
Sensitivity Analysis of Constrained Linear $L_1$ Regression:  
Perturbations to Constraints, Addition and Deletion of Observations  
Mark A. Lukas$^a$, Mingren Shi$^b$  

$^a$ Mathematics and Statistics, Murdoch University, Murdoch, W.A. 6150, Australia  
Fax: 61893606332 E-mail: M.Lukas@murdoch.edu.au  

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

Abstract. This paper extends the direct sensitivity analysis of Shi and Lukas (2005) of linear $L_1$ (least absolute deviations) regression with linear equality and inequality constraints on the parameters. Using the same active set framework of the reduced gradient algorithm (RGA), we investigate the effect on the $L_1$ regression estimate of small perturbations to the constraints (constants and coefficients). It is shown that the constrained estimate is stable, but not uniformly stable, and in certain cases it is unchanged. We also consider the effect of addition and deletion of observations and determine conditions under which the estimate is unchanged. The results demonstrate the robustness of $L_1$ regression and provide useful diagnostic information about the influence of observations. Results characterizing the (possibly non-unique) solution set are also given. The sensitivity results are illustrated with numerical simulations on the problem of derivative estimation under a concavity constraint.  

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 $y_i = x_i^T \beta + \varepsilon_i$, $i = 1, \ldots, n$, where the $p \times 1$ parameter vector $\beta$ is known to satisfy additional linear equality constraints $x_i^T \beta - y_i = 0$ and/or inequality constraints $x_i^T \beta - y_i \leq 0$. Important applications occur in 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)). In many such problems there is some extra information about the unknown curve or solution that can be used to constrain the parameters. In particular, if the solution (or some linear functional, e.g. an integral) is known to have a certain fixed value at some point, we obtain an equality constraint on the parameters. If the solution is known to be positive, monotone, concave or convex, this leads to certain linear inequality constraints on the parameters (see Wahba (1982) and O’Leary and Rust (1986)). Constrained regression problems also arise in certain biometric and econometric models (see Judge et al. (1985)).  

Using the method of least absolute deviations, we define the constrained $L_1$ regression estimate of $\beta$ to be the solution of the problem (denoted $LL_1$)  

$$\begin{align*}
\text{minimize} & \quad S(\beta) = \sum_{i=1}^{n} |x_i^T \beta - y_i|, \quad \beta \in \mathbb{R}^p, \\
\text{subject to} & \quad x_i^T \beta - y_i = 0, \quad i \in \mathcal{E} = \{n + 1, \ldots, n + n_{\mathcal{E}}\}, \quad (1.1A) \\
& \quad x_i^T \beta - y_i \leq 0, \quad i \in \mathcal{I} = \{n + n_{\mathcal{E}} + 1, \ldots, n + n_{\mathcal{E}} + n_{\mathcal{I}}\}, \quad (1.1C)
\end{align*}$$

where $\mathcal{E}$ and $\mathcal{I}$ refer to equalities and inequalities, respectively, and we assume that $n_{\mathcal{E}} < p < n + n_{\mathcal{E}} + n_{\mathcal{I}}$.  
Shi and Lukas (2005) investigated the sensitivity of the solution to (1.1) with respect to perturbations in the
responses $y_i$ and row vectors $x_i$, $1 \leq i \leq n$, of the design matrix $X$. The analysis was done using the active set framework of the reduced gradient algorithm (RGA) developed in Shi and Lukas (2002). In this paper we use the same framework and extend the sensitivity analysis to cover perturbations to the constraints (both the constants and coefficients) and the addition and deletion of observations.

It is well known that unconstrained $L_1$ regression is robust; the $L_1$ estimate is resistant to outliers in the response $y$ and may not be changed at all by some large perturbations in $y$. The same is true for constrained $L_1$ regression, and the estimate is stable, but not uniformly stable, with respect to small perturbations in $y$ and $x_i$, $i = 1, \ldots, n$ (Shi and Lukas (2005), Ellis (1998)). Here we show that similar results apply for perturbations to the constraints in (1.1): the constrained estimate is stable for small perturbations and it may not even change with some large perturbations in the constraints. The stability is not uniform but depends on the degree of ill-conditioning and, for the constraint coefficient vectors, also on how close the estimate is to being non-unique.

For unconstrained $L_1$ regression, the effect of deletion of observations has been investigated by Narula and Wellington (1985), Dušačová (1992) and Castillo et al. (2001). These works are based on the LP formulation of the $L_1$ regression problem. Here we use the direct active set approach on the general $LL_1$ problem (1.1) to find conditions under which an existing solution remains optimal after the deletion of an observation. We also show how the RGA can be used to find a new solution to the problem with a deleted observation, starting from the original solution. The $L_1$ version of the Cook distance can then be computed efficiently to determine the influence of each observation on the $L_1$ regression estimate. This provides important diagnostic information about the model. A result about the addition of an observation shows that constrained $L_1$ regression is robust with respect to the new response.

A brief description of the RGA framework and optimality conditions for (1.1) is given in Section 2. It is known that the solution of an $L_1$ regression problem can be non-unique (Bloomfield and Steiger (1983)). In Section 3 we derive results characterizing the solution set of (1.1) and show that it can be computed efficiently using the RGA. In Sections 4 and 5, we investigate the effect on the solution to (1.1) of a perturbation to the constants and coefficients of the constraints, respectively. Results relating to the deletion and addition of an observation are derived in Sections 6 and 7, respectively.

In Section 8 we consider the problem of estimating the derivative $f(t) = g(t)$ from data $y_i = g(t_i) + \varepsilon_i$, $i = 1, \ldots, n$, under the concavity constraint $f''(t) \leq 0$. The estimation of derivatives arises in many applications, in particular in the analysis of growth curves (Gasser et al. (1984) and Eubank (1988)) and pharmacokinetic response curves (Song et al. (1995)). Numerical differentiation is an ill-posed problem, meaning that the solution is sensitive to errors in the data (see Anderssen and Bloomfield (1974)), and such problems were a major motivation for this work. We use a truncated trigonometric series for the estimate $f_p(t)$ of the derivative and find the estimate by solving a constrained linear $L_1$ regression problem. Results of numerical simulations illustrate the sensitivity results from Sections 4, 5 and 6, and also a result from Shi and Lukas (2005) on perturbations to the responses. We also consider the effect of increasing the number of points at which the the constraint $f''_p(t) \leq 0$ is evaluated.

2. Active set framework

We use the same notation, definitions and optimality results as in Shi and Lukas (2005), but we repeat them here for completeness (see also Bloomfield and Steiger (1983), and Osborne (1985)). An inequality constraint $x_i^T \beta - y_i \leq 0$ is said to be active at $\beta$ if $x_i^T \beta - y_i = 0$ (equality constraints are automatically
active). Similarly we say that a model equation \( y_i = x_i^T \beta + \varepsilon_i \) is an active equation at \( \beta \) if the residual is 0.

Let the **active set** at a given point \( \beta \) and its associated sets be

\[
A = A(\beta) = \{ i | r_i(\beta) \equiv x_i^T \beta - y_i = 0, \ 1 \leq i \leq n + n_\varepsilon + n_\zeta \},
\]

\[
A_s = A \cap \{1, \ldots, n\}, \quad A_{tr} = A \cap T, \quad A_{s}^c = \{1, \ldots, n\} \setminus A_s.
\]

(2.1A)

(2.1B)

Note that the residual \( r_i(\beta) \) is the opposite of the usual definition.

**Definition 2.1** (1) Denoting \( \text{rank}\{x_i\}_{i=1}^{n-r_\varepsilon-n_\zeta} = r \leq p, \) we say \( \beta \) is a base point of problem (1.1) if the rank of the active gradient vectors \( \{x_i| i \in A\} \) is \( r \). (2) A feasible base point \( \beta \) is nondegenerate if the active gradient vectors \( \{x_i| i \in A\} \) 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 \( \beta \) to be a solution to (1.1) are that \( \beta \) is feasible, i.e.

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

and there exist multipliers \( \lambda_i, i \in A \), such that

\[
\mathbf{c} \equiv \sum_{i \in A_{s}^c} \sigma_i \mathbf{x}_i = \sum_{i \in A_s} \lambda_i \mathbf{x}_i + \sum_{i \in \mathcal{E}} \lambda_i \mathbf{x}_i + \sum_{i \in \mathcal{T}} \lambda_i \mathbf{x}_i,
\]

\[
|\lambda_i| \leq 1, \quad i \in A_s, \quad \lambda_i \leq 0, \quad i \in A_t,
\]

(2.3A)

(2.3B)

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

To illustrate this result, consider the simplest unconstrained, univariate case with \( X = [1, \ldots, 1]^T \). Here we have \( r_i = \beta - y_i \) so \( c \) is just the number of values \( y_i \) that are less than \( \beta \) minus the number that are greater than \( \beta \). For optimality of \( \beta \), condition (2.3) requires this number \( c \) to be less than or equal to one in magnitude, i.e. \(-1, 0\) or \(1\). Clearly, if \( c = 0 \), the unique solution is the median of the data, while if \( c = \pm 1 \), all \( \beta \) between the middle two data values are optimal.

We will also use the following 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.)

The sensitivity results will be derived in the active set framework of the reduced gradient algorithm (RGA) as developed in Shi and Lukas (2002) (see also Bloomfield and Steiger (1983) and Osborne (1985,1987)). The RGA is a descent method which moves from one feasible base point to an adjacent one. We only consider the nondegenerate case here (see Shi and Lukas (2002) for the degenerate case). The main steps of the RGA are as follows.

**Step (0)** Find a feasible starting point \( \tilde{\beta}^{(0)} \). For an unconstrained problem, an arbitrary point \( \tilde{\beta} = (\tilde{\beta}_1, \ldots, \tilde{\beta}_p)^T \) (e.g. \( \tilde{\beta} = 0 \)) can serve as the starting point. Then we add artificial constraints \( e_i^T \tilde{\beta} - \tilde{\beta}_i = \beta_i - \tilde{\beta}_i \leq 0, \ i \in \{1, 2, \ldots, p\} \) to (1.1A) to form an augmented problem for which \( \tilde{\beta} \) is a base point (with rank \( p \)). For the general constrained \( LL_1 \) problem, we solve an auxiliary LP problem (see Shi and Lukas (2002)) to find a feasible starting point. At the start an \( LL_1 \) problem may also be augmented with artificial constraints. The active set is \( \hat{A} = \hat{A}_S \cup \mathcal{E} \cup \hat{A}_T \cup \hat{A}_0 \), where \( \hat{A}_S \) and \( \hat{A}_T \) are defined in (2.1) and \( \hat{A}_0 \) denotes a set of indices of artificial constraints. Note that these artificial constraints are only a computational device; it will be seen in step (3) that they are not enforced by the algorithm. Set \( k = 0 \) and go to step (1).
Step (1) Compute the “cost” vector and multipliers. Let \( \beta^{(k)} = \hat{\beta} \) and let its active set, the base matrix \( \hat{\mathbf{A}} \) and its inverse \( \hat{\mathbf{D}} \) be, respectively,

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

where \( \hat{\mathbf{d}}_i \) is the \( i \)th row vector of \( \hat{\mathbf{D}} \). As in (2.3A), compute the “cost” vector \( \hat{\mathbf{c}} \) and the multipliers \( \hat{\lambda}_i \) by

\[
\hat{\mathbf{c}} = \mathbf{c}(\hat{\beta}) \equiv \sum_{i \in \hat{\mathbf{A}}} \sigma_i \mathbf{x}_i = \sum_{i=1}^{p} \hat{\lambda}_i \mathbf{x}_i \quad \text{and} \quad \hat{\lambda}_i = \hat{\mathbf{d}}_i^T \hat{\mathbf{c}}, \quad i = 1, 2, \ldots, p, \tag{2.4}
\]

where \( \sigma_i = \text{sign}(r_i(\hat{\beta})) \).

Step (2) Test the optimality conditions. Let

\[
\max\{\hat{\lambda}_i \mid j_i \in \hat{\mathbf{A}}_0\} = \hat{\lambda}_{j_0}, \quad \max\{\hat{\lambda}_i \mid j_i \in \hat{\mathbf{A}}_s\} = \hat{\lambda}_s, \quad \text{and} \quad \max\{\hat{\lambda}_i \mid j_i \in \hat{\mathbf{A}}_t\} = \hat{\lambda}_t.
\]

If

\[
|\hat{\lambda}_{j_0}| = 0, \quad |\hat{\lambda}_s| \leq 1 \quad \text{and} \quad \hat{\lambda}_t \leq 0, \tag{2.5}
\]

then, from (2.3), \( \hat{\beta} \) is optimal and the algorithm is terminated. Otherwise, go to step (3).

Step (3) Choose a feasible descent direction \( \mathbf{d} \). Any one of the choices

\[
\mathbf{d} = \begin{cases} 
-\text{sign}(\hat{\lambda}_{j_0}) \hat{\mathbf{d}}_{j_0}, & \text{if} \quad |\hat{\lambda}_{j_0}| \neq 0, \\
-\text{sign}(\hat{\lambda}_{j_s}) \hat{\mathbf{d}}_{j_s}, & \text{if} \quad |\hat{\lambda}_{j_s}| > 1, \\
-\hat{\mathbf{d}}_{j_t}, & \text{if} \quad \hat{\lambda}_{j_t} > 0,
\end{cases} \tag{2.6}
\]

is a feasible descent direction at \( \hat{\beta} \) for (1.1) (see Shi and Lukas (2002)). Note that any direction is regarded as feasible for the artificial constraints. A move from \( \hat{\beta} \) 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 \( P_j = \{\beta| x_j^T \beta - y_j = 0\} \)) but remain on the intersection of the other \( p - 1 \) boundary planes \( P_{j_{j_i}} \), \( j_i \in \hat{\mathbf{A}} \setminus \{j_q\} \). 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 \( |\hat{\lambda}_{j_0}| \neq 0 \), we choose \( q = q_0 \) since we want to delete the artificial constraints.

Step (4) Compute the search step length. Let the shortest step to the boundary planes \( P_j \) associated with inactive equations and with inactive inequality constraints be, respectively,

\[
\delta' = \delta'(\mathbf{d}) = \min\{\delta_j \equiv -r_j(\hat{\beta})/x_j^T \mathbf{d} \mid \delta_j > 0, \quad n \geq j \in \hat{\mathbf{A}}_s, \quad x_j^T \mathbf{d} \neq 0\}, \tag{2.7A}
\]

\[
\alpha' = \alpha'(\mathbf{d}) = \min\{\alpha_j \equiv -r_j(\hat{\beta})/x_j^T \mathbf{d} \mid \alpha_j > 0, \quad n < j \in \hat{\mathbf{A}} \setminus \hat{\mathbf{A}}_s, \quad x_j^T \mathbf{d} > 0\}, \tag{2.7B}
\]

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

Step (5) Update the base point and its inverse matrix. Let \( \mathbf{\tilde{\beta}} = \hat{\beta} + \alpha \mathbf{d} \) be the new base point with base matrix \( \mathbf{\tilde{A}} \), and let \( m \) be the index of the new active equation or constraint. The new inverse matrix \( \mathbf{\tilde{D}} = \mathbf{\tilde{A}}^{-1} = [\mathbf{\tilde{d}}_1, \mathbf{\tilde{d}}_2, \ldots, \mathbf{\tilde{d}}_p]^T \) can be obtained from \( \hat{\mathbf{D}} \) using the formula

\[
\mathbf{\tilde{d}}_i = \begin{cases} 
\frac{1}{x_m^T \mathbf{d}} \mathbf{\tilde{d}}_i = \frac{1}{a_m^T \mathbf{d}} \hat{\mathbf{d}}_i, & \text{if} \quad i = q, \\
\mathbf{\tilde{d}}_i - (x_m^T \mathbf{\tilde{d}}_i) \frac{1}{x_m^T \mathbf{d}} \hat{\mathbf{d}}_i = \mathbf{\tilde{d}}_i - (x_m^T \hat{\mathbf{d}}_i) \hat{\mathbf{d}}_i, & \text{if} \quad i \neq q.
\end{cases} \tag{2.8}
\]
Since $\alpha \leq \beta'$, the new value of $S$ is

$$S(\bar{\beta}) = S(\beta') + \alpha (\bar{c}^T d + \sum_{i \in \hat{A}} |x_i^T d|). \quad (2.9)$$

Define $\beta^{(k+1)} = \bar{\beta}$. Set $k = k + 1$ and go to step (1).

Since there are finitely many base points and the RGA is a descent method, it terminates at an optimal base point.

3. **Structure of non-unique solutions**

For the unconstrained $L_1$ regression problem, Bloomfield and Steiger (1983) (Sec. 1.3) derived several important properties of the (possibly non-unique) solution set. For the general problem (1.1), we will characterize the solution set $\mathcal{M}$ in the framework of the RGA and show how it can be computed easily.

It is easy to show that the function $S(\beta)$ is convex. Because the feasible region defined by the constraints (1.1B) and (1.1C) is convex, we immediately have the following result.

**Lemma 3.1** The solution set $\mathcal{M}$ of (1.1) is convex, so if $\mathcal{M}$ contains more than one point, it contains infinitely many.

Since the solution set $\mathcal{M}$ is convex, the same argument used in Bloomfield and Steiger (1983) (Sec. 1.3 Thm. 4) applies to give the following result (the case of inactive constraints is obvious).

**Theorem 3.1** Any two points $\beta$ and $\bar{\beta}$ of $\mathcal{M}$ are on the same side of every boundary plane corresponding to an inactive equation or constraint, i.e. $r_i(\beta)r_i(\bar{\beta}) > 0$ for all $i \in \mathcal{A}'(\beta) \cap \mathcal{A}'(\bar{\beta})$.

To characterize the solution set, there are two cases to consider. Let $X = [x_1 \cdots x_{n+n_x+n_z}]^T$ be the extended design matrix consisting of all the row vectors for the model equations and constraints. First we consider the case where $X$ has full rank $p$. In this case, for any base point $\beta$, the active gradients $x_i$, $i \in A_0(\beta) \cup E \cup A_0(\beta)$, span $\mathbb{R}^p$. From Theorem 2.1, we know that the solution set $\mathcal{M}$ of (1.1) contains a base point. Using the same argument as in Bloomfield and Steiger (1983) (Sec. 1.3 Thm. 5) with Theorem 3.1, we have the following.

**Theorem 3.2** If $X$ has full rank, the solution set $\mathcal{M}$ of (1.1) is the convex hull of the base points contained in it.

Thus, to find all the non-unique solutions of (1.1) it is necessary only to compute all the optimal base points. This can be done efficiently using the RGA above. First we obtain one optimal base point $\bar{\beta}$ and compute the corresponding multipliers $\lambda_i$, $i = 1, \ldots, p$, $j_i \in \hat{A}$. Since $\bar{\beta}$ is not unique, Theorem 2.1 implies there is a multiplier satisfying either $|\lambda_{j_i}| = 1$ for some $j_i \in \hat{A}_0$ or $\lambda_{j_i} = 0$ for some $j_i \in \hat{A}_0$. Then the direction $d = -\text{sign} (\lambda_{j_i}) \tilde{d}_{j_i}$ or $d = -\tilde{d}_{j_i}$ is a feasible direction for which the values of $S(\beta + \alpha d)$ are constant for all sufficiently small $\alpha$. With this change to step (3) of the RGA above, for each “constant” search direction we continue the algorithm to find a new optimal base point. Then, for each new point, we go to step (1) and continue, making sure that a new “constant” search direction $d$ is not simply the opposite of the previous search direction. In this way the algorithm will find all the optimal base points.

**Example 3.1** minimize $\ S(\beta) = |2\beta_1 + \beta_2 - 4| + | - \beta_1 + \beta_2 - 1| + |2\beta_1 - 5| + |\beta_1 + \beta_2 - 2|$. It is easy to check that $\beta = (2,0)^T$ is an optimal base point with $A = \{1,4\}$ and multipliers $(\lambda_1, \lambda_2) = (0,-1)$. Since $|\lambda_2| = 1$, the solution is not unique and the direction $d = -\text{sign} (\lambda_{j_2}) \tilde{d}_{j_2} = +(-1,2)^T$ is a “constant” search direction. Moving from $(2,0)^T$ along this direction yields the new optimal base point $(1,2)^T$ with active set $\{1,2\}$. The solution set is the closed line segment between $(2,0)^T$ and $(1,2)^T$. 

5
In some situations it may be desirable to single out just one point from an infinite solution set. A plausible criterion is to choose (if possible) the solution that has the best balance of its residuals in that the sum is closest to zero. In Example 3.1 the residuals at the two optimal base points are \( r_i((2,0)^T) = [0,-3,-1,0] \) and \( r_i((1,2)^T) = [0,0,-3,1] \), and for any optimal solution the residuals are \( r_i((2,0)^T + \alpha(1,2)^T) = [0,-3 + 3\alpha,-1 - 2\alpha,\alpha] \), \( 0 \leq \alpha \leq 1 \). So \((1,2)^T\) is the preferred optimal solution.

The second case to consider is where \( \mathbf{X} \) does not have full rank. If \( \text{rank}(\mathbf{X}) = r < p \), we can choose \( r \) linearly independent columns and store them as \( \mathbf{X}_1 \), and store the remaining columns as \( \mathbf{X}_2 \). Without loss of generality, reorder the columns of \( \mathbf{X} \) so that \( \mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2] \). The matrix \( \mathbf{X}_2 \) is related to \( \mathbf{X}_1 \) by \( \mathbf{X}_2 = \mathbf{X}_1 \mathbf{B} \) for some \( r \times (p-r) \) matrix \( \mathbf{B} \). Hence, if \( \mathbf{b}^T = [\mathbf{b}_1^T, \mathbf{b}_2^T] \), with \( \mathbf{b}_1 \in \mathbb{R}^r \), then

\[
\mathbf{X}\mathbf{b} = \mathbf{X}_1\mathbf{b}_1 + \mathbf{X}_2\mathbf{b}_2 = \mathbf{X}_1(\mathbf{b}_1 + \mathbf{B}\mathbf{b}_2).
\]

(3.1)

Suppose \( \hat{\mathbf{b}} \) is an optimal base point of (1.1) with active set \( \hat{\mathbf{A}}_1 \subseteq \mathbf{X}_1 \). Without loss of generality we can assume that \( \hat{A}_i = 0 \), \( i = r + 1, \ldots, p \), (since from (3.1), \( ([\hat{\mathbf{b}}_1 + \mathbf{B}\hat{\mathbf{b}}_2]^T, \mathbf{0}^T)^T \) also solves (1.1)). Then we can consider \( \hat{\mathbf{b}} \) as an optimal base point of (1.1) with active artificial constraints \( \hat{A}_i \leq 0 \), \( i = r + 1, \ldots, p \), i.e. \( \hat{\mathbf{A}}_0 = \{r + 1, \ldots, p\} \). With the active set \( \hat{\mathbf{A}} = \hat{\mathbf{A}}_1 \cup \mathbf{X}_1 \cup \hat{\mathbf{A}}_0 \), the base matrix and its inverse are of the form

\[
\hat{\mathbf{X}} = \begin{bmatrix}
\hat{\mathbf{X}}_1 \\
\hat{\mathbf{X}}_2
\end{bmatrix}
\quad \text{and} \quad
\hat{\mathbf{X}}^{-1} = \begin{bmatrix}
\hat{\mathbf{X}}_1^{-1} \\
\hat{\mathbf{X}}_2^{-1}
\end{bmatrix},
\]

(3.2)

where \( A_1 \) is \( r \times r \) and nonsingular, and \( I = I_{(p-r)} \). Consider the problem (1.1) obtained by setting the variables \( \hat{A}_i = 0 \), \( i = r + 1, \ldots, p \), in (1.1), so that the extended design matrix is \( \mathbf{X}_1 \). Since \( \mathbf{X}_1 \) has full rank, the solution set of (1.1) is the convex hull of its optimal base points as in Theorem 3.2.

The following result shows that the solution set \( \mathcal{M} \) of (1.1) is unbounded in this case.

**Theorem 3.3** Suppose \( \text{rank}(\mathbf{X}) = \text{rank}(\mathbf{X}_1) = r < p \) and let \( \mathbf{\theta} \) be any optimal base point of (1.1). Then \( \mathbf{\hat{\theta}} = [\mathbf{\theta}_1^T, \mathbf{0}^T]^T \) is an optimal base point of (1.1) with (after reordering) base matrix and inverse in (3.2). Moreover, for any \( \kappa_j \in \mathbb{R} \), the point

\[
\hat{\mathbf{\theta}} = [\mathbf{\theta}_1^T, \mathbf{0}^T]^T + \sum_{j \in \hat{\mathbf{A}}_0} \kappa_j \hat{\mathbf{d}}_j,
\]

(3.3)

where \( \hat{\mathbf{d}}_j \), \( j = r + 1, \ldots, p \), is the \( j \)th row of \( \hat{\mathbf{D}} \) in (3.2), is also a solution to (1.1). If (1.1) has a unique solution \( \hat{\mathbf{\theta}} \), then the solution set \( \mathcal{M} \) of (1.1) is the \( (p-r) \)-dimensional affine set defined by the right hand side of (3.3) with arbitrary \( \kappa_j \).

**Proof** As above, the matrix \( \mathbf{X}_2 \) is related to \( \mathbf{X}_1 \) by \( \mathbf{X}_2 = \mathbf{X}_1 \mathbf{B} \) for some \( r \times (p-r) \) matrix \( \mathbf{B} \). Clearly \( \hat{\mathbf{\theta}} \) is a feasible base point for (1.1), and it is optimal since, from (3.1), \( S(\mathbf{\hat{\theta}}) = S_1(\mathbf{\hat{\theta}}) \leq S_1(\mathbf{\hat{\theta}} + \mathbf{B\hat{\theta}}) = S(\mathbf{\theta}) \) for any \( \mathbf{\theta} \in \mathbb{R}^p \), where \( S_1 \) is the objective function for (1.1). Restricting attention to the first \( r \) rows (corresponding to the active set) in \( \mathbf{X}_2 = \mathbf{X}_1 \mathbf{B} \), clearly we have \( \hat{\mathbf{A}}_2 = \mathbf{B}^T \hat{\mathbf{A}}_1 \), so \( \mathbf{B}^T = \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_1^{-1} \). Then, from \( \hat{\mathbf{D}} \) in (3.2), \( [-\mathbf{B}^T I] = [-\mathbf{d}_{r+1}, \ldots, \mathbf{d}_p]^T \) and (3.3) can be rewritten as \( \hat{\mathbf{b}} = [\mathbf{\theta}_1^T, \mathbf{0}^T]^T + [-\mathbf{B}^T I]^T \mathbf{b}_2 \), where \( \mathbf{b}_2 = (\kappa_{r+1}, \ldots, \kappa_p)^T \). Since \( \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}\mathbf{b} \) (from \( \mathbf{X}_2 = \mathbf{X}_1 \mathbf{B} \)), it follows that \( \hat{\mathbf{\theta}} \) is a solution to (1.1) for arbitrary \( \mathbf{b}_2 \in \mathbb{R}^{p-r} \).

Now suppose \( \hat{\mathbf{\theta}} \) is the unique solution to (1.1) so \( \hat{\mathbf{\theta}} = [\mathbf{\theta}_1^T, \mathbf{0}^T]^T \) is an optimal base point of (1.1). If \( \hat{\mathbf{b}} = [\mathbf{b}_1^T, \mathbf{b}_2^T]^T \) is any optimal solution to (1.1), then \( \mathbf{b}_1 + \mathbf{B}\mathbf{b}_2 \) is feasible for (1.1) since \( \mathbf{X}_1(\mathbf{b}_1 + \mathbf{B}\mathbf{b}_2) = \mathbf{X}\hat{\mathbf{b}} \). Furthermore it is optimal for (1.1) since \( S_1(\mathbf{b}_1 + \mathbf{B}\mathbf{b}_2) = S(\mathbf{\hat{b}}) = S(\hat{\mathbf{\theta}}) \). Because \( \hat{\mathbf{\theta}} \) is unique, we must have \( \mathbf{b}_1 + \mathbf{B}\mathbf{b}_2 = \hat{\mathbf{\theta}} \) and so \( \hat{\mathbf{b}} = [\mathbf{\theta}_1^T, \mathbf{0}^T]^T + [-\mathbf{B}^T I]^T \mathbf{b}_2 \), which is of the form in (3.3). ■

An example of an unconstrained problem illustrating Theorem 3.3 is discussed in Shi and Lukas (2002) (Example 5.1 and Remark 5.1). Here we give a simple constrained example.
Example 3.2 Consider the $LL_1$ problem
\[
\begin{align*}
\text{minimize} & \quad S(\beta) = |\beta_1 + \beta_3| + |(\beta_1 + \beta_2 + \beta_3 + \beta_4)/2 - 1| + |(\beta_1 - \beta_2 + \beta_3 - \beta_4)/2 - 1| \\
\text{subject to} & \quad \beta_2 + \beta_4 \leq 0 \quad \text{and} \quad \beta_1 + \beta_3 \leq 1.
\end{align*}
\]
By writing down $X$ and letting $\overline{X}$ be the first two columns, it is obvious that $\text{rank}(X) = \text{rank}(\overline{X}) = 2$ here. It is not hard to confirm that $\hat{x} = 0$ is an optimal base point with the optimal value $S(\hat{x}) = 2$ and two artificial constraints $\beta_i \leq 0$, $i = 3, 4$. Its base matrix is $\hat{A} = [e_1, e_2, e_4, e_5]$ (where $e_i$ is the $i$th coordinate vector), the inverse base matrix is
\[
\hat{D} = [d_1, d_2, d_3, d_4] = [e_1, e_2, -e_1 + e_3, -e_2 + e_4]^T,
\]
and the corresponding multipliers are $(-1, 0, 0, 0)^T$. Hence $d = +\hat{d}_s$ and $d = -\hat{d}_t$ are “constant” search directions. First, considering the direction $-\hat{d}_t$, we can show that the points $\beta = \hat{x} + \kappa t (\hat{d}_s)$ are feasible and $S(\beta) = 0 + 1 + 1 = 2$ since $0 \leq \kappa t \leq 2$. Note, however, if $\kappa t > 2$, the point $\beta$ is not optimal. With the same reasoning, if $0 \leq \kappa t \leq 2$, then $(0, -\kappa t)^T$ is a solution to the corresponding problem defined by setting $\beta_3 = \beta_4 = 0$. In particular, with $\kappa t = 0$ or $\kappa t = 2$, respectively, $\bar{x} = (0, 0)^T$ or $\bar{x} = (0, -2)^T$ is an optimal solution and $\beta$ is of the form in (3.3).

Similarly, by considering the “constant” direction $\hat{d}_s$, we can show that the points $\beta = \hat{x} + \kappa s (\hat{d}_s)$ are solutions to the problem with $\beta_3 = \beta_4 = 0$. So $\bar{x}$ is also of the form in (3.3). By Lemma 3.1, any convex combination of $\bar{x}$ and $\beta$ is also a solution.

4. Perturbations to the constants of constraints

In this and the next section, we will analyse a perturbation to the constraints of the $LL_1$ problem. Let $\bar{x}$ be an optimal base point for (1.1) (not necessarily unique). Suppose that the $j$th constraint, $x_j^T \beta - y_j = 0$ (or $\leq 0$), is perturbed to $x_j^T \beta - (y_j + \Delta y_j) = 0$ (or $\leq 0$), where $\Delta y_j \neq 0$, and denote the new problem by (1.1)[\[. The effect of the perturbation depends on the type of constraint.

4.1. Equality constraints

In this case the point $\bar{x}$ is not feasible for the new problem (1.1)[\[. However, a certain move from $\bar{x}$ yields a new feasible and optimal point, as shown below.

**Theorem 4.1** If $\ell = j_0 \in E \subset \bar{A}$, let the new point be $\bar{x} = \bar{x} + |\Delta y_j| \bar{d}_j$, where $\bar{d}_j = \text{sign}(\Delta y_j) \bar{d}_j$. If the perturbation satisfies $|\Delta y_j| < \min\{\delta'(d), \alpha'(d)\}$ (see (2.7)), then $\bar{x}$ is an optimal base point for the new problem (1.1)[\[. The new optimal point and optimal value are, respectively,

\[
\bar{x} = \beta + |\Delta y_j| \bar{d}_j \quad \text{and} \quad S[\beta] = S(\bar{x}) + \Delta y_j \lambda_j.
\]

**Proof** By direct substitution, $\bar{x}$ satisfies the new equality constraint since $x_j^T \bar{d}_j = 1$. This, together with the condition $|\Delta y_j| < \min\{\delta'(d), \alpha'(d)\}$ and the fact that the direction $\bar{d}_j$ satisfies $x_j^T \bar{d} = 0$, $\forall i \in \bar{A} \setminus \{\ell\}$, ensures that $\bar{x}$ is a feasible base point. Moreover, the active set $\bar{A}$ for (1.1)[\[ at $\bar{x}$ is the same as $\bar{A}$ and hence the inverse base matrix is not changed. Since the move does not encounter any boundary planes of inactive equations, the “cost” vector and multipliers at $\bar{x}$ for (1.1)[\[ remain the same as those at $\beta$ for (1.1). Thus $\bar{x}$ is an optimal point for (1.1)[\[. Since $|\Delta y_j| < \delta'(d)$ and $x_j^T \bar{d} = 0$, $\forall i \in \bar{A}_s$, from (2.9) and (2.4) we obtain the new optimal value

\[
S[\beta] = S(\beta) + |\Delta y_j| (\bar{c}^T \bar{d}_j + \sum_{i \in \bar{A}_s} |x_i^T \bar{d}_j|) = S(\beta) + |\Delta y_j| \text{sign}(\Delta y_j) \lambda_j = S(\bar{x}) + \Delta y_j \lambda_j.
\]
which completes the proof. ■

**Remark 4.1** If the base matrix at \( \bar{\beta} \) is ill-conditioned, then the change \( \bar{\beta} - \beta = \Delta y_i \bar{d}_i \) may be much larger than \( \Delta y_i \) in magnitude, since \( ||\bar{d}_i|| \) could be large. Similarly, the bound on \( \Delta y_i \) may be small, since \( |x_i^T \bar{d}_i| \) in (2.7) could be large. This highly sensitive behaviour is the same as occurs for perturbations to the responses \( y_i \), as discussed in Shi and Lukas (2005).

### 4.2. Inactive inequality constraints

For a perturbation to an inactive inequality constraint, we have the following result.

**Theorem 4.2** If \( l \in I \setminus \bar{J}_B \), then \( \bar{\beta} \) is also an optimal base point for (1.1)<\( l \) provided \( \Delta y_l \in [\gamma_l(\bar{\beta}), \infty) \).

**Proof** Under the condition on \( \Delta y_l \) we have \( r_l(\bar{\beta}) \equiv x_i^T \bar{\beta} - (y_l + \Delta y_l) = r_l(\bar{\beta}) - \Delta y_l \leq 0 \). Hence \( \bar{\beta} \) is still feasible and therefore it is an optimal base point for (1.1)<\( l \). ■

### 4.3. Active inequality constraints

For a perturbation to an active inequality constraint, the situation is similar to that for an equality constraint, as the next result shows.

**Theorem 4.3** Let \( l = j_q \in \bar{J}_B \) and let \( \bar{d}_q = -\bar{d}_q \). If the perturbation \( \Delta y_q \) satisfies \( |\Delta y_q| < \min\{\delta'(\bar{d}), \alpha'(\bar{d})\} \) (see (2.7)), then the new base point \( \bar{\beta} = \bar{\beta} - \Delta y_q \bar{d} = \bar{\beta} + \Delta y_q \bar{d}_q \) is a solution to (1.1)<\( l \) with the same optimal value as shown in (4.1).

**Proof** The proof is similar to that of Theorem 4.1. ■

### 5. Perturbations to the coefficients of constraints

Now we study the effect on the optimal solution of a perturbation to the gradient (coefficient vector) of a constraint. Suppose that \( \bar{\beta} \) is the unique solution to the problem (1.1) and the \( k \)th constraint \( x_i^T \bar{\beta} - y_k = 0 \) (or \( \leq 0 \)) is perturbed into \( (x_k + \Delta x_k)^T \bar{\beta} - y_k = 0 \) (or \( \leq 0 \)). The perturbed problem is denoted by (1.1)<\( k \).

#### 5.1. Equality constraints

When the gradient \( x_k \) is active (i.e. for an equality constraint or active inequality constraint), the perturbed base matrix is not the same as the original one and we need to ensure that it is nonsingular. Clearly, 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 5.1** Let the matrix \( \bar{A} \) be formed by replacing the \( q \)th column \( x_k \), \( k = j_q \), of the nonsingular matrix \( \tilde{A} \) by \( x_k + \Delta x_k \). Then \( \bar{A} \) is still nonsingular provided \( \|\Delta x_k\| < 1/||\bar{d}_q|| \). Furthermore, the \( i \)th row of the inverse \( \bar{D} = \bar{A}^{-1} \) is equal to the transpose of

\[
\bar{d}_i = \begin{cases} 
\frac{1}{1 + \Delta x_i^T \bar{d}_i} \bar{d}_i, & i = q, \\
\bar{d}_i - (\Delta x_i^T \bar{d}_i) \bar{d}_q, & i \neq q.
\end{cases}
\]  

(5.1)

See Shi and Lukas (2005) for the proof. The following result and its proof are similar to those for perturbations to the gradients of equations in Shi and Lukas (2005). But here the direction is not necessarily descent, and the optimality conditions have no limitation on the multipliers corresponding to equality constraints.

**Theorem 5.1** Let \( k = j_k \in E \subset \bar{A} \) and suppose \( r_k^{(E)}(\bar{\beta}) \equiv (x_k + \Delta x_k)^T \bar{\beta} - y_k = \Delta x_k^T \bar{\beta} \neq 0 \). (In the case \( \Delta x_k^T \bar{\beta} = 0 \) including \( \bar{\beta} = 0 \), the point \( \bar{\beta} \) remains the solution to the new problem.) Let \( \delta' \) and \( \alpha' \) be the shortest steps to the boundary planes corresponding to the inactive equations and inactive constraints, respectively (see (2.7)), from \( \bar{\beta} \) along the direction
\[ \mathbf{d} = -\theta_k \tilde{d}_i, \quad \theta_k \equiv \text{sign}(\Delta x_k^T \hat{\beta}). \] (5.2)

If

\[ \|\Delta x_k\| < \eta \equiv \min\{\eta_0, \eta_1, \eta_2\}, \] (5.3A)

\[ \eta_0 \equiv \min\left\{ \frac{1}{\|d_i\|}, \delta', \frac{\alpha'}{\|\beta\| + \delta'\|d_i\|}, \frac{\alpha'}{\|\beta\| + \|\alpha'd_i\|} \right\}, \] (5.3B)

\[ \eta_1 \equiv \min_{j \in \tilde{A}_t} \frac{1 - |\lambda_i|}{\lambda_i \cdot \|d_i\| + (1 - |\lambda_i|)\|\tilde{d}_i\|}, \] (5.3C)

\[ \eta_2 \equiv \min_{j \in \tilde{A}_t} \frac{|\lambda_i|}{\lambda_i \cdot \|d_i\| + |\lambda_i| \cdot \|\tilde{d}_i\|}, \] (5.3D)

then

\[ \hat{\beta} = \beta + \alpha^{<k>} \mathbf{d}, \quad \alpha^{<k>} = \frac{\|\Delta x_k^T \hat{\beta}\|}{1 + \Delta x_k^T \tilde{d}_i} \] (5.4)

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

\[ S^{<k>}(\beta) = S(\beta) - \frac{\langle \Delta x_k^T \hat{\beta}, \lambda_i \rangle}{1 + \Delta x_k^T \tilde{d}_i}. \]

**Proof** First we show that \(\beta\), defined by (5.4), is feasible for \((1.1)^{<k>}\). For the new equality constraint, (5.4) and (5.2) imply that

\[ r_i^{<k>}(\beta) = (x_k + \Delta x_k)^T (\beta - \theta_k \alpha^{<k>} \mathbf{d}_i) - y_k = r_i(\beta) + \Delta x_k^T \hat{\beta} - \theta_k \alpha^{<k>}(x_k + \Delta x_k)^T \mathbf{d}_i \]

\[ = 0 + \Delta x_k^T \hat{\beta} - \theta_k \alpha^{<k>}(1 + \Delta x_k^T \mathbf{d}_i) = 0. \]

For the other equality and inequality constraints at \(\beta\), we have

\[ r_i^{<k>}(\beta) = x_i^T (\beta + \alpha^{<k>} \mathbf{d}) - y_i = r_i(\beta) + \alpha^{<k>} x_i^T \mathbf{d}. \] (5.5)

For all \(i \in \mathcal{E} \cup \tilde{A}_t \setminus \{k\}\), it follows from (5.5) and \(x_i^T \mathbf{d} = 0\) that \(r_i^{<k>}(\beta) = 0\). For all inactive (inequality) constraints at \(\beta\), from \(\|\Delta x_k\| < \eta_0\) in (5.3), we have \(1 + \Delta x_k \mathbf{d}_i \geq 1 - \|\Delta x_k\| \cdot \|\tilde{d}_i\| > 0\), and then

\[ 0 < \alpha^{<k>} \leq \frac{\|\Delta x_k\| \cdot \|\hat{\beta}\|}{(1 - \|\Delta x_k\| \cdot \|\tilde{d}_i\|)} < \alpha'. \]

Hence, if \(x_i^T \mathbf{d} > 0\), then \(r_i^{<k>}(\beta) < r_i(\beta) + \alpha' x_i^T \mathbf{d} \leq 0\) by (5.5) and (2.7), and, if \(x_i^T \mathbf{d} \leq 0\), then obviously \(r_i^{<k>}(\beta) < r_i(\beta) < 0\). Therefore \(\beta\) is feasible for \((1.1)^{<k>}\). The above argument also shows that \(\beta\) is a base point of \((1.1)^{<k>}\) with the same active set as \(\beta\) for \((1.1)\), since the perturbation does not affect any equation.

Now we show that \(\hat{\beta}\) is optimal for \((1.1)^{<k>}\). From above \(\alpha^{<k>} < \alpha'\), and similarly we obtain \(\alpha^{<k>} < \delta'\). Hence the move from \(\hat{\beta}\) along \(\mathbf{d}\) with step length \(\alpha^{<k>}\) does not cross any boundary planes associated with the inactive equations, so the “cost” vector for \((1.1)^{<k>}\) is \(\bar{c} \equiv \bar{c}(\beta) = \bar{c}\). But the base matrix at \(\hat{\beta}\) is \(\tilde{A}\) as defined in Proposition 5.1, since \(k = j_\beta \in \tilde{A}\), the active set at \(\hat{\beta}\) for \((1.1)^{<k>}\). From (5.3) and Proposition 5.1, \(\tilde{A}\) is nonsingular, and its inverse is given by (5.1). Consequently, the multipliers at \(\hat{\beta}\) for \((1.1)^{<k>}\) are

\[ \bar{\lambda}_i = \bar{x}_i^T \bar{c} = \bar{x}_i^T \bar{c} = \begin{cases} \frac{\bar{\lambda}_i}{1 + \Delta x_k^T \mathbf{d}_i}, & i = q (k = j_\beta), \\ \frac{\bar{\lambda}_i}{1 + \Delta x_k^T \mathbf{d}_i} - \frac{(\Delta x_k^T \mathbf{d}_i) \bar{\lambda}_i}{1 + \Delta x_k^T \mathbf{d}_i}, & i \neq q. \end{cases} \]
Now we estimate the values of the multipliers. We do not need to check the value of \( \lambda_j \) corresponding to an equality constraint. For the other multipliers, since \( \| \Delta x_k \| < \eta_1 \) in (5.3),

\[
|\lambda_i| \leq |\lambda_i| + \frac{|\lambda_j| \cdot \| \Delta d_i \| \cdot \| \Delta x_k \|}{1 - \| \Delta d_i \| \cdot \| \Delta x_k \|} < 1, \quad j_i \in \bar{A}_s = \bar{A}_s,
\]

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

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

Hence \( \bar{\beta} \) is the unique optimal point for (1.1)\(^{<k>}\).

Finally, since the active set is unchanged and \( \text{sign}(r_i^{<k>}(\bar{b})) = \text{sign}(r_i(\bar{\beta})) \equiv \bar{\sigma}_i, \quad i \in \bar{A}_s \), the new optimal value is

\[
S^{<k>}(\bar{\beta}) = (\sum_{i \in \bar{A}_s} + \sum_{i \in \bar{A}_s^c})(|x_i^T (\bar{\beta} - \theta_k \alpha^{<k>} \bar{d}_i) - y_i|)
\]

\[
= \sum_{i \in \bar{A}_s} |0 - \theta_k \alpha^{<k>} x_i^T \bar{d}_i| + \sum_{i \in \bar{A}_s^c} \bar{\sigma}_i (x_i^T \bar{\beta} - y_i) + \sum_{i \in \bar{A}_s^c} (-\theta_k \bar{\sigma}_i \alpha^{<k>} x_i^T \bar{d}_i)
\]

\[
= 0 + S(\bar{\beta}) - \theta_k \alpha^{<k>} \bar{c}^T \bar{d}_i = S(\bar{\beta}) - \frac{(\Delta x_k^T \bar{\beta}) \lambda_i}{1 + \Delta x_k^T \bar{d}_i}
\]

which completes the proof. ■

**Remark 5.1** This result indicates there is potentially high sensitivity if either the base matrix at \( \bar{\beta} \) is ill-conditioned or if \( \bar{\beta} \) is close to being non-unique (with \( |\lambda_i| \approx 1, \quad j_i \in \bar{A}_s \), or \( \lambda_i \approx 0, \quad j_i \in \bar{A}_i \)). In these cases the bound on \( \| \Delta x_k \| \) in (5.3) may be small while \( \| \bar{\beta} - \bar{\beta} \| \) could be much larger. These highly sensitive cases are the same as those for perturbations to the row vectors \( x_i^T \) of the design matrix, as discussed in Shi and Lukas (2005).

5.2. **Inactive inequality constraints**

**Theorem 5.2** Let \( k \in I \setminus \bar{A}_i \) so \( r_k(\bar{\beta}) < 0 \). Then \( \bar{\beta} \) is also an optimal base point for (1.1)\(^{<k>}\) provided \( \Delta x_k^T \bar{\beta} \leq -r_k(\bar{\beta}) \). **Proof** With the condition on \( \Delta x_k \), we have \( r_k^{<k>}(\bar{\beta}) = r_k(\bar{\beta}) + \Delta x_k^T \bar{\beta} \leq 0 \). Hence \( \bar{\beta} \) is still feasible and the result follows. ■

5.3. **Active inequality constraints**

**Theorem 5.3** Let \( k = j_0 \in \bar{A}_i \) and suppose \( r_k^{<k>}(\bar{\beta}) = \Delta x_k^T \bar{\beta} \neq 0 \). If \( \| \Delta x_k \| \) satisfies (5.3), then the conclusion of Theorem 5.1 is also valid here.

**Proof** The proof of this case is similar to that of Theorem 5.1. In fact, regarding this active inequality constraint as an equality constraint of (1.1) does not affect the solution of (1.1). So, with the same arguments as those for Theorem 5.1, we derive the same conclusion. ■

6. **Deletion of an observation**

In assessing an \( L_1 \) regression estimate, the effect of deleting each observation provides very useful diagnostic information. It identifies the most influential observations and others that have little or no effect on the solution. The effect of deleting one or more observations is also central in cross-validation for model selection, and in jackknife estimation of variance (for \( L_1 \) regression, a delete-\( d \) jackknife should be used since the estimator is not smooth (see Efron and Tibshirani (1993), Sec. 11.6 and 11.7)).
In this section, let \( (1.1)^{(-t)} \) denote the problem obtained by deleting the \( t \)-th observation from (1.1). Let \( \tilde{\beta} \) be an optimal base point for (1.1) (not necessarily unique).

Since deletion (and addition) of an observation is a discrete change to the problem, there can be a sizeable change in the optimal solution. There may be no change to the active set, a change of one active equation or constraint or perhaps even several. For this reason it is hard to find a general estimate or bound on the change in the solution. However, it is possible to identify cases for which there will be no change in the solution. Also, in general, one can use the RGA algorithm starting at \( \beta \) to find the new solution, which is far more efficient than solving the new problem from the beginning.

6.1. Inactive observation

First we consider the case where the deleted observation (equation) is inactive.

**Theorem 6.1** If \( t \in A'_s \), then \( \tilde{\beta} \) remains a solution of (1.1)\(^{(-t)}\) if and only if

\[
\begin{align*}
\lambda_i - \sigma_i \tilde{d}_i^T x_i & \leq 1, \quad \text{if } j_i \in \bar{A}_s, \\
\lambda_i - \sigma_i \tilde{d}_i^T x_i & \leq 0, \quad \text{if } j_i \in \bar{A}_t, \\
\lambda_i - \sigma_i \tilde{d}_i^T x_i & = 0, \quad \text{if } j_i \in \bar{A}_o,
\end{align*}
\]

(6.1)

where \( \sigma_i = \text{sign}(r_i(\tilde{\beta})) \).

**Proof** In this case, from (2.3A) the new “cost” vector is

\[
c^{(t)} = \tilde{c} - \sigma_i x_i,
\]

and the inverse of the base matrix is not changed. Hence the new multipliers for \( (1.1)^{(-t)} \) at \( \tilde{\beta} \) are

\[
\begin{align*}
\lambda_i^{(-t)} = \tilde{d}_i^T c^{(t)} &= \tilde{d}_i^T \tilde{c} - \sigma_i \tilde{d}_i^T x_i = \lambda_i - \sigma_i \tilde{d}_i^T x_i.
\end{align*}
\]

(6.3)

Comparing (6.3) with (2.5) shows that (6.1) is a necessary and sufficient condition for \( \tilde{\beta} \) to remain a solution of \( (1.1)^{(-t)} \).

6.2. Active observation

If the deleted equation is active, we have the following.

**Theorem 6.2** Suppose \( t = j_q \in \bar{A}_s \). Then \( \tilde{\beta} \) remains a solution for \( (1.1)^{(-t)} \) if and only if the corresponding multiplier \( \lambda_q = 0 \). If \( \lambda_q \neq 0 \), the direction

\[
d = -\text{sign}(\lambda_q) \tilde{d}_q
\]

(6.4)

is a feasible descent direction. Furthermore, if index \( p \) is the new element of the active set \( A^{(-t)} \) at the new (feasible) base point \( \tilde{\beta}^{(-t)} = \tilde{\beta} + \alpha d \) obtained by a line search from \( \tilde{\beta} \) along \( d \) (with \( \alpha \) chosen as in step (4) of the RGA in Section 2), the new multipliers are

\[
\lambda_i^{(-t)} = \begin{cases} \frac{\lambda_i}{x_i^T \tilde{d}_q}, & i = q, \\ \lambda_i + (x_i^T \tilde{d}_q / x_q^T \tilde{d}_q) \lambda_q, & i \neq q. \end{cases}
\]

(6.5)

So \( \beta^{(-t)} \) is a solution to \( (1.1)^{(-t)} \) provided all \( \lambda_i^{(-t)} \) satisfy (2.5).

**Proof** In this case, we associate the index \( t \) with an artificial constraint \( x_i^T \tilde{\beta} - y_i \leq 0 \) for \( (1.1)^{(-t)} \), which is active at \( \tilde{\beta} \). With this point of view, \( \tilde{\beta} \) is also a base point of \( (1.1)^{(-t)} \) with the same inverse base matrix \( \tilde{D} \), and the “cost” vector of \( (1.1)^{(-t)} \), \( c^{(-t)} \), is the same as the original one, i.e. \( c^{(-t)} = \tilde{c} \). Hence, \( \lambda_i^{(-t)} = \lambda_i \), \( i = 1, \ldots, n \). Since the \( t \)-th equation for \( (1.1)^{(-t)} \) is artificial, \( \tilde{\beta} \) remains a solution for \( (1.1)^{(-t)} \) if and only if the corresponding multiplier \( \lambda_q^{(-t)} = \lambda_q = 0 \).
If \( \lambda \neq 0 \), the direction defined in (6.4) is a feasible descent direction according to (2.6). Furthermore, from (2.4) and (2.8), the new multipliers are defined by (6.5) and the result follows. □

**Example 6.1** (Shi (1997) Example 2.4.1 and Example 7.2.1) Consider the \( LL_1 \) problem:

\[
\begin{align*}
\text{minimize} & \quad S(\beta) = |2\beta_1 + \beta_2 - 4| + | - \beta_1 + \beta_2 - 1 | + | \beta_1 - 3 | + | \beta_1 + \beta_2 - 2 | + | \beta_1 - 4 | \\
\text{subject to} & \quad -\beta_1 + 2\beta_2 - 2 \leq 0, \quad -\frac{1}{2}\beta_1 - \beta_2 \leq 0, \quad \beta_1 \geq -1, \quad \beta_2 \geq -2.5.
\end{align*}
\]

It is easy to check that an optimal solution, its active set and multipliers are \( \bar{\beta} = (2, 0)^T, A = \{1, 4\} \) and \( (\lambda_1, \lambda_2) = (0, -1) \), respectively. The point \( \bar{\beta} \) remains optimal for (1.1)(-1) since \( \lambda_1^{(-1)} = \lambda_1 = 0 \), but it is not a solution of (1.1)(-4) since \( \lambda_2^{(-4)} = \bar{\lambda}_2 = -1 \neq 0 \). It is easy to solve the latter problem from the beginning using the RGA. However, it is more efficient to start from \( \bar{\beta} \). From (6.4), a descent direction at \( \bar{\beta} \) for (1.1)(-4) is \( d = \bar{d}_2 = (-1, 2)^T \). The search along the edge of \( P_1 \) will reach the optimal point \( \beta^{(-4)} = (1.2, 1.6)^T \) with active set \( A^{(-4)} = \{1, 6\} \) and multipliers \( \lambda_1^{(-4)} = -3/5, \lambda_2^{(-4)} = -1/5 \).

**Remark 6.1** This example and Theorems 6.1 and 6.2 show that if \( \bar{\beta} \) is not a solution of (1.1)(-1), we can use \( \beta \) as a starting base point for the RGA with the feasible descent direction defined by (6.4) if \( t \in A^5 \) and by (2.6) and (6.3) if \( t \in A^6 \). This is more efficient than solving (1.1)(-1) from the very beginning.

For the model \( y = X\beta + \varepsilon \), where \( X \) is \( n \times p \), the influence of the \( t \)-th observation on a regression estimate \( \hat{\beta} \) can be measured by \( V_t \equiv ||X\hat{\beta} - X\beta||_2 \), where \( \beta ||_2 \) is the regression estimate obtained with the \( t \)-th observation deleted. Apart from a scale factor, \( V_t \) is the Cook distance, which was introduced for least squares regression by Cook (1977) (see also Cook and Weisberg (1982)). In the least squares case, there is a formula for \( V_t \), which means one doesn’t have to solve \( n \) additional regression problems to find \( V_t \), \( t = 1, \ldots, n \).

For \( L_1 \) regression, although there is no simple formula for \( V_t \), it can still be computed efficiently by computing \( \beta^{(-t)} \) using \( \beta \) as a starting point. For an unconstrained \( L_1 \) regression problem, Castillo et al. (2001) computed \( V_t \) using the LP formulation and a variant of the dual simplex method (called the Exterior Point Method) to compute \( \beta^{(-t)} \) starting from \( \beta \). For the general \( LL_1 \) problem (1.1), it is clear from Remark 6.1 that the RGA can be used to efficiently compute \( V_t \).

### 7. Addition of an Observation

Suppose we have a new observation \( y_t = x_t^T\beta + \varepsilon_t \), and the term \( |x_t^T\beta - y_t| \) is added to \( S(\beta) \) in (1.1A) forming a new problem (1.1)(+1). Let \( r_t(\beta) = x_t^T\beta - y_t \) be the new residual.

First we show that if the new observation has a certain property then an infinite solution set becomes smaller.

**Theorem 7.1** Let \( \mathcal{M} \) be the solution set of the original problem and suppose the new observation divides \( \mathcal{M} \), i.e. there are two points \( \beta \) and \( \beta \) of \( \mathcal{M} \) such that \( r_t(\beta) \) and \( r_t(\beta) \) are of opposite sign, then the solution set of (1.1)(+1) is \( \mathcal{M} \cap P_t \), where \( P_t = \{\beta| x_t^T\beta - y_t = 0\} \) is the new boundary plane.

**Proof** Let \( S^{(+1)}(\beta) = S(\beta) + |r_t(\beta)| \) be the new sum of absolute deviations at \( \beta \). Clearly, for any \( \beta \in \mathcal{M} \setminus P_t \), there is a \( \bar{\beta} \in \mathcal{M} \setminus P_t \) for which the corresponding residuals have opposite signs. Since \( \mathcal{M} \) is convex, the line between \( \beta \) and \( \bar{\beta} \) lies in \( \mathcal{M} \) and intersects \( P_t \) at some point \( \beta^* \), and \( S^{(+1)}(\beta^*) = S(\beta^*) = S(\beta) < S^{(+1)}(\beta) \). Also, for any \( \beta \notin \mathcal{M} \), we have \( S^{(+1)}(\beta^*) = S(\beta) \leq S^{(+1)}(\beta) \). Since \( S^{(+1)}(\beta) = S^{(+1)}(\beta^*) \) for all \( \beta \in \mathcal{M} \cap P_t \), the result follows. □

Now we derive conditions under which a solution remains optimal. Let \( \bar{\beta} \) be an optimal base point for (1.1) (not necessarily unique).
Theorem 7.2  Suppose we have a new observation \( y_t = x_t^T \beta + \varepsilon_t \), giving a new problem \((1.1)^{++}\). If \( r_t(\beta) = 0 \), then \( \beta \) remains a solution of \((1.1)^{++}\). If \( r_t(\beta) \neq 0 \), with \( \sigma_t = \text{sign}(r_t(\beta)) \), then \( \beta \) remains a solution of \((1.1)^{++}\) if and only if

\[
\begin{align*}
|\lambda_i + \sigma_t d_t^T x_i| &\leq 1, \quad \text{if } j_i \in \tilde{A}_s, \\
\lambda_j + \sigma_t d_t^T x_j &\leq 0, \quad \text{if } j_i \in \tilde{A}_f, \\
\lambda_i + \sigma_t d_t^T x_i & = 0, \quad \text{if } j_i \in \tilde{A}_o.
\end{align*}
\]  

(7.1)

Proof  In this case \( t \in \tilde{A}_s \) and the new “cost” vector is \( c^{++} = \bar{c} + \sigma_t x_i \), which is different from \( c^{(-)} \) in (6.2) only in the middle operation. The result follows by the same argument as for Theorem 6.1. \( \blacksquare \)

This result demonstrates the robustness of \( L_1 \) regression with respect to additional data values, since the condition (7.1) depends on \( y_t \) only through the sign of the residual \( r_t(\beta) \). As a special case, consider an unconstrained problem with solution \( \beta \) and a new observation \( y_t = x_t^T \beta + \varepsilon_t \), where \( x_t \) is the same as an active gradient vector \( x_{j_0}, j_0 \in \tilde{A} \). Then, since \( d_t^T x_t = \delta_{i_0} \), the condition (7.1) simplifies to \( |\lambda_{i_0} + \sigma_t| \leq 1 \). Hence \( \beta \) remains a solution if and only if \( \lambda_{i_0} = 0 \) or \( \lambda_{i_0} \) and \( r_t(\beta) \) have opposite signs (independent of the magnitude of \( y_t \)).

Example 7.1  Consider \( S(\beta) \) in Example 3.1 and add the term \( |2\beta_1 + \beta_2 - y_t| \). Then \( x_t = x_1 \), which is active for both optimal base points \((2,0)^T \) and \((1,2)^T \) in Example 3.1. For \((2,0)^T \) the corresponding multiplier is \( \lambda_1 = 0 \) so \((2,0)^T \) remains a solution of the new problem for arbitrary values of \( y_t \). The same is true for \((1,2)^T \), and so the whole solution set remains optimal for the new problem. (In fact, the new problem has a larger solution set, unlike the situation in Theorem 7.1.)

8. Numerical results

Example 8.1  Derivative estimation under a concavity constraint

We will consider the derivative estimation problem discussed in Shi and Lukas (2005). We wish to estimate the derivative of a function \( g(t), 0 \leq t \leq 1 \), satisfying \( g(0) = 0 \), given discrete noisy data \( y_i = g(t_i) + \varepsilon_i, \ i = 1, \ldots, n \). This problem arises in many applications, including the analysis of growth curves (Gasser et al. (1984) and Eubank (1988)) and pharmacokinetic response curves (Song et al. (1995)). The estimated derivative can be used to determine characteristic time points like the point of maximum growth rate or the point of peak response.

Numerical differentiation is an ill-posed problem (see Anderssen and Bloomfield (1974)) and requires some stabilization to obtain reasonable solutions. A simple approach to this is to approximate \( g'(t) \) in a low dimensional subspace (by some convenient parametric form), where the dimension plays the role of a smoothing parameter.

Clearly the true solution \( f(t) = g'(t) \) satisfies \( \int_0^t f(T) dT = g(t) \), since \( g(0) = 0 \). As in Shi and Lukas (2005) we approximate \( f(t) \) by the truncated trigonometric series

\[
f_p(t) = c_0 + \sum_{k=1}^{p_1} c_k \cos 2k\pi t + \sum_{k=1}^{p_2} d_k \sin 2k\pi t, \quad (8.1)
\]

where \( p = p_1 + p_2 \ll n \), and evaluate the indefinite integral at the points \( t_i \), giving the overdetermined system

\[
\int_0^{t_i} f_p(T) dT = c_0 t_i + \sum_{k=1}^{p_1} c_k \sin 2k\pi t_i/(2k\pi) + \sum_{k=1}^{p_2} d_k (1 - \cos 2k\pi t_i)/(2k\pi) \approx y_i, \quad i = 1, \ldots, n. \quad (8.2)
\]
Applying $L_1$ regression, we have a problem of the form of (1.1A) with $\mathbf{b} = (c_0, c_1, \ldots, c_p, d_1, \ldots, d_p)^T$.

In many situations there is additional information about the unknown true derivative. In the case of child growth curves, obviously the growth rate (derivative) is non-negative for all time. In the case of pharmacokinetic response curves, the underlying reaction kinetics model determines properties of the reaction rate $r = dc/dt$; e.g., typically the (negative) rate increases monotonically after a certain time, so $dr/dt > 0$. It is worth trying to incorporate such extra information into the estimation procedure above.

For our simulations we will assume it is known that the true derivative $f(t) = g'(t)$ is concave, i.e. $f''(t) \leq 0$ for all $t$. A concrete example is the estimation of the speed of a vehicle from time and position data, given that the acceleration is decreasing. To use this information, we evaluate $f''_p(t) \leq 0$ at finitely many values of $t$, resulting in constraints of the form in (1.1C). Thus we obtain a constrained $L_1$ regression problem.

For the numerical experiments, we take $g(t) = 10(-t^3/3 + t^2/2)$ and so $g'(t) = 10(-t^2 + t)$. The data are $y_i = g(t_i) + \varepsilon_i$, $i = 1, \ldots, n$, where $n = 50$, $t_i = i/n$ and the $\varepsilon_i$ are pseudo-random numbers distributed normally with mean 0 and standard deviation 0.01, i.e. approximately 1% noise. The $L_1$ estimate of $g'(t)$ is computed by solving the constrained $L_1$ regression problem defined above with $p_1 = p_2 = 5$ so $p = 11$. 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 as given in Barrodale and Roberts (1978). The results were confirmed using the Fortran 77 routine for the RGA described in Shi and Lukas (2002).

The function $g(t)$ and data are shown in Fig. 1(a). Fig. 1(b) shows the derivative $g'(t)$ and the unconstrained $L_1$ regression estimate of $g'(t)$.

![Graph](image_url)

**Fig. 1(a)** $g(t) = 10(-t^3/3 + t^2/2)$ (solid), and data (+)
**Fig. 1(b)** $g'(t) = 10(-t^2 + t)$ (solid), and unconstrained $L_1$ regression estimate (dashed)

Now we use the information that $f(t) = g'(t)$ is concave. With 5 constraints $f''_p(T_i) \leq 0$, where $T_i = 0.4 + 0.05(i - 1)$, $i = 1, 2, \ldots, 5$, we obtain the $L_1$ estimate of $g'(t)$ shown in Fig. 2(a). Clearly, the constrained estimate is better than the unconstrained one in Fig. 1(b). For the constrained estimate, two of the five constraints are active, i.e. $f''_p(0.4) = 0$ and $f''_p(0.55) = 0$, as shown by the two + marks on the graph of $f''_p(t)$ in Fig. 2(b).
Fig. 2(a) \( g'(t) = 10(-t^2 + t) \) (solid), and constrained \( L_1 \) regression estimate \( f_p(t) \) (dashed). Fig. 2(b) \( f_p''(t) \) for constraint points 0.4, 0.45, …, 0.6

Let \( \tilde{\beta} \) be the optimal solution of the constrained \( L_1 \) regression problem defining the coefficients of \( f_p(t) \). In the following subsections we will consider different types of perturbations to this \( L_1 \) regression problem, possibly resulting in a new optimal solution \( \tilde{\beta} \). If \( \tilde{f}_p(t) \) and \( \tilde{T}_p(t) \) are the estimates of \( g'(t) \) corresponding to \( \beta \) and \( \tilde{\beta} \), respectively, then it is not hard to show from (8.1) that \( ||f_p - \tilde{T}_p||_\infty \leq ||\beta - \tilde{\beta}||_1 \). Therefore, a bound on \( ||\beta - \tilde{\beta}||_1 \) will also serve as a bound on the change in the estimate of \( g'(t) \).

8.1. Perturbations to responses

First we investigate the sensitivity of the constrained \( L_1 \) solution \( \tilde{\beta} \) for perturbations to the response vector \( y \). This case is covered by Theorem 3.1 of Shi and Lukas (2005). In simplified form, this states that for a sufficiently small perturbation vector \( \Delta y \), the solution \( \tilde{\beta} \) of the perturbed problem satisfies

\[
||\tilde{\beta} - \beta||_1 \leq p||D||_\infty ||\Delta y||_\infty.
\]  

(8.3)

This result was illustrated in Shi and Lukas (2005) by simulations with the unconstrained \( L_1 \) regression estimate of \( g'(t) \). Here we illustrate it with the constrained \( L_1 \) regression estimate. As in Shi and Lukas (2005), the simulations involved taking 40 random perturbation vectors \( \Delta y \) with components \( \Delta y_i = \pm 10^{-13} \), where the sign is pseudo-random with probability 0.5. To see the effect of the original data set, this was repeated 10 times with different data vectors \( y \) generated by taking different sets of pseudo-random errors \( \varepsilon_i \) (all normally distributed with mean 0 and standard deviation 0.01). The size of the perturbations was chosen to be \( 10^{-13} \) to ensure that the conditions of Theorem 3.1 of Shi and Lukas (2005) hold for all the problems in the simulation. The resulting \( L_1 \) norm \( ||\tilde{\beta} - \beta||_1 \) of errors in the solution and the bound in (8.3) are plotted in Fig. 3. Clearly, (as in the unconstrained case in Shi and Lukas (2005)) the bound is quite good and follows the behaviour of the errors over the 10 replicates.
8.2. Perturbations to constants of constraints

For the optimal solution $\bar{\beta}$ to the constrained $L_1$ regression problem above there are 2 active constraints, corresponding to $f_p''(0.4) = 0$ and $f_p''(0.55) = 0$. To investigate the sensitivity of this solution and illustrate Theorem 4.3, we perturb one inequality constraint to become $f_p''(0.55) \leq -0.001$ and compute the new solution $\bar{\beta}$ and the error $||\bar{\beta} - \beta||_1$. This is repeated for 40 replicates of the data vector $y$, and the resulting errors (for the 19 replicates in which this constraint was active) are plotted in Fig. 4. Clearly, the solution is not sensitive to this perturbation. These results are consistent with Theorem 4.3; for each replicate the perturbation $\Delta y_i = -0.001$ satisfies the condition in Theorem 4.3 and so the error is simply $||\bar{\beta} - \beta||_1 = 0.001||d_i||_1$, where $l = j_i$.

8.3. Perturbations to coefficients of constraints

For the optimal solution $\bar{\beta}$ to the constrained $L_1$ regression problem above, 3 of the 5 constraints are inactive. In particular, the last constraint, i.e. $f_p''(0.6) \leq 0$, is inactive, with the left hand side equal to $r_k(\bar{\beta}) = x_k^T\bar{\beta} - y_k = -36.448$, where $k = 55$.

From Theorem 5.2, for a perturbation $\Delta x_k^T$ of $x_k^T$ in an inactive constraint, the solution $\bar{\beta}$ remains optimal if the perturbation satisfies $r_k^{<k>}(\bar{\beta}) = r_k(\bar{\beta}) + \Delta x_k^T\bar{\beta} \leq 0$. We will consider the important perturbation resulting from a change in a constraint point, in particular from $T_5 = 6$ to $T_5$. For this it is easiest to find $r_k^{<k>}(\bar{\beta})$, $k = 55$, directly as $r_k^{<k>}(\bar{\beta}) = f_p''(T_5)$. From the graph of $f_p''(t)$ in Fig. 2(b), it is clear that the solution $\bar{\beta}$ remains optimal if $T_5$ belongs to the set $\mathcal{V}_5 = [0.063, 0.164] \cup [0.244, 0.391] \cup [0.4, 0.704] \cup [0.798, 0.912]$ (to 3 decimal places), which is marked in Fig. 2(b) by a thick line.

Clearly, with the constraint points $0.4, 0.45, \ldots, 0.6$, the $L_1$ regression estimate $f_p(t)$ satisfies the desired condition $f_p''(t) \leq 0$ for all $t \in \mathcal{V}_5$. The set $\mathcal{V}_5$ covers much of the interval $[0, 1]$ (the length of $\mathcal{V}_5$ is 0.667) but not close to all of it. This raises the design question of where to place additional constraint points so that the the condition is satisfied over a larger set. It is clear from above that to change the solution, any shifted or new constraint point must be outside $\mathcal{V}_5$.

To investigate the effect of additional constraints, we expanded the set of constraints to $f_p''(T_i) \leq 0$, $T_i = 0.1 + 0.05(i - 1)$, $i = 1, \ldots, 17$, and recomputed the constrained $L_1$ regression solution. The new $L_1$ regression estimate $f_p(t)$, shown in Fig. 5(a), appears to be a better fit than the estimate in Fig. 2(a) based on 5 constraints. From Fig. 5(b) it is clear that for the new estimate, the second derivative satisfies

Fig. 3 $L_1$ norm of error in coefficients $\beta(+)$ and bound (o) for constrained $L_1$ estimates of $g'(t)$

Fig. 4 $L_1$ norm of error in coefficients $\beta(+)$ for perturbation $f_p''(0.55) \leq -0.5$
\[ f_p''(t) \leq 0 \] on a larger set \( V_{17} \) (with length 0.715) than \( V_5 \) in Fig. 2(b). Now there are 4 active constraints corresponding to the constraint points 0.25, 0.3, 0.75 and 0.8, marked by + signs in Fig. 5(b).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5a.png}
\caption{\( g'(t) = 10(-t^2 + t) \) (solid), and constrained \( L_1 \) estimate \( f_p(t) \) (dashed)}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig5b.png}
\caption{\( f_p''(t) \) for constraint points 0.1, 0.15, ..., 0.9}
\end{figure}

The computations above were replicated 10 times using 10 different data sets \( y_i = g(t_i) + \varepsilon_i, \ i = 1, \ldots, n \), with different pseudo-random errors (but with the same mean 0 and standard deviation 0.01). For each data set, we computed the constrained \( L_1 \) regression estimate \( f_p(t) \) using the constraint points 0.4, 0.45, ..., 0.6 and the estimate using the constraint points 0.1, 0.15, ..., 0.9. For the two estimates we computed the lengths of the sets \( V_5 \) and \( V_{17} \) of points \( t \) such that \( f_p''(t) \leq 0 \), and these lengths are plotted in Fig. 6(a). Clearly, the length of \( V_{17} \) is significantly greater than that of \( V_5 \) for all the replicates. The estimates themselves were also compared using the following norm (a discrete \( L_1 \) Sobolev norm) of the error:

\[ \| f_p - f \|_W = \frac{1}{1001} \left( \sum_{i=0}^{1000} |f_p(i/1000) - f(i/1000)| + \sum_{i=0}^{1000} |f_p''(i/1000) - f''(i/1000)| \right), \]

where \( f(t) = 10(-t^2 + t) \) and \( f''(t) = -20 \). Note that this error norm takes into account not only deviations in function values but also deviations in curvature as determined by the second derivative. For each of the 10 replicates, the error norms for the two constrained estimates are plotted in Fig. 6(b). Clearly, for all the replicates, the estimate using the constraint points 0.1, 0.15, ..., 0.9 is significantly more accurate with respect to this norm.

It is clear from Fig. 5(b) that the optimal solution with the 17 constraint points 0.1, 0.15, ..., 0.9 is the same as the optimal solution with the four constraint points 0.25, 0.3, 0.75, 0.8 (the ones corresponding to active constraints). In fact, for all ten replicates, the optimal solution with the 17 constraint points is the same as the optimal solution with a smaller set of (different) constraint points, of sizes 4, 6, 5, 4, 5, 3, 4, 5 and 3 points (i.e. the number of active constraints).

If the set of constraint points is expanded to \{0.05, 0.1, 0.15, ..., 0.9, 0.95\} (i.e. two extra points), there is a dramatic change in the optimal solution in that it becomes a constant \( f_p(t) = 1.784 \), which is obviously a poor estimate of \( g'(t) \). Essentially, the constraints have now completely knocked out the sinusoidal terms from the estimate \( f_p(t) \) in (8.1), leaving only the term \( c_0 t \) in (8.2) to fit to the data by least absolute deviations. For this constant solution all the constraints are active. It is clear that this problem would not
occur if the number $p$ of terms in $f_p(t)$ were increased sufficiently (i.e. increasing the degrees of freedom). In practice, however, $p$ would usually be chosen first and then the constraint points would be decided. The example shows that some care should be taken in the selection of the constraint points.

Fig. 6(a) Lengths of $\{t : f_p^n(t) \leq 0\}$ for $0.4, 0.45, \ldots, 0.6$ (+) and $0.1, 0.15, \ldots, 0.9$ (x)

Fig. 6(b) Errors $||f_p - f||_W$ for constraint points $0.4, 0.45, \ldots, 0.6$ (+) and $0.1, 0.15, \ldots, 0.9$ (x)

8.4. Deletion of observations

Lastly we investigate the influence of each observation in Fig. 1(a) on the constrained $L_1$ regression estimate with the constraint points $0.4, 0.45, \ldots, 0.6$ (shown in Fig. 2(a)), by deleting each observation as discussed in Section 6. For each observation number $t = 1, \ldots, 50$, we compute the (unscaled) Cook distance $V_t = \|X_B - X_B^{(o)}\|$ using both the $L_2$ norm and $L_1$ norm in the definition. The results are plotted in Fig. 7. Note that for this data set the most influential observations occur near the middle and near the right hand end point of the interval $[0, 1]$. Using Theorems 6.1 and 6.2 we find that the $L_1$ regression estimate remains unchanged precisely if we delete one of the seven observations, numbers 1, 2, 4, 10, 11, 16 or 17, which is consistent with Fig. 7.

Fig. 8.7 Values of $V_t$ defined by $L_2$ (o) and $L_1$ (+) norms

References


