# Robust counterparts of errors-in-variables problems

G. A. Watson,
Department of Mathematics,
University of Dundee,
Dundee DD1 4HN, Scotland.

**Abstract**

Of interest here are linear data fitting problems with uncertain data which lie in a given uncertainty set. 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. The purpose of this paper is to consider robust counterparts of the problems considered as errors-in-variables problems, when it is appropriate to work directly with the uncertain variable values. It is shown how the original problems can be replaced by convex optimization problems in fewer variables for which standard software may be applied.

## 1 Introduction

Let data $(y_i, \mathbf{x}_i), i = 1, \ldots, m$, be available, where $y_i \in R$ and $\mathbf{x}_i \in R^t$, and all values contain errors. Let these 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. Defining $A \in R^{m \times n}$ by

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

this 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 total least squares can be used to find values of the parameters.

In certain application areas, the size of allowable perturbations is restricted and it may not be possible for (2) to hold exactly. Then it may be appropriate to solve a **robust counterpart** [3] [4] of the problem, for example to find

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

where the norms and the values of $\rho_1, \rho_2$ are given. This may be interpreted as the requirement to minimize $\|\tilde{\mathbf{y}} - \tilde{A}\mathbf{a}\|$ with respect to $\mathbf{a}$ over the **worst** of perturbations defined by $(\tilde{\mathbf{y}}, \tilde{A}) \in \mathcal{E}$, where $\mathcal{E}$ is the uncertainty set

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

This is precisely analogous to robust counterparts identified in many other areas, see for example [3], [9], [10], [14].

When the norms are $l_2$ norms (Frobenius norms on matrices), good methods are available to compute solutions (for example, [7], [10]), where "hidden convexity" is exploited, and $\mathbf{a}$ solving (3) also minimizes the convex objective function

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

The components of the perturbation matrix $E$ of $A$ are regarded as independent of each other, although some structure can be imposed (for example [2], [10], [15], [16]). However, instead of perturbations of the elements of $A$, an alternative is to consider perturbations of the variables $\mathbf{x}_i, i = 1, \dots, m$, on which $A$ depends. So the allowable perturbed values would become $\mathbf{y} + \mathbf{r}$ and $\mathbf{x}_i + \mathbf{s}_i, i = 1, \dots, m$, giving a direct errors-in-variables interpretation, and the analogue of (2) would be

$$y_i + r_i = \sum_{j=1}^{n} a_j \phi_j(\mathbf{x}_i + \mathbf{s}_i), i = 1, \dots, m. \tag{4}$$

The minimization of the sum of squares of the perturbations in this case is known as orthogonal distance regression [5]. Errors-in-variables models (see for example [8]) arise most frequently in the statistical and psychometric literature, but problems are also encountered in signal processing and time series analysis.

The purpose of this paper is to consider robust counterparts of these errors-in-variables (or orthogonal distance regression) problems. It is assumed that the perturbed values are restricted to lie in an uncertainty set $\mathcal{E}$, and therefore it may not be possible for the equations (4) to hold exactly. For given $\mathbf{a}$, let

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

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 $\mathcal{E}$. Thus the basic problem is to find

$$\min_{\mathbf{a}} \max_{(\mathbf{y}+\mathbf{r}, \mathbf{x}_i + \mathbf{s}_i, i=1,\ldots,m) \in \mathcal{E}} \|\mathbf{v}\|. \quad (5)$$

An interpretation of this is that if

$$\tilde{z}_i = \tilde{y}_i - \sum_{j=1}^{n} a_j \phi_j(\tilde{\mathbf{x}}_i), i = 1, \ldots, m, \quad (6)$$

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

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

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 $\mathbf{s}_i$, (5) can be stated as

$$\min_{\mathbf{a}} \max_{(\mathbf{y}+\mathbf{r},\mathbf{x}_i+\mathbf{s}_i,i=1,\dots,m)\in\mathcal{E}} \|\mathbf{w}\|. \tag{7}$$

Although we would ideally like to solve (5), for reasons of tractability we will restrict attention here to (7), and we will examine the problems which arise when different kinds of uncertainty set are used. The emphasis is on robust counterparts having the form (7) for which good algorithms can be identified. In particular, it is shown how the original problems can be replaced by convex optimization problems in fewer variables for which standard software may be applied.

We will be mainly concerned here with the $l_2$ norm, and any unadorned norm will henceforth imply this norm. However, the problems can be defined for other norms (in particular the $l_1$ and $l_\infty$ vector norms may be of interest), and so the use of other norms will be mentioned in Remarks from time to time.

# 2 Pointwise uncertainty

Consider now the case where the level of uncertainty is known for each point. In fact, assume first that it is known for each variable value. Define

$$\mathbf{s} = (\mathbf{s}_1^T, \dots, \mathbf{s}_m^T)^T.$$

Let

$$S_1 = \{(\mathbf{r},\mathbf{s}) : |r_i| \le \rho_i, |(\mathbf{s}_i)_j| \le \gamma_{ij}, i = 1,\dots,m, j = 1,\dots,t\},$$

where $\rho_i, \gamma_{ij}, i = 1,\dots,m, j = 1,\dots,t$ are given. Consider the uncertainty set

$$\mathcal{E}_1 = \{\mathbf{y}+\mathbf{r}, \mathbf{x}_i+\mathbf{s}_i, i = 1,\dots,m : (\mathbf{r},\mathbf{s}) \in S_1\}.$$

Then in this case (7) becomes

$$\min_{\mathbf{a}} \max_{(\mathbf{r},\mathbf{s})\in S_1} \|\mathbf{w}\|. \tag{8}$$

Define

$$G_i = \begin{bmatrix} \nabla\phi_1(\mathbf{x}_i) \\ \nabla\phi_2(\mathbf{x}_i) \\ .. \\ \nabla\phi_n(\mathbf{x}_i) \end{bmatrix} \in R^{n\times t}, i = 1,\dots,m,$$

and let

$$D_i = \text{diag}\{\gamma_{i1}, \ldots, \gamma_{it}\}, i = 1, \ldots, m.$$

**Theorem 1** Let $\mathbf{a}^* \in R^n$ be a solution to the problem

$$\text{minimize } \|\mathbf{c}\|, \text{ where } c_i = |z_i| + \rho_i + \|D_i G_i^T \mathbf{a}\|_1, i = 1, \ldots, m. \tag{9}$$

Then $\mathbf{a}^*$ solves (8).

**Proof** Let $\mathbf{a} \in R^n$ be arbitrary, and let $\mathbf{r}$ and $\mathbf{s}$ lie in $S_1$. Now

$$w_i = z_i + r_i - \mathbf{s}_i^T G_i^T \mathbf{a}, i = 1, \ldots m.$$

Thus

$$\begin{aligned} |w_i| &\leq |z_i| + \rho_i + \sum_{j=1}^{t} \gamma_{ij} |(G_i^T \mathbf{a})_j|, i = 1, \ldots, m, \\ &= |z_i| + \rho_i + \|D_i G_i^T \mathbf{a}\|_1, i = 1, \ldots, m. \end{aligned}$$

Let $i$ be such that $z_i \neq 0$. Then for such $i$, choose

$$r_i = \rho_i \frac{z_i}{|z_i|},$$

$$(\mathbf{s}_i)_j = -\gamma_{ij} \frac{z_i}{|z_i|} \text{sign}((G_i^T \mathbf{a})_j), j = 1, \ldots, t.$$

For $i$ such that $z_i = 0$ let $r_i = \rho_i, (\mathbf{s}_i)_j = -\gamma_{ij}\text{sign}((G_i^T \mathbf{a})_j), j = 1, \ldots, t$. Then $\mathbf{r}$ and $\mathbf{s}$ are in $S_1$. Also if $z_i \neq 0$,

$$w_i = z_i + \rho_i \frac{z_i}{|z_i|} + \sum_{j=1}^{t} \gamma_{ij} \frac{z_i}{|z_i|} |(G_i^T \mathbf{a})_j|,$$

so that

$$|w_i| = |z_i| + \rho_i + \|D_i G_i^T \mathbf{a}\|_1.$$

The same equation holds for any remaining values of $i$, and the result follows. ■

The problem (9) can be stated as

$$\begin{aligned} &\text{minimize } \mathbf{u}^T \mathbf{u} \qquad \text{subject to} \\ |z_i| + \rho_i + \|D_i G_i^T \mathbf{a}\|_1 &\leq u_i, i = 1, \ldots m, \end{aligned}$$

which can be reformulated as a quadratic programming problem. Of course, when $t = 1$, the $l_1$ norm just becomes a single modulus term.
**Remark 1** The problem is easily solved when the norm in (8) (and so in (9)) is the $l_1$ norm, because then we require just to find $\mathbf{a}^*$ to solve the linear $l_1$ approximation problem

$$\text{minimize } \|\mathbf{y} - A\mathbf{a}\|_1 + \|M\mathbf{a}\|_1,$$

where $M$ is the $(mt \times n)$ matrix formed by

$$M = \begin{bmatrix} D_1 G_1^T \\ ... \\ D_m G_m^T \end{bmatrix}.$$

**Remark 2** For the $l_\infty$ norm, the problem for $\mathbf{a}^*$ simplifies to

$$\text{minimize } h \text{ subject to}$$

$$|z_i| + \rho_i + \|D_i G_i^T \mathbf{a}\|_1 \leq h, i = 1, \ldots, m,$$

which can be reformulated as a linear programming problem.

Next consider the set

$$S_2 = \{(\mathbf{r}, \mathbf{s}) : |r_i| \leq \rho_i, \|\mathbf{s}_i\| \leq \gamma_i, i = 1, \ldots, m\},$$

where $\rho_i, \gamma_i, i = 1, \ldots, m$ are given. This is of course just the same as before if $t = 1$, so we are interested in the cases when $t > 1$. The uncertainty set becomes

$$\mathcal{E}_2 = \{\mathbf{y} + \mathbf{r}, \mathbf{x}_i + \mathbf{s}_i, 1 = 1, \ldots, m : (\mathbf{r}, \mathbf{s}) \in S_2\},$$

and the problem (8) becomes

$$\min_{\mathbf{a}} \max_{(\mathbf{r},\mathbf{s}) \in S_2} \|\mathbf{w}\|. \tag{10}$$

In what follows, the usual notation $\partial$ is used to denote the subdifferential of a convex function; in particular, for a norm $\|.\|_A$ on $R^m$:

$$\partial \|\mathbf{v}\|_A = \{\mathbf{u} \in R^m : \|\mathbf{v}\| = \mathbf{v}^T \mathbf{u}, \|\mathbf{u}\|_A^* \leq 1\},$$

where $\|.\|_A^*$ denotes the norm dual to $\|.\|$. For $l_2$ norms, this set is just the singleton $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ whenever this is defined. The following result can be established by a proof which closely follows that of Theorem 1.

**Theorem 2** Let $\mathbf{a}^* \in R^n$ be a solution to the problem

$$\text{minimize } \|\mathbf{c}\|, \text{ where } c_i = |z_i| + \rho_i + \gamma_i \|G_i^T \mathbf{a}\|, i = 1, \ldots, m. \tag{11}$$

Then $\mathbf{a}^*$ solves (10). Taking $z_i^*$ to mean $z_i(\mathbf{a}^*)$, the optimal perturbations are

$$r_i = \rho_i \frac{z_i^*}{|z_i^*|}, \text{ if } z_i^* \neq 0,$$

otherwise $r_i = \rho_i$, and

$$\mathbf{s}_i = -\gamma_i \frac{z_i^*}{|z_i^*|} \mathbf{u}_i, \text{ if } z_i^* \neq 0,$$

otherwise $\mathbf{s}_i = -\gamma \mathbf{u}_i$, where $\mathbf{u}_i \in \partial \|G_i^T \mathbf{a}^*\|, i = 1, \ldots, m$.

The problem (11) can be stated as

$$\text{minimize } \mathbf{u}^T \mathbf{u} \text{ subject to}$$

$$|z_i| + \rho_i + \gamma_i \|G_i^T \mathbf{a}\| \leq u_i, \; i = 1, \ldots, m.$$

The presence of the $l_2$ norm in the constraints suggests the alternative formulation as

$$\text{minimize } h \text{ subject to } \|\mathbf{u}\| \leq h,$$

and

$$|z_i| + \rho_i + \gamma_i \|G_i^T \mathbf{a}\| \leq u_i, \; i = 1, \ldots, m.$$

Alternatively, this is

$$\text{minimize } h \text{ subject to}$$

$$\begin{aligned} \|\mathbf{u}\| &\leq h, \\ \gamma_i \|G_i^T \mathbf{a}\| &\leq u_i - z_i - \rho_i, \; i = 1, \ldots, m, \\ \gamma_i \|G_i^T \mathbf{a}\| &\leq u_i + z_i - \rho_i, \; i = 1, \ldots, m. \end{aligned}$$

This is a second order cone programming problem (the minimization of a linear function over the intersection of an affine set and the (Cartesian) product of second order cones.) Good interior point methods exist for such problems (see, for example [6], [11], [12]), and solutions can be obtained in polynomial time.

**Remark 3** If for every $i$, the bound on $\mathbf{s}_i$ is the more general one $\|\mathbf{s}_i\|_A \leq \gamma_i$ where $\|.\|_A$ is any norm on $R^t$, then $c_i$ in (11) becomes

$$c_i = |z_i| + \rho_i + \gamma_i \|G_i^T \mathbf{a}\|_A^*, i = 1, \ldots, m,$$

where $\|.\|_A^*$ denotes the norm which is dual to $\|.\|_A$. For the optimal perturbations, $\mathbf{u}_i$ is defined by

$$\mathbf{u}_i \in \partial \|G_i^T \mathbf{a}\|_A^*, i = 1, \ldots, m.$$

If the norm in (10) is the $l_1$ norm, and the norm $\|.\|_A$ is the $l_\infty$ norm (so that its dual is the $l_1$ norm), then the problem to be solved is the minimization with respect to $\mathbf{a}$ of

$$\sum_{i=1}^m (|z_i| + \rho_i + \gamma_i \|G_i^T \mathbf{a}\|_1),$$

or equivalently the minimization of

$$\|\mathbf{y} - A\mathbf{a}\|_1 + \sum_{i=1}^m \gamma_i \|G_i^T \mathbf{a}\|_1.$$

This is just a linear $l_1$ approximation problem.
**Remark 4** If the norm in (10) is the $l_\infty$ norm, then the problem can be expressed as

$$\text{minimize } h \ \text{ subject to}$$

$$|z_i| + \rho_i + \gamma_i \|G_i^T \mathbf{a}\| \leq h, \ i = 1, \ldots, m,$$

or

$$\text{minimize } h \ \text{ subject to}$$

$$\begin{aligned} \gamma_i \|G_i^T \mathbf{a}\| &\leq h - z_i - \rho_i, \ i = 1, \ldots, m, \\ \gamma_i \|G_i^T \mathbf{a}\| &\leq h + z_i - \rho_i, \ i = 1, \ldots, m. \end{aligned}$$

This is just a second order cone programming problem.

# 3 Normwise uncertainty

Assume now that the uncertainty is only known with respect to norms of the variables. First, define

$$S_3 = \{(\mathbf{r}, \mathbf{s}) : \|\mathbf{r}\| \leq \rho, \|\mathbf{s}_i\| \leq \gamma_i, i = 1, \ldots, m\},$$

and the uncertainty set

$$\mathcal{E}_3 = \{\mathbf{y} + \mathbf{r}, \mathbf{x}_i + \mathbf{s}_i, i = 1, \dots, m : (\mathbf{r}, \mathbf{s}) \in S_3\}.$$

Then the problem of minimizing $\|\mathbf{w}\|$ with respect to $\mathbf{a}$ over the worst of perturbations which keep perturbed values in $\mathcal{E}_3$ can be stated as

$$\min_{\mathbf{a}} \max_{(\mathbf{r},\mathbf{s}) \in S_3} \|\mathbf{w}\|. \tag{12}$$

Since

$$\mathbf{w} = \mathbf{z} + \mathbf{r} - \sum_{i=1}^{m} \mathbf{e}_i \mathbf{a}^T G_i \mathbf{s}_i,$$

for any $\mathbf{a}$ and $(\mathbf{r}, \mathbf{s}) \in S_3$, it is clear that

$$\|\mathbf{w}\| \leq \|\mathbf{z}\| + \rho + \sum_{i=1}^{m} \gamma_i \|G_i^T \mathbf{a}\|.$$

The problem of minimizing the right hand side is one of minimizing a sum of Euclidean norms and good methods are available, for example [1]. However, unlike previous results, it is not possible to choose $(\mathbf{r}, \mathbf{s}) \in S_3$ so that this inequality is attained, so this approach will not solve (12) (except approximately). A similar difficulty occurs if a bound $\|\mathbf{s}\| \leq \gamma$ is used in $S_3$.

Perhaps a natural question to ask is this. For the problem (3), as has already been noted, a solution is obtained by minimizing

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

What does this correspond to in the context of (4)? The answer is given in the following theorem, whose proof again closely follows that of Theorem 1. For any $\mathbf{s}$, define $M$ by

$$M^T = [G_1 \mathbf{s}_1, \dots, G_m \mathbf{s}_m].$$

**Theorem 3**. Let $\mathbf{a}^*$ minimize

$$\|\mathbf{y} - A\mathbf{a}\| + \gamma \|\mathbf{a}\|. \tag{13}$$

Then if $G_i^T \mathbf{a}^* \neq 0, i = 1, \dots, m$, $\mathbf{a}^*$ solves the problem

$$\min_{\mathbf{a}} \max_{\|\mathbf{r}\| \leq \rho, \frac{\|M\mathbf{a}\|}{\|\mathbf{a}\|} \leq \gamma} \|\mathbf{w}\|. \tag{14}$$

Then, taking $\mathbf{z}^*$ to denote $\mathbf{z}(\mathbf{a}^*)$, optimal perturbations are given by choosing

$$\mathbf{r} = \rho \frac{\mathbf{z}^*}{\|\mathbf{z}^*\|},$$

and choosing $\mathbf{s}_i$ to satisfy

$$\mathbf{s}_i^T G_i^T \mathbf{a}^* = -\gamma \frac{z_i^*}{\|\mathbf{z}^*\|} \|\mathbf{a}^*\|, i = 1, \ldots, m.$$

The bound on $\mathbf{s}$ here leads to a rather artificial uncertainty set. To return to more natural situations, it seems that different techniques are needed. For example let us define

$$\mathbf{d}^T = [r_1, \ldots, r_m, \mathbf{s}_1^T, \ldots, \mathbf{s}_m^T],$$

so that $\mathbf{d} \in R^{m(1+t)}$, and define the uncertainty set

$$\mathcal{E}_4 = \{\mathbf{y} + \mathbf{r}, \mathbf{x} + \mathbf{s}, \|\mathbf{d}\| \leq \rho\}.$$

Then the problem of minimizing $\|\mathbf{w}\|$ with respect to $\mathbf{a}$ over the worst of perturbations which keep values of the variables in $\mathcal{E}_4$ can be stated as that of finding

$$\min_{\mathbf{a}} \max_{\|\mathbf{d}\| \leq \rho} \|\mathbf{w}\|. \tag{15}$$

Let

$$E(\mathbf{s}) = \begin{bmatrix} \mathbf{s}_1^T G_1^T \\ .. \\ \mathbf{s}_m^T G_m^T \end{bmatrix}.$$

As before, since

$$w_i = z_i + r_i - \mathbf{s}_i^T G_i^T \mathbf{a}, i = 1, \ldots .m,$$

then

$$\begin{aligned} \mathbf{w} &= \mathbf{z} + \mathbf{r} - E(\mathbf{s})\mathbf{a} \\ &= \mathbf{y} + \mathbf{r} - (A + E(\mathbf{s}))\mathbf{a} \\ &= \mathbf{b}(\mathbf{d}) - G(\mathbf{d})\mathbf{a} \end{aligned}$$

where

$$G(\mathbf{d}) = A + \sum_{i=1}^{m} \mathbf{e}_i \sum_{j=1}^{t} (\mathbf{e}_j^T G_i^T)(\mathbf{e}_j^T \mathbf{s}_i),$$

$$\mathbf{b}(\mathbf{d}) = \mathbf{y} + \sum_{i=1}^{m} \mathbf{e}_i r_i.$$

Thus (15) is solved by finding the solution to the problem

$$\min_{\mathbf{a}} \max_{\|\mathbf{d}\| \le \rho} \|G(\mathbf{d})\mathbf{a} - \mathbf{b}(\mathbf{d})\|. \tag{16}$$

The problem (16) is a special case of a structured robust problem (see, for example, [10]), which shows that the present treatment is a way of imposing extra structure naturally. Now

$$G(\mathbf{d})\mathbf{a} - \mathbf{b}(\mathbf{d}) = A\mathbf{a} - \mathbf{y} + [-I : \; D(\mathbf{a})]\mathbf{d},$$

where

$$D(\mathbf{a}) = \mathrm{diag}\{\mathbf{a}^T G_1, \ldots, \mathbf{a}^T G_m\} \in R^{m \times tm}.$$

Setting

$$M(\mathbf{a}) = [-I : \; D(\mathbf{a})],$$

it follows that

$$\|G(\mathbf{d})\mathbf{a} - \mathbf{b}(\mathbf{d})\|^2 = \begin{bmatrix} 1 \\ \mathbf{d} \end{bmatrix}^T \begin{bmatrix} \|A\mathbf{a} - \mathbf{y}\|^2 & (A\mathbf{a} - \mathbf{y})^T M(\mathbf{a}) \\ M(\mathbf{a})^T (A\mathbf{a} - \mathbf{y}) & M(\mathbf{a})^T M(\mathbf{a}) \end{bmatrix} \begin{bmatrix} 1 \\ \mathbf{d} \end{bmatrix}.$$

Let $\mathbf{a}$ be fixed. Then in this form, Lemma 2.1 of [10] can be used (see also Theorem 2 of [14] and the general theory of [3]) to give the maximum of this function subject to $\|\mathbf{d}\| \le \rho$ as the solution of the semi-definite programming problem (SDP) in the two variables $\lambda$ and $\tau$:

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

$$\begin{bmatrix} \lambda - \rho^2 \tau - \|A\mathbf{a} - \mathbf{y}\|^2 & -(A\mathbf{a} - \mathbf{y})^T M(\mathbf{a}) \\ -M(\mathbf{a})^T (A\mathbf{a} - \mathbf{y}) & \tau I - M(\mathbf{a})^T M(\mathbf{a}) \end{bmatrix} \ge 0.$$

Using the Schur complement, the positive semi-definite constraint can be restated as

$$\begin{bmatrix} \lambda - \rho^2 \tau & \mathbf{0}^T & (A\mathbf{a} - \mathbf{y})^T \\ \mathbf{0} & \tau I_{m+tm} & M(\mathbf{a})^T \\ A\mathbf{a} - \mathbf{y} & M(\mathbf{a}) & I_m \end{bmatrix} \ge 0. \tag{17}$$

The solution to (16) can now be found by determining $\lambda, \tau, \mathbf{a}$ to solve

$$\text{minimize } \lambda \text{ subject to (17).}$$

Because the problem matrix depends affinely on the free parameters, there exist good interior point methods for solving problems of this type: see for example [13]. The structure of the above matrix in fact makes it easy to write down conditions which must be satisfied if it is to be positive semi-definite. When $t = 1$, these can be used to write down a quadratic programming problem (otherwise a second order cone programming problem) whose solution approximates that of the SDP [18].

**Remark 5** The $l_2$ norm on the perturbation size implies a correlated bound on the perturbation. If this norm (only) is replaced by the $l_\infty$ norm in (16), then an upper bound on the worst case residual can be minimized by posing the problem again as an SDP (for example, [10]). When both the norms in (16) are $l_1$ or $l_\infty$ norms, then (16) can be posed as a linear programming problem [14].

# 4 Concluding remarks

Of concern here has been the treatment of robust counterparts of errors-in-variables problems. For the linearized approximation given by (7), various uncertainty sets have been considered. For such sets, and the $l_2$ norm, it has been shown that there is hidden convexity, and the problem (7) is often equivalent to a different problem which can be posed as a standard convex optimization problem in fewer variables (for example a second order cone programming problem or a semi-definite programming problem) and for which efficient interior point methods are available.

An indication has also been given of how the results can be extended to problems involving $l_1$ and $l_\infty$ norms. Mainly, these generate linear programming problems, for which good interior point methods are again available.

The treatment of (7) rather than (5) is of course a compromise, which enables tractable problems to emerge. It may be that (5) can in fact be solved through the solution of a sequence of problems of the form (7), where the solution to the simpler (linearised) problem is used as a means of updating an approximation to the solution of (5). However, this remains to be investigated. The emphasis here has been on the replacement of the original problems by ones with fewer variables which can be solved by standard software, and we have not addressed statistical aspects, although obviously the form of the problems has implications for any statistical analysis.

**Acknowledgement** I am grateful to the referees for helpful suggestions.

# References

[1] K. D. Anderson, E. Christiansen, A. R. Conn and M. L. Overton. An efficient primal-dual interior point method for minimizing a sum of Euclidean norms, *SIAM J. Scient. Comp.* 22: 243–262 (2000).

[2] 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).

[3] A. Ben-Tal and A. Nemirovski. Robust convex optimization, *Mathematics of Operations Research*, 23: 769–803, 1998.

[4] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs, *Operations Research Letters*, 25: 1–13, 1999.

[5] P. T. Boggs, R. H. Byrd, and R. B. Schnabel. Stable and efficient algorithm for nonlinear orthogonal distance regression, *SIAMJ. Sci. Stat. Comput.*, 8: 1052–1078, 1987.

[6] S. Boyd and L. Vandenbergh. Convex Optimization, Cambridge University Press (2004).

[7] 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.

[8] W. A. Fuller. Measurement error models, Wiley, New York (1987).

[9] 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*, 161–170, SIAM, Philadelphia, PA, 1997.

[10] L. El Ghaoui and H. Lebret. Robust solutions to least squares problems with uncertain data, *SIAM J. Matrix Anal. Appl.* 18: 1035–1064, 1997.

[11] M. Grant, S. Boyd and Y. Ye. Disciplined convex programming, in Leo Liberti and Nelson Maculan (eds.), *Global Optimization: From Theory to Implementation*, Kluwer (Nonconvex Optimization and its Applications series), Dordrecht, 2005.

[12] M. Grant, S. Boyd and Y. Ye. CVX: Matlab software for disciplined convex programmimng, Version 1.0 beta 2, April, 2006.

[13] C. Helmberg.

http://www-user.tu-chemnitz.de/ helmberg/semidef.html

[14] H. A. Hindi and S. P. Boyd. Robust solutions to $l_1$, $l_2$ and $l_\infty$ linear approximation problems using convex optimization, *Proc. American Control Conf.* Philadelphia, 3487–3491, June 1998.

[15] 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*, 215–238, SIAM, Philadelphia, PA, 1997.

[16] 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.

[17] G. A. Watson. Data fitting problems with bounded uncertainties in the data, *SIAM J. Matrix Anal. Appl.* 22: 1274–1293, 2001.

[18] G. A. Watson. Robust solutions to a general class of approximation problems, *SIAM J. Sci. Comp.* 25: 1448–1460, 2004.