Robust solutions to linear data fitting problems

G A Watson

Mathematics Division, University of Dundee, Dundee DD1 4HN, Scotland

Abstract. An important problem is that of fitting a linear model to data which are uncertain. If the data are known to lie in a given uncertainty set, then a robust counterpart of such a problem may be interpreted as the problem of finding a solution which is best over all possible perturbations of the data which lie in the set. In particular, robust counterparts of total least squares problems have been studied and good algorithms are available. This work is reviewed, and consideration is then given to robust counterparts of linear data fitting problems considered as errors-in-variables problems, when it is appropriate to work directly with the uncertain variable values. For such problems, and different choices of uncertainty set, the original problems can be replaced by convex optimization problems in fewer variables for which standard software may be applied. This means that robust counterparts of errors-in-variables problems may also be solved by interior point methods in polynomial time.

Keywords: linear data fitting, robustness, least squares
PACS: 07.05.Kf, 02.60.Ed

1. INTRODUCTION

Let data $(y_i, \mathbf{x}_i), i = 1, \ldots, m$, be generated, where $y_i \in R$ and $\mathbf{x}_i \in R^t$. Let these data points be related through a linear model, containing n parameters, so that we can write

$$y_i \approx \sum_{j=1}^{n} a_j \phi_j(\mathbf{x}_i), i = 1, \ldots, m, \tag{1}$$

where $\mathbf{a} \in R^n$ is the vector of free parameters. Normally the data are uncertain, because of errors, so (1) cannot be satisfied exactly, and the issue then is the determination of an appropriate vector $\mathbf{a}$. This means taking account of the nature of the errors, and indeed any further information which is available about these.

Sometimes information is available about the level of uncertainty in the data, and in some application areas (control theory, system identification), there is an emphasis on **robustness**: a solution $\mathbf{a}$ is said to be robust if it solves the "worst case" problem. Put another way, it is best over all allowable perturbations of the data, those which keep the perturbed values in a given "uncertainty set". The resulting problem is often refererred to as a **robust counterpart** of the underlying problem [2], [3].

The most commonly used criterion in data fitting problems is least squares. When there are assumed to be errors in all the variables, then this criterion generalizes to total least squares (TLS). A systematic investigation of TLS and the development of good numerical methods was initiated in [10], and this lead to a huge upsurge of interest. However, statisticians (in particular) have studied analogous problems, so-called errors-in-variables problems, for over a hundred years. This gives a different way of dealing with the errors.

We will consider those criteria in the next two sections, and show how both problems can lead to robust counterparts, which are tractable, and which can be solved by standard software. Section 2 considers the problem from a TLS point of view, and Section 3 gives an errors-in-variables perspective.

2. ROBUST COUNTERPARTS OF TLS PROBLEMS

Defining $A \in R^{m \times n}$ by

$$A_{ij} = \phi_j(\mathbf{x}_i), i = 1, \ldots, m, j = 1, \ldots, n,$$

then (1) can be written in matrix-vector form as $\mathbf{y} \approx A\mathbf{a}$, and if the model equations are interpreted as

$$\mathbf{y} + \mathbf{r} = (A + E)\mathbf{a}, \tag{2}$$

where $\mathbf{r}$ and E are perturbations, then TLS can be used to find values of the parameters. Define the uncertainty set for the data

$$\mathscr{E} = \{(\mathbf{y}+\mathbf{r}, A+E) : \|\mathbf{r}\| \le \rho_1, \|E\| \le \rho_2\},$$

where ρ_1 and ρ_2 are given values and the norms are least squares norms. Then a robust counterpart of the TLS problem is to find

$$\min_{\mathbf{a}\in I\!R^n} \max_{\|\mathbf{r}\|\le\rho_1, \|E\|\le\rho_2} \|\mathbf{y}+\mathbf{r}-(A+E)\mathbf{a}\|. \tag{3}$$

An interpretation of this is that we require to minimize $\|\tilde{A}\mathbf{a}-\tilde{\mathbf{y}}\|$ with respect to $\mathbf{a}$ over the worst of all perturbations defined by

$$(\tilde{A}, \tilde{\mathbf{y}}) \in \mathscr{E} = \{(\mathbf{y}+\mathbf{r}, A+E) : \|\mathbf{r}\| \le \rho_1, \|E\| \le \rho_2\}.$$

Good methods are a vailable to compute solutions (for example, [6], [9]), where "hidden convexity" is exploited, and $\mathbf{a}$ solving (3) also minimizes the convex objective function

$$\|\mathbf{y}-A\mathbf{a}\| + \rho_2\|\mathbf{a}\|. \tag{4}$$

The minimization of this function can be achieved by solving the second order cone programming problem (SOCP)

$$\text{minimize } \lambda + \mu \text{ subject to}$$

$$\begin{aligned} \|\mathbf{y}-A\mathbf{a}\| &\le \lambda, \\ \rho_2\|\mathbf{a}\| &\le \mu, \end{aligned}$$

for which efficient (interior point) methods are available (for example [5], [11], [12]). Alternatively, exploiting the fact that (4) is normally differentiable, a method which involves the Singular Value Decomposition of A is given in [9] and in [6], which requires the solution of a "secular equation" in one variable, followed by the solution of a system of linear equations for $\mathbf{a}$.

Other uncertainty sets have been suggested, for example,$\mathscr{E}$ may be defined by

$$\mathscr{E} = \{(\mathbf{y}+U\mathbf{d}, A+UP) : \|U\| \le 1\},$$

where P and $\mathbf{d}$ are given. With this kind of restriction on the perturbations (and variants in similar spirit), the problem is said to have ellipsoidal uncertainty. Such problems are considered in [3] in the context of a wide range of optimization problems; applications to the present context, and generalizations to more general norms, are considered in [20]. In particular, for least squares, the problems can often be reduced to ones like (4), or to SOCPs, to which standard software can be applied.

In many applications, the perturbations E and $\mathbf{r}$ have a known structure, which might be inherited from A and $\mathbf{b}$. Thus structured problems may be considered. For example we may have

$$E = \sum_{i=1}^{k} u_i A_i, \quad \mathbf{r} = \sum_{i=1}^{k} u_i \mathbf{r}_i,$$

where $A_i, i = 1,\ldots,k$, are given matrices having the required structure, and $\mathbf{r}_i, i = 1,\ldots,m$, are given vectors. Defining $\mathbf{u} = [u_1,\ldots,u_k]^T$, a robust counterpart to the TLS problem, appropriate to this case, might require us to find

$$\min_{\mathbf{a}} \max_{\|\mathbf{u}\|\le\rho} \|\mathbf{y}+\mathbf{r}-(A+E)\mathbf{a}\|,$$

where ρ is given. This is considered in [8], [9], where it is shown that such problems can be posed and solved as semi-definite programming problems (SDPs); see also [15], [17], [16]. Good software is again available for SDPs, with interior point methods effective in solving the problems in polynomial time (see, for example, [13], [14]).

So it is possible to deal with a variety of robust counterparts of TLS problems, which reduce the problems to convex problems for which good software is available. The question arises as to whether the same can be said for robust counterparts of errors-in-variables interpretations of the problems, and this is addressed in the next section.

3. ROBUST COUNTERPARTS OF ERRORS-IN-VARIABLES PROBLEMS

The components of the perturbation matrix E of A introduced in the previous section are regarded as independent of each other, although as indicated there, some structure can be imposed. However, instead of perturbations of the elements of A, an alternative is to consider perturbations of the variables $\mathbf{x}_i, i = 1, \ldots, m$, on which A depends. So the allowable perturbed values would become $\mathbf{y} + \mathbf{r}$ and $\mathbf{x}_i + \mathbf{s}_i, i = 1, \ldots, m$, giving a direct errors-in-variables interpretation, and the analogue of (2) would be

$$y_i + r_i = \sum_{j=1}^{n} a_j \phi_j(\mathbf{x}_i + \mathbf{s}_i), i = 1, \ldots, m. \tag{5}$$

The minimization of the sum of squares of the perturbations in this case is known as orthogonal distance regression [4]. Errors-in-variables models (see for example [7]) arise most frequently in the statistical and psychometric literature, but problems are also encountered in signal processing and time series analysis.

For given $\mathbf{a}$, let

$$v_i = y_i + r_i - \sum_{j=1}^{n} a_j \phi_j(\mathbf{x}_i + \mathbf{s}_i), i = 1, \ldots, m.$$

Given an uncertainty set $\mathscr{E}$, inside which perturbed values of the variables lie, that is

$$(y_i + r_i, \mathbf{x}_i + \mathbf{s}_i, i = 1, \ldots, m) \in \mathscr{E},$$

then of interest is the problem of minimizing $\|\mathbf{v}\|$ with respect to $\mathbf{a}$ over the **worst** case perturbations of y_i and $\mathbf{x}_i$, $i = 1, \ldots, m$, which keep the perturbed values in $\mathscr{E}$. Thus the basic problem, analogous to (3), is to find

$$\min_{\mathbf{a}} \max_{(\mathbf{y}+\mathbf{r}, \mathbf{x}_i+\mathbf{s}_i, i=1,\ldots,m) \in \mathscr{E}} \|\mathbf{v}\|. \tag{6}$$

An interpretation of this is that if

$$\tilde{z}_i = \tilde{y}_i - \sum_{j=1}^{n} a_j \phi_j(\tilde{\mathbf{x}}_i), i = 1, \ldots, m, \tag{7}$$

then we require to minimize $\|\tilde{\mathbf{z}}\|$ with respect to $\mathbf{a}$ over the worst case perturbations defined by

$$(\tilde{y}_1, \ldots, \tilde{y}_m, \tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_m) \in \mathscr{E}.$$

This is a robust counterpart of the errors-in-variables problem.

For given $\mathbf{a} \in R^n$, let

$$z_i = y_i - \sum_{j=1}^{n} a_j \phi_j(\mathbf{x}_i), i = 1, \ldots, m.$$

Assume that $\phi_i \in C^1, i = 1, \ldots, m$. Then

$$v_i = w_i + O(\|\mathbf{s}_i\|^2), i = 1, \ldots, m,$$

where

$$w_i = z_i + r_i - \sum_{j=1}^{n} a_j \nabla \phi_j(\mathbf{x}_i) \mathbf{s}_i, i = 1, \ldots, m,$$

and so to first order in $\mathbf{s}_i$, (6) can be stated as

$$\min_{\mathbf{a}} \max_{(\mathbf{y}+\mathbf{r}, \mathbf{x}_i+\mathbf{s}_i, i=1,\ldots,m) \in \mathscr{E}} \|\mathbf{w}\|. \tag{8}$$

Although we would ideally like to solve (6), for reasons of tractability we will restrict attention here to (8). The emphasis is on robust counterparts having this form, for which good algorithms can be identified. A number of different uncertainty sets will be considered, involving both pointwise and normwise bounds on r_i and $\mathbf{s}_i, i = 1, \ldots, m$, and the corresponding robust counterparts can be posed as convex optimization problems in fewer variables which have the form of standard problems: quadratic programming problems or SOCPs or SDPs. In all cases efficent algorithms are available, with interior point methods able to compute solutions in polynomial time. One formulation results in a special case of the structured robust counterpart met in Section 2, and so may be interpreted as a way of imposing extra structure on the problems in a natural manner. As Stewart puts it in [18] (where by the singular value approach he means TLS), the errors-in-variables approach "has a lot to offer numerical analysts because it folds in the structure of the error from the beginning, whereas in the singular value approach structured errors are grafted on as an afterthought."

REFERENCES

1. A. Beck and A. Ben-Tal. *A global solution for the structured total least squares problem with block circulant matrices, SIAM J. Matrix Anal. Appl.* **27**, 238–255 (2005).
2. A. Ben-Tal and A. Nemirovski, *Robust convex optimization, Mathematics of Operations Research*, **23**, 769–803 (1998).
3. A. Ben-Tal and A. Nemirovski, *Robust solutions to uncertain linear programs via convex programming, Operations Research Letters*, **25**, 1–13 (1999).
4. P. T. Boggs, R. H. Byrd, and R. B. Schnabel. *Stable and efficient algorithm for nonlinear orthogonal distance regression, SIAM J. Sci. Stat. Comput.*, **8**, 1052–1078 (1987).
5. S. Boyd and L. Vandenbergh. *Convex Optimization*, Cambridge University Press (2004).
6. S. Chandrasekaran, G. H. Golub, M. Gu and A. H. Sayed. *An efficient algorithm for a bounded errors-in-variables model, SIAM J. Matrix Anal. Appl.* **20**, 839–859 (1999).
7. W. A. Fuller. *Measurement error models*, Wiley, New York (1987).
8. L. El Ghaoui and H. Lebret. *Robust solutions to least squares problems with uncertain data*, in S. Van Huffel, editor, *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, SIAM, Philadelphia, PA, 1997, pp. 161–170.
9. L. El Ghaoui and H. Lebret. *Robust solutions to least squares problems with uncertain data, SIAM J. Matrix Anal. Appl.* **18**, 1035–1064 (1997).
10. G. H. Golub and C. F. Van Loan, *Matrix Computations, 2nd Ed.*, Johns Hopkins University Press, Baltimore, MD, 1989.
11. M. Grant, S. Boyd and Y. Ye. *Disciplined convex programming*, in Leo Liberti and Nelson Maculan (eds.), *Global Optimization: From Theory to Implementation*, Nonconvex Optimization and Its Application Series, Vol. 84, Springer, Berlin, 2006, pp. 155–210.
12. M. Grant, S. Boyd and Y. Ye. *CVX: Matlab software for disciplined convex programming, Version 1.0RC3*, February, 2007. (webpage: http://www.stanford.edu/∼boyd/cvx/)
13. C.Helmberg, *Semidefinite Programming*, European Journal of Operational Research, **137**, 461–482 (2002).
14. C. Helmberg, *Semidefinite Programming (webpage)*, http://www-user.tu-chemnitz.de/ helmberg/semidef.html
15. H. A. Hindi and S. P. Boyd. *Robust solutions to* l_1, l_2 *and* l_∞ *linear approximation problems using convex optimization, Proc. American Control Conf.* Philadelphia, PA, 1998, pp. 3487–3491.
16. P. Lemmerling, S. Van Huffel and B. De Moor. *Structured total least squares problems: formulations, algorithms and applications*, in S. Van Huffel, editor, *Recent Advances in Total Least Squares Techniques and Errors-in-Variables Modeling*, SIAM, Philadephia, PA, 1997, pp. 215–238.
17. J. B. Rosen, H. Park and J. Glick. *Total least norm formulation and solution for structured problems, SIAM J. Matrix Anal. Appl.* **17**, 110–126 (1996).
18. G. W. Stewart, *Errors in variables for numerical analysts*, in in S. Van Huffel, editor, *Recent Advances in Total Least Squares Technique s and Errors-in-Variables Modeling*, SIAM, Philadephia, PA, 1997, pp. 3–10.
19. G. A. Watson. *Data fitting problems with bounded uncertainties in the data, SIAM J. Matrix Anal. Appl.* **22**, 1274–1293 (2001).
20. G. A. Watson. *Robust solutions to linear approximation problems under ellipsoidal uncertainty*, in S. Van Huffel and P. Lemmerling, eds. *T*otal Least Squares and Errors-in -Variables Modeling, Kluwer, Dordrecht, 213–222, 2002.
21. G. A. Watson. *Robust solutions to a general class of approximation problems, SIAM J. Sci. Comp.* **25**, 1448–1460 (2004).