta}}$ is a feasible base point of (1.1), possibly augmented with artificial constraints. Let its active set, the base matrix and its inverse be, respectively,

$$ \mathcal{A} = \{j_1, j_2, \ldots, j_p\}, \quad \mathcal{A} = [x_j, x_{j_2}, \ldots, x_{j_p}], \quad \hat{D} = \hat{A}^{-1} = [\hat{d}_1, \hat{d}_2, \ldots, \hat{d}_p]^T, $$

where $\hat{d}_j$ is the $i$th row vector of $\hat{D}$. As in (2.3A), we define a vector $\hat{c}$ (which can be thought of as a “cost” vector at $\mathbf{\hat{\beta}}$) and express $\hat{c}$ as a linear combination of the columns of $\hat{A}$, where the multipliers $\hat{\lambda}_i$ are given by

$$ \hat{c} = \sum_{i \in \mathcal{A}_S} \sigma_i x_i = \sum_{i = 1}^p \hat{\lambda}_i x_{j_i}, \quad \hat{\lambda}_i = \hat{d}_j^T \hat{c}, \quad i = 1, 2, \ldots, p. $$

Here $\mathcal{A}_S = \{1, 2, \ldots, n\} \setminus \mathcal{A}$ and $\sigma_i = \sigma_i(\mathbf{\hat{\beta}}) = \text{sign}(r_i(\mathbf{\hat{\beta}})).$ Let

$$ \max\{|\hat{\lambda}_i| : \ i \in \mathcal{A}_0\} = |\hat{\lambda}_{\hat{\beta}_0}|, \quad \max\{|\hat{\lambda}_i| : \ i \in \mathcal{A}_S\} = |\hat{\lambda}_{\hat{\beta}_S}|, \quad \text{and} \quad \max\{|\hat{\lambda}_i| : \ i \in \mathcal{A}_I\} = |\hat{\lambda}_{\hat{\beta}_I}|. $$

Then it is easy to see from Theorem 2.1 that the point $\mathbf{\hat{\beta}}$ is optimal provided

$$ |\hat{\lambda}_{\hat{\beta}_0}| = 0, \quad |\hat{\lambda}_{\hat{\beta}_S}| \leq 1 \text{ and } |\hat{\lambda}_{\hat{\beta}_I}| \leq 0. $$

(2.5)
After computing the multipliers, the optimality conditions (2.5) are tested. If they hold, the current feasible base point \( \hat{\beta} \) is optimal and the algorithm is terminated. Otherwise any of the choices

\[
d = \begin{cases} 
-\text{sign}(\hat{\lambda}_{q_0}) \hat{d}_{q_0}, & \text{if } |\hat{\lambda}_{q_0}| \neq 0, \\
-\text{sign}(\hat{\lambda}_{q_1}) \hat{d}_{q_1}, & \text{if } |\hat{\lambda}_{q_1}| > 1, \\
-\hat{d}_{q_2}, & \text{if } \hat{\lambda}_{q_2} > 0,
\end{cases}
\]

(2.6)
is a feasible descent direction at \( \hat{\beta} \). It is descent since, for any direction \( \mathbf{v} \) and sufficiently small \( \alpha > 0 \), we have

\[
S(\hat{\beta} + \alpha \mathbf{v}) = S(\hat{\beta}) + \alpha \left( \mathbf{e}^T \mathbf{v} + \sum_{i \in A_0} |x_i^T \mathbf{v}| \right),
\]

(2.7)
and the term in brackets is negative if \( \mathbf{v} = \mathbf{d} \). It is easy to check that the direction \( \mathbf{d} \) is feasible. Note that any direction is feasible with respect to the artificial constraints. A move in the direction \( \mathbf{d} \), with corresponding index \( j_q, \ q \in \{q_0, q_1, q_2\} \), will leave the boundary plane \( P_{j_q} \) (defined by \( P_{j_q} = \{ \mathbf{b} | \mathbf{x}_j^T \mathbf{b} = y_j = 0 \} \)) but remain on the intersection of the other \( n - 1 \) boundary planes \( P_{j_i}, \ j_i \in \hat{A} \setminus \{j_q\} \). Hence the index \( j_q \) will become inactive. Note that more than one of the three choices of \( \mathbf{d} \) in (2.6) may be possible. But if \( |\hat{\lambda}_{q_0}| \neq 0 \), we choose \( q = q_0 \) since we want to delete the artificial constraints.

Now we compute the search step length. Let the shortest step to the boundary planes \( P_j \) associated with \textit{inactive equations} (i.e. non-zero residuals) and with \textit{inactive inequality constraints} be, respectively,

\[
\delta' = \delta'(\mathbf{d}) = \min \{ \delta_j \equiv -r_j(\hat{\beta})/x_j^T \mathbf{d} | \delta_j > 0, \ \ n \geq j \in \hat{A}_S, \ x_j^T \mathbf{d} \neq 0 \},
\]

(2.8A)

\[
\alpha' = \alpha'(\mathbf{d}) = \min \{ \alpha_j \equiv -r_j(\hat{\beta})/x_j^T \mathbf{d} | \alpha_j > 0, \ \ n \leq j \in I \setminus \hat{A}_I, \ x_j^T \mathbf{d} > 0 \},
\]

(2.8B)

where \( r_j \) is defined in (2.1A). If \( \alpha' < \delta' \), then we choose the step length \( \alpha = \alpha' \). Otherwise, we choose \( \alpha = \delta' \).

Let \( \bar{\beta} = \hat{\beta} + \alpha \mathbf{d} \) be the new base point, and let \( m \) be the index of the new active equation or constraint. In order to compute the new multipliers, the inverse of the base matrix \( \bar{\mathbf{A}} \) at \( \bar{\beta} \) is required. From above, \( \bar{\mathbf{A}} \) is obtained by replacing the \( q \)th column vector of \( \hat{\mathbf{A}} \) by the vector \( \mathbf{x}_m \), where \( q \) is one of \( q_i, \ i = 0, 1, 2 \), according to the choice of \( \mathbf{d} \) in (2.6). So \( \bar{\mathbf{A}} \) can be represented as \( \bar{\mathbf{A}} = \hat{\mathbf{A}} + (\mathbf{x}_m - \mathbf{x}_{j_q}) \mathbf{e}_q^T \), where \( \mathbf{e}_q \) is the \( q \)th coordinate vector. Using the Sherman-Morrison formula (see McCormick (1983)) and letting the \( i \)th row of \( \bar{\mathbf{A}}^{-1} \) and \( \bar{\mathbf{A}}^{-1} \) be \( \bar{\mathbf{d}}_i \) and \( \bar{\mathbf{d}}_i \), respectively, we have

\[
\bar{\mathbf{d}}_i = \begin{cases} 
\frac{1}{\mathbf{x}_m^T \mathbf{d}} \mathbf{d} = \frac{1}{\mathbf{x}_m^T \mathbf{d}_q} \hat{\mathbf{d}}_q, & i = q, \\
\mathbf{d}_i - (\mathbf{x}_m^T \mathbf{d}_i) \frac{1}{\mathbf{x}_m^T \mathbf{d}} \mathbf{d} = \mathbf{d}_i - (\mathbf{x}_m^T \hat{\mathbf{d}}_i) \bar{\mathbf{d}}_q, & i \neq q.
\end{cases}
\]

(2.9)

This completes the iteration of the algorithm. Because it is a descent method and there are only finitely many base points, the algorithm terminates at an optimal base point.

Note that in order to use the sensitivity results in the following sections, it is not necessary to compute the original optimal solution using the RGA above. Suppose we are given a nondegenerate solution \( \beta^* \) to (1.1), computed by any algorithm. From Theorem 2.2 we know that if the solution is unique, then \( \bar{\beta} = \beta^* \) is a base point, and otherwise there is an optimal base point \( \bar{\beta} \) with the same degree of fit as \( \beta^* \) (i.e. \( S(\bar{\beta}) = S(\beta^*) \)). We find the active equations and constraints and if there are less than \( n \), we augment them with the required number of artificial constraints (taken from those of form \( \beta_i \leq \beta_i^* \) with gradients the standard unit vectors) to make a total of \( n \) active equations and constraints. This defines the active set \( \bar{\mathbf{A}} = \bar{\mathbf{A}}_s \cup \mathcal{E} \cup \bar{\mathbf{A}}_l \cup \bar{\mathbf{A}}_0 = \{j_1, j_2, \ldots, j_p\} \) and base matrix \( \bar{\mathbf{A}} = [\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \ldots, \mathbf{x}_{j_p}] \) and its inverse \( \bar{\mathbf{D}} = \bar{\mathbf{A}}^{-1} = [\bar{\mathbf{d}}_1, \bar{\mathbf{d}}_2, \ldots, \bar{\mathbf{d}}_p]^T \). The optimal base point \( \bar{\beta} \) is defined from the active equations and constraints by \( \bar{\mathbf{A}}^T \bar{\beta} = [y_{j_1}, y_{j_2}, \ldots, y_{j_p}]^T \). In practice this may give a more accurate optimal solution than the
original computed solution. From above, a move from $\mathbf{\bar{\beta}}$ in a direction $\pm \mathbf{d}_i$, $i=1, \ldots, p$, is a move towards a neighbouring base point, and since $\mathbf{\bar{\beta}}$ is optimal the direction is not descent.

3. Perturbations to response variable values

In what follows we suppose that $\mathbf{\bar{\beta}} = (\bar{\beta}_1, \bar{\beta}_2, \ldots, \bar{\beta}_p)^T$ is an optimal base point for the problem (1.1). Let the active set, base matrix and the inverse at $\mathbf{\bar{\beta}}$ be, respectively,

$$\mathcal{A} = \mathcal{A}_S \cup \mathcal{E} \cup \mathcal{A}_I \cup \mathcal{A}_0 = \{ j_1, j_2, \ldots, j_p \},$$

(3.1A)

$$\mathcal{A} = [x_{j_1}, x_{j_2}, \ldots, x_{j_p}] \quad \text{and} \quad D = A^{-1} = [d_1, d_2, \ldots, d_p]^T.$$  

(3.1B)

First we consider a perturbation to a single response. Assume that $y_{l}$ is changed to $y_{l} + \Delta y_{l}$. In the following three lemmas, we will analyse the relation between the optimal points of the original problem (1.1) and the perturbed problem (denoted by (1.1)$^{[l]}$) in three different cases which cover all the possibilities. Note that the residuals at $\mathbf{\bar{\beta}}$ for (1.1)$^{[l]}$ are the same as those for (1.1) except for the $l$th residual which is

$$r_l^{[l]}(\mathbf{\bar{\beta}}) = x^T_l \mathbf{\bar{\beta}} - (y_l + \Delta y_l) = r_l(\mathbf{\bar{\beta}}) - \Delta y_l.$$  

(3.2)

Lemma 3.1 Suppose the $l$th equation is not active at $\mathbf{\bar{\beta}}$, i.e. $l \not\in \mathcal{A}$. If $\Delta y_l$ satisfies

$$\Delta y_l \in \begin{cases} (\infty, r_l(\mathbf{\bar{\beta}})) & \text{if} \quad r_l(\mathbf{\bar{\beta}}) > 0, \\ (r_l(\mathbf{\bar{\beta}}), \infty) & \text{if} \quad r_l(\mathbf{\bar{\beta}}) < 0, \end{cases}$$

(3.3)

or equivalently sign$(r_l(\mathbf{\bar{\beta}})) \Delta y_l \leq |r_l(\mathbf{\bar{\beta}})|$, then $\mathbf{\bar{\beta}}$ remains an optimal point for (1.1)$^{[l]}$, with new optimal value

$$S^{[l]}(\mathbf{\bar{\beta}}) = S(\mathbf{\bar{\beta}}) - \text{sign}(r_l(\mathbf{\bar{\beta}})) \Delta y_l.$$  

(3.4)

Proof It is easy to see from (3.2) that if (3.3) holds, then $r_l^{[l]}(\mathbf{\bar{\beta}}) r_l(\mathbf{\bar{\beta}}) \geq 0$. (In fact the two are equivalent since $\text{sign}(r_l(\mathbf{\bar{\beta}})) \Delta y_l > |r_l(\mathbf{\bar{\beta}})|$ implies $r_l(\mathbf{\bar{\beta}}) \Delta y_l > 0$ and $|\Delta y_l| > |r_l(\mathbf{\bar{\beta}})|$, so from (3.2) we have $r_l^{[l]}(\mathbf{\bar{\beta}}) r_l(\mathbf{\bar{\beta}}) < 0$.) Since $r_l^{[l]}(\mathbf{\bar{\beta}})$ is either 0 or has the same sign as $r_l(\mathbf{\bar{\beta}})$, then the same (necessary) optimality conditions (see Theorem 2.1 and (2.5)) for (1.1) are also valid for (1.1)$^{[l]}$ (sufficient condition), so $\mathbf{\bar{\beta}}$ is also a solution to the new problem. Its optimal value will be

$$S^{[l]}(\mathbf{\bar{\beta}}) = \sum_{i \neq l} |x^T_l \mathbf{\bar{\beta}} - y_l| + |r_l(\mathbf{\bar{\beta}}) - \Delta y_l|$$

$$= \sum_{i \neq l} |x^T_l \mathbf{\bar{\beta}} - y_l| + |r_l(\mathbf{\bar{\beta}})| - \text{sign}(r_l(\mathbf{\bar{\beta}})) \Delta y_l = S(\mathbf{\bar{\beta}}) - \text{sign}(r_l(\mathbf{\bar{\beta}})) \Delta y_l.$$  

$$\blacksquare$$

Remark 3.1 (1) The intervals in (3.3) are the same as those defined by Narula and Wellington (1985, 2002). They are also consistent with the sufficiency condition in Dupačová (1992, eq. (13)), but the necessary condition in Dupačová (1992, eq. (15)) is incorrect; see part (3) below and Lemma 3.2.

(2) Geometrically, a perturbation of $y_l$ causes a parallel (i.e. affine) move of the plane $P_l = \{ \mathbf{\beta} \mid x^T_l \mathbf{\beta} = y_l \}$ along its normal direction $\pm x_l$. If $\Delta y_l$ is within the interval (3.3), $\mathbf{\bar{\beta}}$ remains on the same side of the shifted plane $P_l^{[l]} = \{ \mathbf{\beta} \mid x^T_l \mathbf{\beta} = y_l + \Delta y_l \}$ (when $\Delta y_l = r_l(\mathbf{\bar{\beta}})$ the plane $P_l^{[l]}$ passes through $\mathbf{\bar{\beta}}$). Hence the directions from $\mathbf{\bar{\beta}}$ to the new base points lying on $P_l^{[l]}$ are the same as the directions to the original ones lying on $P_l$. Therefore no direction from $\mathbf{\bar{\beta}}$ becomes a new descent direction, implying that $\mathbf{\bar{\beta}}$ remains an optimal base point.

(3) The inequality (3.3) is not a necessary condition for $\mathbf{\bar{\beta}}$ to remain optimal; see Example 3.1. It can be seen from Example 3.1 that for an unconstrained problem with $n = p + 1$ and $l \not\in \mathcal{A}$, the optimal point $\mathbf{\bar{\beta}}$ is not changed provided the scale $||x_i||$ of the vector is small enough. This follows because, in this case, the
original and perturbed “cost” vector differ only in sign and so the absolute values of the multipliers are not changed.

**Example 3.1** \( S(\bar{\beta}) = |2\beta_1 + \beta_2 - 4| + |\beta_1 + \beta_2 - 2| + |(\beta_1 + \beta_2) - y_3|, \ |\bar{\varepsilon}| < 1/3. \)

It is easy to check that \( \bar{\beta} = (2,0)^T \) is an optimal base point with \( \bar{A} = \{1,2\} \) and \( r_3(\bar{\beta}) = -2\bar{\varepsilon} - y_3. \) The “cost” vector is \( \bar{c} = \sigma_3\varepsilon(-1,1)^T, \) and the multipliers are \( \lambda_1 = \frac{\bar{d}}{\bar{c}}c = (1,-1)\sigma_3\varepsilon(-1,1)^T = -2\sigma_3\varepsilon \) and \( \lambda_2 = \frac{\bar{d}_2}{\bar{c}} = (-1,2)\sigma_3\varepsilon(-1,1)^T = 3\sigma_3\varepsilon. \) (If \( r_3(\bar{\beta}) = 0, \) then set \( \lambda_3 = 0. \)) Now for any \( \Delta y_3, \) the new “cost” vector is \( \bar{c}^{[1]} = \pm\sigma_3\varepsilon(-1,1)^T \) and the new multipliers have absolute values \( |\bar{\lambda}_1^{[1]}| = 2|\bar{\varepsilon}| < 1 \) and \( |\bar{\lambda}_2^{[1]}| = 3|\bar{\varepsilon}| < 1, \) so \( \bar{\beta} \) remains optimal.

As the next result shows, in a special case it is possible that the solution remains optimal after a perturbation to the response of an active equation.

**Lemma 3.2** Assume that the \( h \)th equation is active at \( \bar{\beta}, \) let \( l = j_q \in \bar{A} \) and let the multipliers at \( \bar{\beta} \) for (1.1) be \( \bar{\lambda}_i, \ i = 1,2,\ldots, p. \) If \( \text{sign}(\Delta y_l) = \bar{\lambda}_q \) (i.e. \( \bar{\lambda}_q = 1 \) and \( \Delta y_l > 0 \), or \( \bar{\lambda}_q = -1 \) and \( \Delta y_l < 0 \); the solution is not unique in this case), then \( \bar{\beta} \) is also an optimal point for (1.1)[l] with new optimal value

\[
S[l](\bar{\beta}) = S(\bar{\beta}) + |\Delta y_l|. \tag{3.5}
\]

**Proof** From (3.2), \( r[l](\bar{\beta}) = -\Delta y_l \neq 0 \) in this case so the index \( j_q \notin \bar{A}[l], \) the active set at \( \bar{\beta} \) for (1.1)[l]. However, it can be regarded as the index of the artificial constraint \( x^T\beta - y_l \leq 0. \) With this point of view, \( \bar{\beta} \) is still a base point and the inverse of the base matrix is the same as shown in (3.1B) for (1.1)[l]. For the perturbed problem, the corresponding “cost” vector at \( \bar{\beta} \) (see (2.4)) will be changed to

\[
c^{[l]} = \bar{c} + \text{sign}(r[l])x_{j_q} = \bar{c} - \text{sign}(\Delta y_l)x_l.
\]

Then the new multipliers will be \( \lambda[l]_i = (Dc[l])_i, \) and from \( \bar{d}_l^T x_{j_q} = \delta_{lq}, \) we have

\[
\lambda[l]_i = \bar{\lambda}_i, \quad \text{if} \quad i \neq q, \quad \text{and} \quad \lambda[l]_q = \bar{\lambda}_q - \text{sign}(\Delta y_l). \tag{3.6}
\]

If \( \lambda_q = 1 \) and \( \Delta y_l > 0, \) then from (3.6) we have \( \lambda[l]_q = 0. \) Furthermore, the multipliers \( \lambda[l]_i, \ i = 1,2,\ldots, p, \) satisfy (2.5) and so \( \bar{\beta} \) is also an optimal point for (1.1)[l]. It is easy to show that (3.5) is valid. If \( \lambda_q = -1 \) and \( \Delta y_l < 0, \) then the same results hold with the same arguments.

We illustrate Lemma 3.2 with the following simple examples.

**Example 3.2** (a) \( S(\bar{\beta}) = |2\beta_1 + \beta_2 - 4| + |\beta_1 + \beta_2 - 2| + |(\beta_1 + \beta_2) - y_3|, \)

It is easy to check that \( \bar{\beta} = (2,0)^T \) is an optimal base point with \( \bar{A} = \{1,4\} \) and multipliers \( \bar{\lambda}_1, \bar{\lambda}_2 = (0,-1). \) Therefore, by Lemma 3.2, if \( \Delta y_4 < 0, \) then \( \bar{\beta} = (2,0)^T \) remains optimal with new optimal value

\[
S[l](\bar{\beta}) = S(\bar{\beta}) + |\Delta y_4| = 4 - \Delta y_4. \quad \text{(Note (1,2)^T is another optimal base point with similar behaviour.)}
\]

(b) \( S(\beta) = |\beta - 2| + |\beta - 2|. \)

It is easy to check that \( \bar{\beta} = 2 \) is an optimal base point with \( \bar{A} = \{1\} \) and multiplier \( \lambda = 1. \) So, by Lemma 3.2, if \( \Delta y_1 > 0, \) then \( \bar{\beta} = 2 \) remains optimal with new optimal value

\[
S[l](\beta) = S(\beta) + |\Delta y_1| = 4 + \Delta y_1. \quad \text{In this 1-dimensional example, the result is easily confirmed by graphing} \ S[l](\beta) = |\beta - (2 + \Delta y_1)| + |\beta - 2|.
\]

The next result shows that, usually, a perturbation to the response of an active equation results in a change in the optimal solution.

**Lemma 3.3** Assume that \( l = j_q \in \bar{A} \) and \( \text{sign}(\Delta y_l) \neq \bar{\lambda}_q. \) Then we have:

1. The signs of \( \Delta y_l \) and \( \lambda[l]_q \) in (3.6) are opposite.
2. If the perturbation satisfies

\[
|\Delta y_l| < \min\{d'(d), \alpha'(d)\}, \quad d = -\text{sign}(\lambda[l]_q)d_l = \text{sign}(\Delta y_l)d_l, \tag{3.7}
\]

7
where $\delta'$ and $\alpha'$ are the shortest steps to the boundary planes associated with the inactive equations and inactive constraints, respectively (see (2.8)), then an optimal point and its optimal value for (1.1)[il] are, respectively,
\[ \tilde{\beta} = \beta + |\Delta y| \mathbf{d} = \beta + \Delta y \mathbf{d}_q \quad \text{and} \quad S^{[l]}(\tilde{\beta}) = S(\beta) + \Delta y \tilde{\lambda}_q. \]
(3.8)

Here $\mathbf{d}_q$ is the $q$th row of the inverse matrix $\tilde{D}$. Also the base matrix does not change.

**Proof** (1) This follows easily from (3.6).

(2) In this case, from (3.6), the multiplier corresponding to the artificial constraint $x_i^T \beta - y_i \leq 0$ is $\lambda_i^{[l]} \neq 0$, so $\tilde{\beta}$ is not a solution to (1.1)[il]. According to (2.6), $\mathbf{d}$ in (3.7) is a descent direction. Under the condition on $\Delta y$, the search from $\tilde{\beta}$ along $\mathbf{d}$ will leave the boundary plane corresponding to the artificial constraint and reach the boundary plane corresponding to the $l$th perturbed equation $x_i^T \beta - (y_i + \Delta y_i) = 0$ without crossing any inactive plane. It follows from (2.8) and (3.7) that the step length is $\delta^{[l]} = -r_i^{[l]}(\tilde{\beta})/\|x_i^T \mathbf{d}\| = |\Delta y_i|$. Furthermore, $\tilde{\beta}$ defined by (3.8) is a feasible base point for (1.1)[il] and $l \in A^{[l]}(\tilde{\beta})$ since $x_i^T \tilde{\beta} - (y_i + \Delta y_i) = 0$. Hence $A^{[l]}(\tilde{\beta}) = \tilde{A}$. Because the artificial constraint and perturbed residual have the same gradient $x_i$, the “cost” vector, the base matrix with its inverse and multipliers do not change. Therefore, from Theorem 2.1, $\tilde{\beta}$ satisfies the sufficient conditions for optimality and so it is a solution to (1.1)[il]. Moreover, it follows from (3.8), (3.7), and (2.7) that
\[
S^{[l]}(\tilde{\beta}) = \sum_{j=1}^n |x_j^T \beta - y_j| = S(\tilde{\beta}) - |x_j^T \tilde{\beta} - y_j|
\]
\[
= S(\beta) + |\Delta y| \text{sign}(\Delta y_i) \mathbf{d}_q - |\Delta y_i|
\]
\[
= S(\beta) + |\Delta y_i| (\text{sign}(\Delta y_i) \tilde{\lambda}_q + 1) - |\Delta y_i| = S(\beta) + \Delta y_i \tilde{\lambda}_q,
\]
which completes the proof. ■

**Remark 3.2** (1) Geometrically, a parallel move of an active plane will make $\tilde{\beta}$ a non-base point and produce a new base point. The condition (3.7) ensures that the new base point is strictly on the same side of all inactive constraint planes (therefore it is still feasible) and all inactive equation planes (therefore still all relative directions are not descent). So the new base point is optimal.

(2) If the base matrix at $\tilde{\beta}$ is ill-conditioned, then $\|\mathbf{d}_q\|$ could be large, which, from (3.8), would make the norm change $\|\tilde{\beta} - \beta\|$ much larger, relatively, than $|\Delta y_i|$. Also $\delta'$ in (2.8A) or $\alpha'$ in (2.8B) may be small, since $x_i^T \mathbf{d}_q$ appears in the denominator, so the upper bound for $|\Delta y_i|$ in (3.7) may be small. This shows there may be substantial sensitivity in this case.

(3) When $|\Delta y_i| = \alpha'(\mathbf{d})$, the point $\tilde{\beta}$ is degenerate but $\tilde{\beta}$ is still a solution to (1.1)[il] (see Shi and Lukas (2002)). If $|\Delta y_i| = r_i(\mathbf{d}) = \delta^{[l]}$, then, since the sign of the $j$-th residual is changed to zero, the “cost” vector is changed. Therefore, we need to check if the degenerate point $\tilde{\beta}$ is a solution to the new problem or not (using say the derived LL1 problem described in Shi and Lukas (2002)).

Now assume that the response vector $\mathbf{y} = (y_1, \ldots, y_n)^T$ is changed to $\mathbf{y} + \Delta \mathbf{y}$. The above lemmas can be used sequentially to obtain the following theorem bounding the $L_1$ norm error due to a small general perturbation vector $\Delta \mathbf{y}$.

**Theorem 3.1** Let $\delta(\tilde{\beta}, \mathbf{d}_k)$ be the shortest step from $\tilde{\beta}$ to the inactive boundary planes (for equations or constraints) along the directions $\pm \mathbf{d}_k$ and let $\delta^*(\tilde{\beta})$ be the shortest of these steps over all $k \in \{1, \ldots, p\}$, i.e.

\[
\delta(\tilde{\beta}, \mathbf{d}_k) = \min_j |r_j(\tilde{\beta})|/|x_j^T \mathbf{d}_k| \quad \text{for} \quad j \in C_k, \quad \delta^*(\tilde{\beta}) = \min_k \delta(\tilde{\beta}, \mathbf{d}_k),
\]
(3.9A)

\[
C_k = \{ j | j \in \tilde{A}_\delta \cup I \setminus \tilde{A}_r, x_j^T \mathbf{d}_k \neq 0 \}.
\]

Given any $0 < \epsilon \leq 1$, if
\[
|\Delta \mathbf{y}|_{\infty} \leq \eta \epsilon, \quad \eta = \min \{ \delta^*(\tilde{\beta})/(\rho K) \}, 
\]
\[
\max_{1 \leq r_1 \leq p} \max_{j \in C_k} \{|x_j^T \mathbf{d}_j|/|x_j^T \mathbf{d}_k|\}.
\]
(3.9B)

\[
K = \max_{1 \leq r_1 \leq p} \max_{j \in C_k} \{|x_j^T \mathbf{d}_j|/|x_j^T \mathbf{d}_k|\}.
\]
(3.9C)
then there is an optimal solution $\overline{\beta}$ to the perturbed problem with optimal value $\mathcal{S}(\overline{\beta})$ that satisfy, respectively,

$$
|\overline{\beta} - \overline{\beta}| \leq \eta \epsilon \|\overline{D}\|_\infty \quad \text{and} \quad |\mathcal{S}(\overline{\beta}) - \mathcal{S}(\overline{\beta})| \leq \eta \epsilon n.
$$

(3.10)

**Proof** First we show that Lemmas 3.1, 3.2 and 3.3 can be applied sequentially. We perturb $y_k$ by $\Delta y_k$, one by one in an order such that the first group of perturbations are of the type in Lemma 3.1 (case I), the second group are of the type in Lemma 3.2 (case II) and the last group are of the type in Lemma 3.3 (case III). If Lemmas 3.1 and 3.2 apply, then the solution is not changed. (i) For a perturbation in case I, the previous perturbations do not affect the current residual $r_l$ (see (3.2)). Hence Lemma 3.1 applies since

$$
|\Delta y_k| \leq \min\{|r_j(\overline{\beta})| : j \in \mathcal{A}_k\} \quad \text{from (3.9)}.
$$

(ii) In case II it is easy to see that the previous perturbations (from case I or case II) do not affect the condition on $\Delta y_k$ in Lemma 3.2, so the conclusion follows. (iii) Suppose the successive perturbations in case III are $\Delta y_l$, $l = l_1, l_2, \ldots, l_t$, $(1 \leq t \leq p)$. We will prove by induction that Lemma 3.3 applies for each $l_r$ to define a sequence of corresponding solutions $\overline{\beta}^{[l_r]}$, $r = 1, 2, \ldots, t$. In fact, we will show that for $r = 1, 2, \ldots, t - 1$ and $l_r = j_{y_k}$,

$$
\overline{\beta}^{[l_r]} = \overline{\beta}^{[l_r-1]} + \Delta y_{l_r} \overline{d}_{l_r}, \quad \Delta y_{l_r} = r_j(\overline{\beta}^{[l_r-1]}) + \Delta y_{l_r} \overline{x}_j^T \overline{d}_{l_r},
$$

(3.11A)

$$
\delta(\overline{\beta}^{[l_r]}, \overline{d}_k) \geq \delta(\overline{\beta}^{[l_r-1]}, \overline{d}_k) - \max\{|\Delta y_{l_r}| |x_j^T \overline{d}_{l_r}| / |x_j^T \overline{d}_k|\} \geq \delta(\overline{\beta}^{[l_r-1]}, \overline{d}_k) - K|\Delta y_{l_r}|,
$$

(3.11B)

$$
\delta^*(\overline{\beta}) / p \leq \delta(\overline{\beta}^{[l_r]}, (p - r)) \leq \delta^*(\overline{\beta}^{[l_r]}) / (p - r),
$$

(3.11D)

and (3.11A) and (3.11B) also hold for $r = t$. For $r = 1$, since $K \geq 1$ it is obvious from (3.9B) that (3.11A) is valid, so $\Delta y_{l_1}$ satisfies (3.7). Then Lemma 3.3 applies to define $\overline{\beta}^{[l_1]}$ as in (3.11B) ($q_1 = q$), and (3.11B) implies (3.11C). Minimizing both sides of (3.11C) over $k$ and using (3.9), we obtain $\delta^*(\overline{\beta}^{[l_1]}) \geq ((p - 1)/p)\delta^*(\overline{\beta})$, and so (3.11D) holds for $r = 1$. Now suppose that (3.11A)-(3.11D) are valid for $r = s - 1$. For $r = s \leq t - 1$, from (3.9), (3.11D) and $s \leq t \leq p$, we have $p - (s - 1) \geq p - (t - 1) \geq 1$ and

$$
|\Delta y_{l_s}| \leq \delta^*(\overline{\beta}) / p \leq \delta^*(\overline{\beta}^{[l_{s-1}]} / (p - s) \leq \delta^*(\overline{\beta}^{[l_{s-1}]}).
$$

Hence, $\Delta y_{l_s}$ satisfies (3.7) so Lemma 3.3 applies to define $\overline{\beta}^{[l_s]}$ as in (3.11B) (note the base matrix is not changed) which implies (3.11C) for $r = s$. Now, minimizing both sides of (3.11C) over $k$ and using (3.9) and the inductive hypothesis for (3.11D) ($r = s - 1$), we obtain

$$
\delta^*(\overline{\beta}^{[l_s]}) \geq \delta^*(\overline{\beta}^{[l_{s-1}]}) - K|\Delta y_{l_s}| \geq ((p - s + 1)/p)\delta^*(\overline{\beta}) - (1/p)\delta^*(\overline{\beta}) = ((p - s - 1)/p)\delta^*(\overline{\beta}).
$$

So (3.11D) holds for $r = s$. Finally, from (3.9) and (3.11D) with $r = t - 1$, we obtain (3.11A) for $r = t$, and Lemma 3.3 defines $\overline{\beta}^{[l_t]}$ as in (3.11B).

Now we derive (3.10). In the cases I and II, the solution is not changed. In case III, $\overline{\beta} \equiv \overline{\beta}^{[l_t]}$ is a solution to the whole perturbed problem. From (3.8) and since $t \leq p$, we have

$$
|\overline{\beta} - \overline{\beta}| \leq \sum_{l=1}^t \|\overline{\beta}^{[l-1]} - \overline{\beta}^{[l-2]}\|_1 + \|\overline{\beta}^{[l-1]} - \overline{\beta}\|_1 \leq \sum_{l=1}^p |\Delta y_{l_r}| \cdot |\overline{D}|_1 \leq \eta \epsilon \|\overline{D}\|_\infty.
$$

From (3.4), (3.5) and (3.8), for the perturbation $\Delta y_k$ of $y_k$, the optimal objective function value changes by no more than $|\Delta y_k|$ in magnitude. Since there are $n$ such perturbations and $|\Delta y_k| \leq \eta \epsilon$, the final optimal value $\mathcal{S}(\overline{\beta})$ satisfies (3.10). \qed

This result shows that the $L_1$ regression solution is stable with respect to arbitrary small perturbations in the response vector. From the $L_1$ error bound in (3.10), the degree of sensitivity is determined by $\|\overline{D}\|_\infty$. If the base matrix is ill-conditioned, then $\|\overline{D}\|_\infty$ will be large, indicating a much more sensitive solution. This will be demonstrated by numerical simulations in Section 5.

4. Perturbations to the values of predictor variables

Now we examine the effect on the $L_1$ regression solution of a perturbation to a row of the design matrix $X$. Assume that $\overline{\beta}$ is the unique solution to (1.1) and the $k$th vector $x_k$ is perturbed to $x_k + \Delta x_k$, $1 \leq k \leq n$. The perturbed problem is denoted by (1.1)$^{<k>}$\(\textcircled{\scriptsize{<}}\). First we consider the case where the corresponding equation is inactive, i.e., the residual is non-zero.
Theorem 4.1 If $k \not\in \mathcal{A}_S$, so $r_k(\tilde{\beta}) \neq 0$, and the perturbation $\Delta x_k$ satisfies
\[
\|\Delta x_k\| < \eta \equiv \min \{ \frac{\|r_k(\tilde{\beta})\|}{\|\tilde{\beta}\|}, \min_{j_i \in \mathcal{A}_S} \frac{1 - |\tilde{\lambda}_i|}{\|\tilde{d}_i\|}, \min_{j_i \in \mathcal{A}_T} \frac{|\tilde{\lambda}_i|}{\|\tilde{d}_i\|} \},
\]
then $\tilde{\beta}$ remains the unique optimal base point for (1.1)$^{<k>}$, but with the new optimal value
\[
S^{<k>}(\tilde{\beta}) = S(\tilde{\beta}) - |r_k(\tilde{\beta})| + |r^{<k>}_k(\tilde{\beta})| = S(\tilde{\beta}) + \text{sign}(r_k(\tilde{\beta})) \Delta x_k^T \tilde{\beta}.
\]
\[\text{Note that } \eta > 0 \text{ since } r_k(\tilde{\beta}) \neq 0 \text{ and the multipliers from the uniqueness conditions for the multipliers. (The case with } \tilde{\beta} = 0, \text{ which implies } \Delta x_k^T \tilde{\beta} = 0 \text{ for arbitrary } \|\Delta x_k\|, \text{ is trivial and the undefined fraction in (4.1) can be replaced by } +\infty.\]

**Proof** First, from (4.1) we have $\|\Delta x_k^T \tilde{\beta}\| \leq \|\Delta x_k\| \cdot \|\tilde{\beta}\| < |r_k(\tilde{\beta})|$. Then $r^{<k>}_k(\tilde{\beta})$ has the same sign as $r_k$ from $r^{<k>}_k(\tilde{\beta}) = r_k(\tilde{\beta}) + \Delta x_k^T \tilde{\beta}$. (or the same value if $\Delta x_k^T \tilde{\beta} = 0$.) Therefore, the active set, the base matrix and its inverse are not changed, and the new “cost” vector and multipliers are $c^{<k>}_k = c + \sigma_k \Delta x_k$ and
\[
\lambda^{<k>}_k = d_k^T c^{<k>} = d_k^T c + \sigma_k d_k^T \Delta x_k = \tilde{\lambda}_i + \sigma_k d_k^T \Delta x_k, \quad (\sigma_k = \text{sign}(r_k)) \tag{4.3}
\]
respectively. If $j_i \in \mathcal{A}_S$, it follows from (4.3) and (4.1) that $|\lambda^{<k>}_k| \leq |\tilde{\lambda}_i| + |\tilde{d}_i| \cdot \|\Delta x_k\| < 1$. Similarly, if $j_i \in \mathcal{A}_T$, we have $\lambda^{<k>}_k \leq \lambda_i + |\tilde{d}_i| \cdot \|\Delta x_k\| < \lambda_i + -\tilde{\lambda}_i = 0$ (because $\tilde{\lambda}_i < 0$, see (2.3B)). Hence $\tilde{\beta}$ is also the unique solution to (1.1)$^{<k>}$.

It is easy to show that (4.2) is valid. \(\blacksquare\)

**Remark 4.1** If the base matrix at $\tilde{\beta}$ is ill-conditioned, then the upper bound for $\|\Delta x_k\|$ in (4.1) may be small since max $|\tilde{d}_i|$ may be large. Similarly, if the solution $\tilde{\beta}$ is close to being non-unique, with $|\tilde{\lambda}_i| \approx 1$ for some $j_i \in \mathcal{A}_S$ or $\tilde{\lambda}_i \approx 0$ for some $j_i \in \mathcal{A}_T$, then the upper bound for $\|\Delta x_k\|$ may be small.

For a perturbation $\Delta x_{kl}$ to a single element $x_{kl}$ in the $k$th row vector, similar arguments as for Theorem 4.1 can be applied to give the following result.

**Corollary 4.1** Suppose $k \not\in \mathcal{A}_S$ and let $R_{il} = \text{sign}(r_k(\tilde{\beta})) \tilde{D}_{il}, i = 1, \ldots, p, l = 1, \ldots, p$. Define
\[
\Delta_{kl}^{\min} = \max \left\{ \left\{ \frac{|r_k(\tilde{\beta})|}{\|\tilde{\beta}\|} : \tilde{\beta} r_k(\tilde{\beta}) > 0 \right\}, \max_{j_i \in \mathcal{A}_S} \frac{-\text{sign}(R_{il}) - \tilde{\lambda}_i}{R_{il}}, \max_{j_i \in \mathcal{A}_T} \frac{-\tilde{\lambda}_i}{R_{il}} \right\}
\]
and
\[
\Delta_{kl}^{\max} = \min \left\{ \left\{ \frac{|r_k(\tilde{\beta})|}{\|\tilde{\beta}\|} : \tilde{\beta} r_k(\tilde{\beta}) < 0 \right\}, \min_{j_i \in \mathcal{A}_S} \frac{\text{sign}(R_{il}) - \tilde{\lambda}_i}{R_{il}}, \min_{j_i \in \mathcal{A}_T} \frac{-\tilde{\lambda}_i}{R_{il}} \right\},
\]
where we replace an undefined term in $\Delta_{kl}^{\min}$ by $-\infty$ (e.g. if $R_{il} = 0$) and an undefined term in $\Delta_{kl}^{\max}$ by $+\infty$. If the perturbation $\Delta x_{kl}$ satisfies $\Delta_{kl}^{\min} < \Delta x_{kl} < \Delta_{kl}^{\max}$, then $\tilde{\beta}$ remains the unique optimal base point for (1.1)$^{<k>}$, with the new optimal value $S^{<k>}(\tilde{\beta}) = S(\tilde{\beta}) + \text{sign}(r_k(\tilde{\beta})) \Delta x_{kl} \tilde{\beta}$.

**Remark 4.2** Note that for an unconstrained problem, the bounds $\Delta_{kl}^{\min}$ and $\Delta_{kl}^{\max}$ on each perturbation $\Delta x_{kl}$ will be finite. This follows since $\tilde{D}$ is nonsingular so for at least one $i$ we have $R_{il} \neq 0$ and, regardless of the sign of this element, $(-\text{sign}(R_{il}) - \tilde{\lambda}_i)/R_{il} < 0$ and $(\text{sign}(R_{il}) - \tilde{\lambda}_i)/R_{il} > 0$ because $-1 < \tilde{\lambda}_i < 1$.

When the $k$th column is active, i.e. the residual is 0, $x_k$ is a column of the base matrix and we need to ensure that the perturbed base matrix is nonsingular. It is well known that the determinant of a matrix is a continuous function with respect to its column vectors. Hence, for a sufficiently small change to a column vector, the nonsingularity of the matrix does not change. In fact we have the following result.

**Proposition 4.1** Let the matrix $\bar{A}$ be formed by replacing the $q$th column $x_k$, $k = j_q$, of the nonsingular matrix $A$ by $x_k + \Delta x_k$. Then $\bar{A}$ is still nonsingular provided
\[
\|\Delta x_k\| < 1/\|\tilde{d}_i\|.
\]
Moreover, the $i$th row of the inverse $\bar{A}^{-1} = \tilde{A}^{-1}$ is equal to the transpose of
\[
d_i = \begin{cases} \frac{1}{\tilde{d}_i}, & i = q, \\ \tilde{d}_i - (\Delta x_k^2 \tilde{d}_i), & i \neq q. \end{cases} \tag{4.5}
\]
Proof Clearly, $\overline{A}$ can be represented as $\overline{A} = \tilde{A} + \Delta x_k e_i^T$. It follows from (4.4) that

$$|e_i^T \tilde{A}^{-1} \Delta x_k + 1| = |d_i^T \Delta x_k + 1| \geq 1 - ||\tilde{d}_i|| \cdot ||\Delta x_k|| > 0,$$

so $e_i^T \tilde{A}^{-1} \Delta x_k + 1 \neq 0$ and the condition of the Sherman-Morrison formula is satisfied. This implies that $\overline{A}$ is nonsingular and then (4.5) is valid from (2.9) and $x_i^T \tilde{d}_i = \delta_{ii}$. ■

**Theorem 4.2** Assume that $k = j_q \in \tilde{A}_s$, so $r_k(\bar{\beta}) = 0$, and $\Delta x_k^T \bar{\beta} \neq 0$. Let $\gamma_t = \Delta x_k^T \tilde{d}_k$, $1 \leq t \leq p$, and then

$$\Delta x_k = \sum_{j_t \in \tilde{A}} \gamma_t x_{j_t},$$

(4.6)

since the matrix $\tilde{A}$ is nonsingular and $x_i^T \tilde{d}_i = \delta_{ii}$.

(1) If

$$\gamma_q = - (\lambda_q + \theta_k)/\theta_k, \quad \theta_k \equiv \text{sign}(\Delta x_k^T \bar{\beta}),$$

and

$$||\Delta x_k|| < \eta \equiv \min\left\{ \min_{j_t \in \tilde{A}_s \setminus \{k\}} \frac{1 - ||\lambda_t||}{||\tilde{d}_t|| \cdot ||\lambda_t||}, \min_{j_i \in \tilde{A}_r} \frac{||\lambda_i||}{||\tilde{d}_i|| \cdot ||\lambda_i||} \right\},$$

(4.7A)

(4.7B)

then $\bar{\beta}$ remains the unique solution to (1.1)$^{<k>}$ with the optimal value

$$S^{<k>}(\bar{\beta}) = S(\bar{\beta}) + ||\Delta x_k^T \bar{\beta}||.$$  \hspace{1cm} (4.8)

(2) As in (2.8), let $\delta'$ and $\alpha'$ be the shortest steps to the boundary planes corresponding to the inactive equations and inactive constraints, respectively, from $\bar{\beta}$ along the direction

$$d = -\theta_k \bar{d}_q, \quad \theta_k \equiv \text{sign}(\Delta x_k^T \bar{\beta}),$$

(4.9)

If

$$||\Delta x_k|| < \eta \equiv \min\{\eta_0, \eta_1, \eta_2\},$$

(4.10A)

$$\eta_0 \equiv \min\left\{ \frac{1 - ||\lambda_q||}{||\tilde{d}_q|| \cdot ||\lambda_q|| + \delta' \cdot ||\tilde{d}_q||}, \frac{1 - ||\lambda_q|| \cdot ||\tilde{d}_q||}{||\lambda_q|| + \alpha' \cdot ||\tilde{d}_q||} \right\},$$

(4.10B)

$$\eta_1 \equiv \min_{j_t \in \tilde{A}_s \setminus \{k\}} \frac{1 - ||\lambda_t||}{||\lambda_t|| \cdot ||\tilde{d}_t||},$$

(4.10C)

$$\eta_2 \equiv \min_{j_i \in \tilde{A}_r} \frac{||\lambda_i||}{||\lambda_i|| \cdot ||\tilde{d}_i||},$$

(4.10D)

then

$$\bar{\beta} = \bar{\beta} + \alpha^{<k>} \tilde{d}_q, \quad \alpha^{<k>} = \frac{||\Delta x_k^T \bar{\beta}||}{1 + ||\Delta x_k^T \bar{\beta}||}.$$ \hspace{1cm} (4.11)

is the unique solution to (1.1)$^{<k>}$ with the new optimal value

$$S^{<k>}(\bar{\beta}) = S(\bar{\beta}) - \frac{(\Delta x_k^T \bar{\beta}) \lambda_q}{1 + \Delta x_k^T d_q}. \hspace{1cm} (4.12)$$

**Proof** In the two cases above, the new residual and “cost” vector at $\bar{\beta}$ are, respectively,

$$r_k^{<k>} = \Delta x_k^T \bar{\beta} \neq 0 \quad \text{and} \quad c^{<k>} = \bar{c} + \theta_k (x_k + \Delta x_k).$$

(4.13)

With the point of view that $k$ is the index of the artificial constraint $x_k^T \bar{\beta} - y_k \leq 0$, $\bar{\beta}$ and $\tilde{D}$ remain a feasible base point and the inverse of the base matrix for (1.1)$^{<k>}$, respectively. Then the new multipliers at $\bar{\beta}$ are

$$\lambda^{<k>} = \tilde{d}_q^T c^{<k>} = \begin{cases} \lambda_q + \theta_k (1 + \gamma_q), & i = q, \\ \lambda_i + \theta_k \gamma_i, & i \neq q, \end{cases}$$

(4.14)
by (4.13) and (4.6).

(1) If (4.7) holds, then \( \lambda_{\ell}^{<k>} \), which is the multiplier associated with the artificial constraint, vanishes. Furthermore, from the definition of \( \gamma_0 \), we have:

\[
|\lambda_{\ell}^{<k>}| \leq |\bar{\lambda}_i| + ||\bar{d}_i|| \cdot ||\Delta x_k|| < 1, \quad j_i \in \bar{A}_S \setminus \{ k \} = \bar{A}_{S}^{<k>} \setminus \{ k \},
\]

\[
\lambda_{\ell}^{<k>} \leq -|\bar{\lambda}_i| + ||\bar{d}_i|| \cdot ||\Delta x_k|| < 0, \quad j_i \in \bar{A}_I = \bar{A}_{I}^{<k>},
\]

where \( \bar{A}_{I}^{<k>} \) denotes the active set for (1.1)<sup>\(k\rangle</sup> at \( \bar{\theta} \). Hence \( \bar{\theta} \) is also the unique optimal point for (1.1)<sup>\(k\rangle</sup>. Clearly, (4.8) is valid.

(2) In this case, using \( \gamma_0 = \Delta x^T_k \bar{d}_q \) and \( ||\Delta x_k|| < \eta_0 \) in (4.10), we have

\[
1 + \gamma_0 \geq 1 - |\bar{\gamma}_i| \geq 1 - ||\Delta x_k|| \cdot ||\bar{d}_i|| > |\bar{\lambda}_i| \geq 0, \quad |\theta_k(1 + \gamma_0)| \geq 1 - |\gamma_0| > |\bar{\lambda}_i|.
\]

(4.15A)

and

\[
|\gamma_0| \leq ||\Delta x_k|| \cdot ||\bar{d}_i|| < 1 - |\bar{\lambda}_i| < 1.
\]

(4.15B)

This implies that \( \lambda_{\ell}^{<k>} \neq 0 \) by (4.14) (so \( \bar{\theta} \) is no longer a solution to (1.1)<sup>\(k\rangle</sup> since \( \lambda_{\ell}^{<k>} \) is the multiplier of an artificial constraint), and \( \text{sign}(\lambda_{\ell}^{<k>}) = \text{sign}(\theta_k(1 + \gamma_0)) = \theta_k \). Therefore, the direction \( \bar{d} \) defined by (4.9) is descent since it matches the first case in (2.6).

Furthermore, the first inequality in (4.15A) implies that the step length defined in (4.11) satisfies \( \alpha^{<k>} > 0 \). Also we have \( \alpha^{<k>} \leq ||\Delta x_k|| \cdot ||\bar{d}_i||/(1 - ||\Delta x_k|| \cdot ||\bar{d}_i||) < \delta' \) and, similarly, \( \alpha^{<k>} < \alpha' \). Therefore, \( \bar{\theta} \), defined by (4.11), is feasible for (1.1)<sup>\(k\rangle</sup> and it is a base point since

\[
r_k^{<k>}(\bar{\theta}) = (x_k + \Delta x_k)^T(\bar{\theta} + \alpha^{<k>} \bar{d}) - y_k = r_k(\bar{\theta}) + \Delta x_k^T \bar{\theta} + \alpha^{<k>} (x_k + \Delta x_k)^T \bar{d} = 0
\]

by (4.11) and (4.9).

Now we show that \( \bar{\theta} \) is optimal for (1.1)<sup>\(k\rangle</sup>. Because the move from \( \bar{\theta} \) along \( \bar{d} \) with step length \( \alpha^{<k>} \) does not cross any boundary planes associated with the inactive equations, then the “cost” vector \( \bar{c} = \tilde{c}(\bar{\theta}) \) for (1.1)<sup>\(k\rangle</sup> satisfies \( \bar{c} = \tilde{c} \). Notice that the base matrix at \( \bar{\theta} \) is \( \bar{A} \) as defined in Proposition 4.1, since \( k = j_q \in \bar{A}, \) the active set at \( \bar{\theta} \) for (1.1)<sup>\(k\rangle</sup>. From Proposition 4.1, \( \bar{A} \) is nonsingular since \( ||\Delta x_k|| < (1 - |\bar{\lambda}_q|)/||\bar{d}_q|| < 1/||\bar{d}_q|| \) (see (4.10) and (4.4)), and the transposes of row vectors of its inverse are shown in (4.5). Consequently, the multipliers at \( \bar{\theta} \) for (1.1)<sup>\(k\rangle</sup> are

\[
\lambda_i = \bar{d}_i^T \bar{c} = \tilde{d}_i^T \tilde{c} = \begin{cases} \frac{\lambda_q}{1 + \Delta x_k^T \bar{d}_q}, & i = q \ (k = j_q), \\ \frac{\lambda_i - (\Delta x_k^T \bar{d}_i) \bar{\lambda}_q}{1 + \Delta x_k^T \bar{d}_q}, & i \neq q. \end{cases}
\]

(4.16)

Now we estimate the values of the multipliers. First, from (4.10), \( |\lambda_q| \leq |\bar{\lambda}_q|/(1 - ||\Delta x_k|| \cdot ||\bar{d}_q||) < 1 \) since \( ||\Delta x_k|| < (1 - |\bar{\lambda}_q|)/||\bar{d}_q|| \). Furthermore, since \( ||\Delta x_k|| < \eta_1 \) in (4.10),

\[
|\lambda_i| \leq |\bar{\lambda}_i| + \frac{|\lambda_q| \cdot ||\bar{d}_i|| \cdot ||\Delta x_k||}{1 - ||\bar{d}_i|| \cdot ||\Delta x_k||} < 1, \quad j_i \in \bar{A}_S \setminus \{ k \} = \bar{A}_S \setminus \{ k \},
\]

while since \( ||\Delta x_k|| < \eta_2 \),

\[
|\lambda_i| \leq -|\bar{\lambda}_i| + \frac{|\lambda_q| \cdot ||\bar{d}_i|| \cdot ||\Delta x_k||}{1 - ||\bar{d}_i|| \cdot ||\Delta x_k||} < 0, \quad j_i \in \bar{A}_I = \bar{A}_I.
\]

Hence \( \bar{\theta} \) is the unique optimal point for (1.1)<sup>\(k\rangle</sup>.
Finally, from (2.7) and since $\bar{A}_s^{<k>}(\bar{\beta}) = \bar{A}_s \setminus \{k\}$, the new optimal value is

$$S^{<k>}(\bar{\beta}) = S^{<k>}(\bar{\beta}) + \alpha^{<k>}(\mathbf{c}^{<k>}^T \mathbf{d} + \sum_{i \in \bar{A}_s \setminus \{k\}} |\mathbf{x}_i^T \mathbf{d}|).$$

(4.17)

The value of $S^{<k>}(\bar{\beta})$ is shown in (4.8), and, by (4.13), (4.9) and (4.7A), we have $\mathbf{c}^{<k>}^T \mathbf{d} = -\bar{\theta}_k \bar{\lambda}_y - (1 + \gamma_0)$.

Substituting these values into (4.17) and, using the fact that $\mathbf{x}_i^T \mathbf{d} = 0$, $\forall i \in \bar{A}_s \setminus \{k\}$, we obtain (4.12). ■

The last case to consider is covered in the following result.

**Theorem 4.3** Assume that $k = j_0 \in \bar{A}_s$, so $r_k(\bar{\beta}) = 0$, and $\Delta x_i^T \bar{\beta} = 0$. If

$$||\Delta x_k|| < \eta \equiv \min\{\eta_1, \eta_2, (1 - |\bar{\lambda}_y|)/||\bar{\lambda}_y||\},$$

where $\eta_1$ and $\eta_2$ are defined by (4.10), then $\bar{\beta}$ is also the solution to (1.1)<sup>^{<k>}\end{sup> with unchanged optimal value.

**Proof** In this case, at $\bar{\beta}$ we have $\mathbf{c}^{<k>} = \bar{\mathbf{c}}$ and $k \in \bar{A}^{<k>} = \bar{A}$. Hence the base matrix at $\bar{\beta}$ for (1.1)<sup>^{<k>}\end{sup> is $\bar{A}$ as defined in Proposition 4.1 and the transposes of row vectors of its inverse are given in (4.5). Therefore, the multipliers $\lambda_i^{<k>}$ at $\bar{\beta}$ are the same as $\bar{\lambda}_i$ in (4.16). The conclusion follows by bounding these multipliers in the same way as in the proof of case (2) of Theorem 4.2. ■

Theorems 4.1, 4.2 and 4.3 show that the $L_1$ regression solution is stable with respect to sufficiently small perturbations to $x_k$. However, the solution is not uniformly stable. In all cases the upper bound on $||\Delta x_k||$ may be small if the base matrix is ill-conditioned or if the solution is close to being non-unique. This allows for the possibility that a small perturbation will lead to a significant change in the solution, i.e. the solution may exhibit instability. This is consistent with the results of Ellis (1998) about the “singular set” for the unconstrained $L_1$ regression problem, i.e. the set of data points at which arbitrarily small perturbations can substantially change the fit. He characterized this set (and its dimension) as the points where the solution is not unique (which includes points at which the design matrix is not of full rank as well as other points). The examples in Figure 2 of Ellis (1998) exhibit this type of instability.

5. **Numerical results**

**Example 5.1** Brownlee stack loss data

**I. Perturbations to the responses**

We consider the well known stack loss data of Brownlee (1965, p 454), which was also analysed by Narula and Wellington (1985). The model is $y_i = \mathbf{x}_i^T \mathbf{\beta} + \varepsilon_i$, $i = 1, \ldots, 21$, $\mathbf{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]^T$, where the response $y_i$ and the first 3 components of the vector $\mathbf{x}_i^T$ are given in row $i$ of Table 5.1. The fourth component of $\mathbf{x}_i^T$ is 1 for all $i$.

To solve for the $L_1$ regression estimate of $\mathbf{\beta}$, we used the LP formulation of the unconstrained problem (1.1A) given in Barrodale and Roberts (1978). This LP problem was solved using the linprog routine in the MATLAB Optimization Toolbox, which by default employs an interior point algorithm. The $L_1$ regression estimate of $\mathbf{\beta}$ is $\mathbf{\beta} = (0.832, 0.574, -0.061, -0.39689)^T$, and it is determined by the active equations, numbers 2, 8, 16 and 18.

First we consider the responses $y_i$ for the inactive equations. Using Lemma 3.1, for each $y_i$ we determine the interval for which the solution $\mathbf{\beta}$ remains optimal. The results are given on the right of Table 5.1, where L.B. and U.B. stand for the lower bound and upper bound of the interval, respectively. These results agree with those in Table 4 of Narula and Wellington (1985).
Table 5.1 Intervals for values of predictor and response variables in stack loss model

Now we consider perturbations to the responses \( y_i \) for the active equations. For each \( i = 2, 8, 16 \) and 18, we perturb \( y_i \) to \( y_i + \Delta y_i \) (leaving the other values fixed), where \( \Delta y_i = 0.01z \) and \( z \) is a pseudo-random number distributed normally with mean 0 and variance 1. This is replicated 40 times. In every case the conditions of Lemma 3.3 are satisfied and so we can use (3.8) to determine the solution \( \overline{\beta} \) of the perturbed problem. For each of the active equations, the \( L_1 \) norm of the error \( \| \overline{\beta} - \hat{\beta} \|_1 \) due to each of the 40 perturbations is plotted in Figure 5.1(a). Note the large difference in the sensitivity; clearly the solution is least sensitive to small perturbations in \( y_2 \) and most sensitive to small perturbations in \( y_{18} \).

![Graphs showing \( L_1 \) norm of error for perturbations to responses and predictors](image)

**Fig. 5.1(a)** \( L_1 \) norm of error in solution (+) for perturbations to responses  
**Fig. 5.1(b)** \( L_1 \) norm of error in solution (+) for perturbations to predictors

II. **Perturbations to the values of predictor variables**

We now consider perturbations to the predictor variables in the stack loss model, which were also analysed by Narula and Wellington (2002). First we consider the inactive equations, i.e. observations with non-zero residuals. We restrict our attention to the first three predictor variables (excluding the fourth variable which...
is identically 1). For each predictor variable we use Corollary 4.1 to determine the interval of values for which the solution $\mathbf{\beta}$ remains optimal. The results are given in Table 5.1, where L.B denotes the lower bound and U.B. the upper bound of the interval, respectively. For most of the variable values the results agree with those in Table 4 of Narula and Wellington (2002). However, three of their values (the U.B. values for $x_{3,2} = 25, x_{10,3} = 80$ and $x_{12,1} = 58$) are incorrect and a few others are inaccurate.

For the active equations, we use Theorem 4.2 to find the $L_1$ norm error $\| \mathbf{\beta} - \mathbf{\beta} \|_1$ resulting from a perturbation to each of the predictor variable values (one at a time) in these equations. For each equation, we again restrict our attention to the first three variables. For each variable we take 40 perturbations defined by $\Delta x_{kl} = 0.001z$, where $z$ is a pseudo-random number distributed normally with mean 0 and variance 1. The conditions of Theorem 4.2 hold for all the perturbed problems. The resulting $L_1$ norm errors are displayed in Figure 5.1(b), where the first vertical display in each group of 3 corresponds to the first variable, and similarly for the second and third. Note that there is significant variation in the sensitivity due to perturbations in the predictor variables; the lowest sensitivity is for the third predictor value in observation number 2 and the greatest is for the first predictor value in observation number 18.

**Example 5.2** Curve fitting and estimation of first and second derivatives

Consider the curve fitting problem of estimating $f(t)$ from data $y_i = f(t_i) + \epsilon_i, i = 1, \ldots, n$. One approach to this is to estimate $f(t)$ by a truncated trigonometric series

$$f_p(t) = c_0 + \sum_{k=1}^{p_1} c_k \cos 2k\pi t + \sum_{k=1}^{p_2} d_k \sin 2k\pi t,$$

(5.1)

where $p = 1 + p_1 + p_2 \ll n$. Here $p$ plays the role of a discrete smoothing parameter. Applying $L_1$ regression, we have a problem of the form of (1.1A) with $\mathbf{\beta} = (c_0, c_1, \ldots, c_{p_1}, d_1, \ldots, d_{p_2})^T$.

Now consider the problem of estimating the derivative of a function $g(t)$, $0 \leq t \leq 1$, satisfying $g(0) = 0$, given discrete noisy data $y_i = g(t_i) + \epsilon_i, i = 1, \ldots, n$. This numerical differentiation problem is known to be ill-posed (see Andersen and Bloomfield (1974)) and hence requires some form of stabilization. A simple approach to this is to seek an approximate solution in a low dimensional subspace. Clearly the true solution $f(t) = g'(t)$ satisfies $\int_0^t f(s) \, ds = g(t)$, since $g(0) = 0$. We approximate $f(t)$ by $f_p(t)$ in (5.1) and evaluate the indefinite integral at the points $t_i$, giving the overdetermined system

$$\int_0^{t_i} f_p(t) \, dt = c_0 t_i + \sum_{k=1}^{p_1} c_k \sin 2k\pi t_i / (2k\pi) + \sum_{k=1}^{p_2} d_k (1 - \cos 2k\pi t_i) / (2k\pi) \approx y_i, \quad i = 1, \ldots, n.$$

Applying $L_1$ regression, we have a problem of the form of (1.1A) with $\mathbf{\beta} = (c_0, c_1, \ldots, c_{p_1}, d_1, \ldots, d_{p_2})^T$.

The above method can also be used to estimate the second and higher derivatives of $g(t)$. If $g(0) = g'(0) = 0$, we approximate $g''(t)$ by $f_p(t)$ in (5.1), integrate twice from 0 and evaluate at the points $t_i$ to get the system

$$c_0 t_i^2 / 2 + \sum_{k=1}^{p_1} c_k (1 - \cos 2k\pi t_i) / (2k\pi)^2 + \sum_{k=1}^{p_2} d_k (2k\pi t_i - \sin 2k\pi t_i) / (2k\pi)^2 \approx y_i, \quad i = 1, \ldots, n.$$

The unknown coefficients are then obtained by $L_1$ regression.

For the numerical experiments, we take $g(t) = 10(-t^3/3 + t^2/2)$ and so $g'(t) = 10(-t^2 + t)$ and $g''(t) = 10(-2t + 1)$. The data are $y_i = g(t_i) + \epsilon_i, i = 1, \ldots, n$, where $n = 50, t_i = i/n$ and the $\epsilon_i$ are pseudo-random numbers distributed normally with mean 0 and standard deviation 0.01, i.e. approximately 1% noise. The $L_1$ estimates of $g(t)$, $g'(t)$ and $g''(t)$ are computed by solving the $L_1$ regression problems defined above with $p_1 = p_2 = 5$ and $p = 11$. The computations were done in MATLAB, using the linprog routine from the Optimization Toolbox to solve the LP formulation of the $L_1$ regression problem.
Fig. 5.2(a) $g(t) = 10(-t^3/3 + t^2/2)$ (solid), data (+) and $L_1$ regression estimate (dashed)

Fig. 5.2(b) $g'(t) = 10(-t^2 + t)$ (solid), and $L_1$ regression estimate (dashed)

Fig. 5.2(c) $g''(t) = 10(-2t + 1)$ (solid) and $L_1$ regression estimate (dashed)

Fig. 5.3(a) $L_1$ norm of error in coefficients (+) and bound (o) for $L_1$ regression estimates of $g(t)$

The computed estimates are plotted in Figures 5.2(a), 5.2(b) and 5.2(c) together with the true functions and the data. Clearly, on the interior of the interval (i.e. excluding the end behaviour caused by the mismatch in boundary values), the estimates get worse as the order of derivative of $g$ increases from 0 to 1 to 2. This is consistent with the fact that second order numerical differentiation is more ill-posed than first order numerical differentiation, which is more ill-posed than the curve fitting problem.

To illustrate Theorem 3.1, we consider a perturbation $\Delta y$ of the vector $y = (y_1, \ldots, y_n)^T$ with $\Delta y_i = \pm 10^{-9}$, where the sign is pseudo-random with probability 0.5. By using 40 different random sequences of the signs, we obtain 40 different trial perturbation vectors $\Delta y$. For each perturbation vector, we compute the solution $\bar{y}$ to the perturbed $L_1$ regression problem and compute the $L_1$ error $\|\bar{y} - \beta\|_1$, where $\beta$ is the solution to the unperturbed problem (i.e. with data vector $y$). Then the bound $\eta \alpha \|D\|_\infty$ from Theorem 3.1 is computed, and the errors are plotted with the bound (on a log scale) giving one of the vertical displays in each of Figures 5.3(a), 5.3(b) and 5.3(c). In order to consider the effect of the random nature of the unperturbed data vector $y$, the above procedure was replicated 10 times, each with a different pseudo-random noise vector $\varepsilon$ (keeping the standard deviation fixed at 0.01). This gives the 10 vertical displays in each of Figures 5.3(a), 5.3(b) and 5.3(c). It is easy to see that with the form (5.1) of $f_p(t)$, the error $\|\bar{y} - \beta\|_1$ in the coefficients is a bound on the $\infty$-norm error in the function estimates of $g(t)$, $g'(t)$ and $g''(t)$. 

16
Fig. 5.3(b) $L_1$ norm of error in coefficients (+) and bound (o) for $L_1$ regression estimates of $g'(t)$

It is clear from these figures that the bound in Theorem 3.1 is a fairly good bound on the error $||\bar{\beta} - \hat{\beta}||_1$, and it generally follows the behaviour of the errors for the 10 different replicates. Also note that both the bound and the errors are larger for the estimation of $g''(t)$ compared to $g'(t)$ compared to $g(t)$. This highlights the greater sensitivity in the more ill-posed problem of second order numerical differentiation. It is quantified in the bound by the size of $||\tilde{D}||_{\infty}$, where $\tilde{D}$ is the inverse of the base matrix defined by the active equations at $\hat{\beta}$. The value of $||\tilde{D}||_{\infty}$ is different for each of the 10 replicates as shown in Table 5.2 below. Clearly the values of $||\tilde{D}||_{\infty}$ are significantly higher for second derivative estimation compared to the other problems.

<table>
<thead>
<tr>
<th>Case</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g(t)$</td>
<td>11.92</td>
<td>2.34</td>
<td>34.86</td>
<td>6.29</td>
<td>7.30</td>
<td>13.97</td>
<td>5.70</td>
<td>43.78</td>
<td>4.71</td>
<td>29.95</td>
<td>16.08</td>
</tr>
<tr>
<td>$g'(t)$</td>
<td>75.37</td>
<td>66.59</td>
<td>67.32</td>
<td>63.54</td>
<td>242.86</td>
<td>68.29</td>
<td>71.98</td>
<td>75.56</td>
<td>55.98</td>
<td>66.74</td>
<td>85.42</td>
</tr>
<tr>
<td>$g''(t)$</td>
<td>2029</td>
<td>12538</td>
<td>2315</td>
<td>990</td>
<td>5558</td>
<td>4257</td>
<td>6343</td>
<td>1632</td>
<td>1569</td>
<td>2921</td>
<td>4015.1</td>
</tr>
</tbody>
</table>

Table 5.2 Values of $||\tilde{D}||_{\infty}$ for the 10 replicates and their mean

Notice that there are no error values plotted for replicate number 5 in Figure 5.3(a). This is because the unperturbed $L_1$ regression problem in this case does not have a unique solution $\hat{\beta}$ (so the error $||\bar{\beta} - \hat{\beta}||_1$ is not well defined). Note that the size $||\Delta y||$ of the perturbations was chosen to be $10^{-9}$ so that the conditions of Theorem 3.1 held in all the simulation problems, allowing comparisons across the problems. For many of these problems a considerably larger perturbation size could have been used with Theorem 3.1.

In the general problems above, if it is known that the desired function $f(t)$ satisfies some constraint, it may be beneficial to impose this constraint on the estimate $f_p(t)$. For example, if in the estimation of $f(t) = g'(t)$ from the data in Figure 5.2(a), if it is known that $f(t)$ is concave, then one should impose the constraint $f''_p(t) \leq 0$, at least at some values of $t$. This leads to a constrained $L_1$ regression problem, which we examine in more detail in Lukas and Shi (2004).

References


