Robust Solutions to l_1, l_2, and l_∞ Uncertain Linear Approximation Problems using Convex Optimization [1]

Haitham A. Hindi Stephen P. Boyd
Department of Electrical Engineering, Durand Building, Stanford University, CA 94305 [2]

Abstract

We present minimax and stochastic formulations of some linear approximation problems with uncertain data in $\mathbf{R}^n$ equipped with the Euclidean (l_2), Absolute-sum (l_1) or Chebyshev (l_∞) norms. We then show that these problems can be solved using convex optimization. Our results parallel and extend the work of El-Ghaoui and Lebret on robust least squares [3], and the work of Ben-Tal and Nemirovski on robust conic convex optimization problems [1]. The theory presented here is useful for desensitizing solutions to ill-contitioned problems, or for computing solutions that guarantee a certain performance in the presence of uncertainty in the data.

1 Introduction

Consider the problem of finding a solution x to the system $A\,x \simeq b$, where the coefficient matrices $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$ are uncertain. Over the past several decades, the sensitivity of the solution x to perturbations in A or b has received much attention. Yet it is only recently that algorithms have been developed that explicitly use specific information about the perturbations to desensitize the solution. These include total least squares (TLS) [4], and more recently, robust least squares (RLS) [3] and robust convex optimization [1, 6].

In this paper we present minimax and stochastic formulations of linear approximation problems with uncertain data in $\mathbf{R}^n$ equipped with the Euclidean, Absolute-sum, or Chebyshev norms. We then show that these problem can all be reduced to convex optimization problems which can be solved very efficiently using interior point methods. Our results parallel and extend the work in [3] on l_2 norms to the l_1 and l_∞ norms, and the work in [1, 6] to more general conic problems.

The engineering applications of robust approximation using the three norms above include desensitizing ill-conditioned problems, system identification in the presence of norm bounded noise [3], robust interpolation [3], steering of charged particle beams [9], and signal processing [2]. Due to space limitations, these will not be discussed here in detail.

[1] This work was supported by DOE contract # DE-AC03-76SF00515.
[2] email: hhindi@isl.stanford.edu, boyd@isl.stanford.edu

2 Problem Formulation

We model the uncertainty in the coefficient matrices as $[A\ b]+[\Delta A\ \Delta b]$, where A and b are the problem data matrices and $[\Delta A\ \Delta b]$ are the perturbations. Note that it is easy to specialize the results here to the case where the uncertainty is only in A or in b.

In addition to standard notation in the literature, we will use the following: If x_1 and x_2 are column vectors in $\mathbf{R}^{n_1}$ and $\mathbf{R}^{n_2}$ respectively, then (x_1, x_2) denotes the column vector in $\mathbf{R}^{n_1+n_2}$ formed by the concatenation of x_1 and x_2. Also if x is a column vector in $\mathbf{R}^n$, then $\mathrm{sgn}(x)$ is the column vector in $\mathbf{R}^n$ whose components are the signs of the corresponding components of x.

2.1 Minimax Formulation

If the perturbations $[\Delta A\ \Delta b]$ are known to lie in some subset Ω of $\mathbf{R}^{m\times(n+1)}$, then following [1, 3], we may define the worst case perturbation function as

$$\varepsilon(x) \triangleq \max_{[\Delta A\ \Delta b]\in\ \Omega} \|(A+\Delta A)x - (b+\Delta b)\|_p \qquad (1)$$

where $1 \le p \le \infty$. This function measures the norm of the worst case residual over the set of all possible coefficient matrices in Ω, for a given x. We may now state the following minimax formulation of the uncertain data linear approximation problem:

$$\min_x \max_{[\Delta A\ \Delta b]\in\ \Omega} \|(A+\Delta A)x - (b+\Delta b)\|_p. \qquad (2)$$

We note that ε is clearly a convex function of x, since it is the max of a parametrized set of convex functions in x. In some cases [1, 3], the evaluation of ε may not be a tractable problem, since it requires the computation of the maximum of a convex function. However, we will show that for the case of the Euclidean, Absolute-sum and the Chebyshev norms, the problem can be solved efficiently for some useful choices of Ω.

In particular, we will focus on two particular choices of Ω. Suppose very little is known about the structure of the perturbations, and all we want to do is desensitize the solution. Then we may model the perturbations as a norm bounded and unstructured. One would then set Ω to be:

$$\mathcal{U} = \{[\Delta A\,\Delta b] \mid \|[\Delta A\,\Delta b]\|_p \le \rho\}$$

where $\rho > 0$ is a measure of the degree of the uncertainty. Using this in definitions (1) and (2) gives the corresponding unstructured worst case perturbation function ε_{up} and unstructured worst case uncertain approximation problem.

If there is some structure to the perturbations, e.g. if A is sparse or Toeplitz, then solving (2) with $\Omega = \mathcal{U}$ can give a conservative solution. In this case one might chose to model the perturbations as structured and norm bounded. Then one would set Ω to be:

$$\mathcal{S} = \{[\Delta A\, \Delta b] \mid \Delta A = \sum_{i=1}^{L} A_i\, \delta_i,\ \Delta b = \sum_{i=1}^{L} b_i\, \delta_i,\ \|\delta\|_p \leq \rho\}$$

where $A_i \in \mathbf{R}^{m\times n}$ and $b_i \in \mathbf{R}^m$ are given fixed matrices, and $A_0 = A$ and $b_0 = b$. Using $\mathcal{S}$ in definitions (1) and (2) gives the corresponding structured worst case perturbation function ε_{sp} and structured worst case uncertain approximation problem.

2.2 Stochastic Formulation

When $[\Delta A\ \Delta b]$ are random variables with known statistics, then it may be more appropriate to minimize the unstructured stochastic residual function:

$$\varepsilon_{us}(x) \triangleq \mathbf{E}\,\|(A + \Delta A)x - (b + \Delta b)\|_p^p \qquad (3)$$

where $\mathbf{E}$ denotes the expectation, which is taken with respect to the perturbations $[\Delta A\ \Delta b]$. This leads to the following (unstructured) stochastic uncertain data approximation problem:

$$\min_x \mathbf{E}\,\|(A + \Delta A)x - (b + \Delta b)\|_p^p. \qquad (4)$$

This stochastic formulation also admits a structured version. Let $A_i \in \mathbf{R}^{m\times n}$ and $b_i \in \mathbf{R}^m$ be given fixed matrices, as above. Let δ be a random vector with known statistics. Then the structured versions of the perturbation function and associated minimization problem are given by:

$$\varepsilon_{ss}(x) \triangleq \mathbf{E}\,\|(A_0 + \sum_{i=1}^{L} A_i\,\delta_i)x - (b_0 + \sum_{i=1}^{L} b_i\,\delta_i)\|_p^p$$

$$\min_x \mathbf{E}\,\|(A_0 + \sum_{i=1}^{L} A_i\,\delta_i)x - (b_0 + \sum_{i=1}^{L} b_i\,\delta_i)\|_p^p, \qquad (5)$$

where the expectation is now taken with respect to δ.

The functions ε_{us} and ε_{ss} are both convex in x, since both are the positively weighted integral of a parametrized set of convex functions [8]. Once again, the evaluation of these functions may not be a simple matter. For example, we do not know how to evaluate (3) for the Absolute-sum and Chebyshev cases. In the Euclidean case however, we will show that both functions take on the form of a convex quadratic function which can be efficiently minimized.

3 The Euclidean Case

The minimax formulation of the uncertain data linear approximation problem for the l_2 case has been solved by El-Ghaoui and Lebret [3]. We simply restate their results here for comparison and for completeness. The results on the stochastic l_2 formulation, though quite straight forward, seem new.

3.1 Solving the Minimax Formulation

Specifying $\Omega = \mathcal{U}$ with $p = 2$ in (1), we obtain the unstructured perturbation function:

$$\varepsilon_{u2}(x) = \max_{\|[\Delta A\ \Delta b]\|_2 \leq \rho} \|(A + \Delta A)x - (b + \Delta b)\|_2. \qquad (6)$$

In this case the relevant results from [3] can be summarized in the following theorem:

Theorem 1 *The unstructured worst-case perturbation function ε_{u2} can be computed as*

$$\varepsilon_{u2}(x) = \|Ax - b\|_2 + \rho\|(x, -1)\|_2. \qquad (7)$$

The solution to the unstructured uncertain data approximation problem can be obtained by solving the following convex optimization problem:

$$\min_x \{\|Ax - b\|_2 + \rho\|(x, -1)\|_2\}, \qquad (8)$$

which can be solved efficiently as a second order cone program (SOCP) [6].

Remark 1 The equation (8) in the theorem has the form of a multi-objective optimization problem, and shows the direct trade off between accuracy and robustness: the robustness requirement adds a term to the cost which penalizes large solutions.

Specifying $\Omega = \mathcal{S}$ with $p = 2$ in (1) yields the structured perturbation function:

$$\varepsilon_{s2}(x) = \max_{\|\delta\|_2 \leq \rho} \left\| (A_0 + \sum_{i=1}^{L} A_i\delta_i)x - (b_0 + \sum_{i=1}^{L} b_i\delta_i) \right\|_2,$$

which can be written as:

$$\varepsilon_{s2}(x) = \max_{\|\delta\|_2 \leq \rho} \|F\,\delta - g\|_2 \qquad (9)$$

$$F(x) = [(A_1\,x - b_1)\ldots(A_L\,x - b_L)], \qquad g(x) = -(A_0\,x - b_0).$$

(When x is fixed, we will sometimes drop the arguments of F and g.)

Theorem 2 *For a fixed x, $\varepsilon_{s2}(x)^2$ can be computed by solving the following semi-definite program (SDP) [8] in the two variables λ and τ:*

$$\text{minimize} \quad \lambda \quad \text{subject to} \begin{bmatrix} \lambda - \rho^2\tau - r & p^T \\ p & \tau I - Q \end{bmatrix} \geq 0,$$

where $Q = F^TF$, $p = F^Tg$, $r = g^Tg$ and F and g are as in (9). The solution to the structured uncertain data linear approximation can be obtained by solving the following SDP in (λ, τ, x):

$$\text{minimize} \quad \lambda \quad \text{subject to} \begin{bmatrix} \lambda - \rho^2\tau & 0 & -g(x)^T \\ 0 & \tau I & F(x)^T \\ -g(x) & F(x) & I \end{bmatrix} \geq 0.$$

Remark 2 Both Theorems 1 and 2 can be obtained as special cases of the general theory presented in [1] on uncertain second order conic problems with ellipsoidal uncertainty.

3.2 Solving the Stochastic Formulation

Let us assume that the perturbations are all random and zero mean, and that their covariances are known. Then we may compute the stochastic unstructured perturbation function as follows:

$$\begin{aligned}\varepsilon_{su}(x) &= \mathbf{E}\,\|(A+\Delta A)x-(b+\Delta b)\|_2^2\\ &= \mathbf{E}\,\|(A\,x-b)+[\Delta A\,\Delta b](x,-1)\|_2^2\\ &= \|A\,x-b\|_2^2+\mathbf{E}\,\|[\Delta A\,\Delta b](x,-1)\|_2^2\end{aligned}$$

The second term can be written as:

$$\begin{aligned}\mathbf{E}\,\|[\Delta A\,\Delta b](x,-1)\|_2^2 &= \mathbf{E}\,(x,-1)^T[\Delta A\,\Delta b]^T[\Delta A\,\Delta b](x,-1)\\ &= (x,-1)^T(\mathbf{E}\,[\Delta A\,\Delta b]^T[\Delta A\,\Delta b])(x,-1).\end{aligned}$$

The matrix $\mathbf{E}\,[\Delta A\,\Delta b]^T[\Delta A\,\Delta b]$ is simply the covariance matrix of the transposed perturbations $[\Delta A\,\Delta b]^T$. Hence we have the following result:

Theorem 3 *Assume that the perturbations $[\Delta A\,\Delta b]$ are random and zero mean. Let $R=\mathbf{E}\,[\Delta A\,\Delta b]^T[\Delta A\,\Delta b]$ denote the covariance matrix of the (transposed) perturbations. Then the stochastic unstructured perturbation function is a quadratic function in x*

$$\varepsilon_{us}(x)=\|A\,x-b\|_2^2+(x,-1)^T R\,(x,-1).$$

The solution to the stochastic unstructured uncertain linear approximation problem can be obtained by solving:

$$\min_x\{\|A\,x-b\|_2^2+(x,-1)^T R\,(x,-1)\} \tag{10}$$

which is a convex quadratic optimization problem.

In a similar manner, one can show the following result for the structured case:

Theorem 4 *Let δ be a zero mean random vector with covariance matrix R_δ. Let $a_{k,i}^T$ denote the ith row of the matrix A_k, and $b_{k,i}$ the ith component of vector b_k. Let $\tilde{A}_i^T=[a_{1,i}\ldots a_{L,i}]$ and $\tilde{b}_i^T=[b_{1,i}\ldots b_{L,i}]$. Then the stochastic structured perturbation function is a quadratic function in x*

$$\varepsilon_{us}(x)=\|A_0\,x-b_0\|_2^2+(x,-1)^T S\,(x,-1)$$

where $S=\left(\sum_{i=1}^L[\tilde{A}_i\,\tilde{b}_i]\right)^T R_\delta\left(\sum_{i=j}^L[\tilde{A}_j\,\tilde{b}_j]\right)$. Hence the solution to the stochastic structured uncertain approximation problem can be obtained by solving:

$$\min_x\{\|A_0\,x-b_0\|_2^2+(x,-1)^T S\,(x,-1)\} \tag{11}$$

which is a convex quadratic optimization problem.

Remark 3 Equations (10) and (11) again show a direct trade off between accuracy and uncertainty in the coefficients.

4 The Absolute Sum Case

We will now obtain some analogous results for the space $(\mathbf{R}^n,\|\cdot\|_1)$. The corresponding norm induced on a matrix $A\in\mathbf{R}^{m\times n}$ is the "max-col-sum" norm:

$$\|A\|_1=\max_{\|x\|_1\le 1}\|A\,x\|_1=\max_{j=1,\ldots,n}\sum_{i=1}^m|a_{ij}|. \tag{12}$$

Solving approximation problems in the l_1-norm may be preferable to the l_2-norm, when it is suspected that the data may contain large "outliers", since the l_1-norm does not put relatively more weight on large errors than on small ones.

4.1 Solving the Minimax Formulation

Specifying $\Omega=\mathcal{U}$ and $p=1$ in (1) we obtain:

$$\varepsilon_{u1}(x)=\max_{\|[\Delta A\ \Delta b]\|_1\le\rho}\|(A+\Delta A)x-(b+\Delta b)\|_1.$$

It follows from the triangle inequality and (12) that for any $[\Delta A\,\Delta b]$ in $\mathcal{U}$:

$$\|(A+\Delta A)x-(b+\Delta b)\|_1\le\|Ax-b\|_1+\rho\|(x,-1)\|_1. \tag{13}$$

The (nonunique) worst case perturbation can now easily be found: it is the perturbation for which equality holds. Let u be any unit 1-norm vector. Then

$$[\Delta A\,\Delta b]_{\rm wc}=\begin{cases}\rho\,\frac{Ax-b}{\|Ax-b\|_1}\mathrm{sgn}((x,-1))^T & ;\text{ if } Ax-b\neq 0\\ \rho u\,(\mathrm{sgn}(x,-1))^T & ;\text{ otherwise.}\end{cases}$$

Thus we conclude that:

Theorem 5 *The unstructured worst case perturbation function can be computed as*

$$\varepsilon_{u1}(x)=\|Ax-b\|_1+\rho\|(x,-1)\|_1 \tag{14}$$

Therefore, the unstructured uncertain approximation problem can be solved as:

$$\min_x\{\|Ax-b\|_1+\rho\|(x,-1)\|_1\},$$

which is a convex optimization problem that can be solved using linear programming or other specialized algorithms.

This result is identical to that obtained for the Euclidean norm, and again exhibits the direct trade off between accuracy and robustness.

Specifying $\Omega=\mathcal{S}$ yields the corresponding structured perturbation function:

$$\varepsilon_{s1}(x)=\max_{\|\delta\|_1\le\rho}\left\|(A_0+\sum_{i=1}^L A_i\delta_i)x-(b_0+\sum_{i=1}^L b_i\delta_i)\right\|_1,$$

which we can write as

$$\varepsilon_{s1}(x)=\max_{\|\delta\|_1\le\rho}\|F\,\delta-g\|_1 \tag{15}$$

with F and g defined as in (9).

In order to compute ε_{s1} and the associated worst case perturbation we will make use of the following lemma:

Lemma 1 *Let $F\in\mathbf{R}^{m\times L}$ and $g\in\mathbf{R}^m$. Let f_i be the columns of $[F\ \ -F]$. Then*

$$\max_{\|\delta\|_1\le\rho}\|F\,\delta-g\|_1\equiv\max_{i=j,\ldots,2L}\|(\rho\,f_i,g_i)\|_1.$$

A maximizing $\delta^\star$ is given by

$$\delta^\star=\alpha\,\rho\,e_{j_0},$$

$$j_0=\arg\max_{j=1,\ldots,2L}\|(\rho\,f_i-g)\|_1,\quad \alpha=\begin{cases}+1 & ;1\le j\le L\\ -1 & ;p<j\le 2L\end{cases}.$$

Proof: First we note that

$$\max_{\|\delta\|_1\le\rho}\|F\,\delta-g\|_1 \equiv \max_{\|\bar\delta\|_1\le 1}\|\rho\,F\,\bar\delta-g\|_1,$$

which is the maximization of a real valued convex function over a compact convex set. Therefore the maximum is attained at an extreme point of the set [7]. The set of extreme points of the set $\{\bar\delta\mid\|\bar\delta\|_1\le 1\}$ is $\mathcal{E}=\{\pm e_1,\ldots,\pm e_L\}$ where e_i are the columns of identity in $\mathbf{R}^L$. Hence

$$\begin{aligned}\max_{\|\bar\delta\|_1\le 1}\|\rho\,F\,\bar\delta-g\|_1 &= \max_{\bar\delta\in\mathcal{E}}\|\rho\,F\,\bar\delta-g\|_1\\ &= \max_{j=1,\ldots,2L}\|\rho\,f_j-g\|_1.\end{aligned}$$

The last expression is maximized by $\delta=\delta^\star$. ■

Applying the Lemma to (15) gives:

$$\boldsymbol{\varepsilon}_{s1}(x)=\max_{j=1,\ldots,2L}\|\rho\,f_j(x)-g(x)\|_1.$$

Using this expression and the definitions of f_j and g, we obtain the following result:

Theorem 6 *The structured worst case perturbation function can be computed as*

$$\boldsymbol{\varepsilon}_{s1}(x)=\max_{j=1,\ldots,2L}\|\tilde A_j\,x-\tilde b_j\|_1 \tag{16}$$

$$[\tilde A_j\ \tilde b_j]=\begin{cases}[(A_0+\rho A_j)\ (b_0+\rho b_j)] & ;1\le j\le L\\ [(A_0-\rho A_j)\ (b_0-\rho b_j)] & ;p<j\le 2L.\end{cases}$$

Thus structured uncertain linear approximation problem can be solved as:

$$\min_x\max_{j=1,\ldots,2L}\|\tilde A_j\,x-\tilde b_j\|_1 \tag{17}$$

Once again, this is a convex optimization problem in x which could be solved as an LP or using other specialized methods [8].

5 The Chebyshev Case

We will now obtain similar results for the space $(\mathbf{R}^n,\|\cdot\|_\infty)$. The corresponding norm induced on a matrix $A\in\mathbf{R}^{m\times n}$ is the "max-row-sum" norm:

$$\|A\|_\infty=\max_{\|x\|_\infty\le 1}\|A\,x\|_\infty=\max_{i=1,\ldots,m}\sum_{j=1}^n|a_{ij}|. \tag{18}$$

Solving approximation problems in the l_∞-norm may be preferable to the l_2-norm, when the perturbations are independent, ie, we have the constraint $\|\delta\|_\infty\le\rho$, rather than the constraint $\|\delta\|_2\le\rho$ which inherently couples the perturbations. In this case, the Euclidean formulation would minimize only an upper bound on the worst case residual, in general [3].

5.1 Solving the Minimax Formulation

Taking $\Omega=\mathcal{U}$ and $p=\infty$ in (1), $\boldsymbol{\varepsilon}$ becomes:

$$\boldsymbol{\varepsilon}_{u\infty}(x)=\max_{\|[\Delta A\ \Delta b]\|_\infty\le\rho}\|(A+\Delta A)x-(b+\Delta b)\|_\infty.$$

It follows from the triangle inequality and definition 18 that for any $[\Delta A\,\Delta b]$ in $\mathcal{U}$:

$$\|(A+\Delta A)x-(b+\Delta b)\|_\infty\le\|Ax-b\|_\infty+\rho\|(x,-1)\|_\infty. \tag{19}$$

Now let $x_{\rm aug}=(x,-1)$, $i_0=\arg\max_{i=1,\ldots,n+1}|{x_{\rm aug}}_i|$ and e_i be the ith column of the identity matrix. Then we can achieve equality in (19) with the following (nonunique) worst case perturbation:

$$[\Delta A\,\Delta b]_{\rm wc}=\begin{cases}\rho\,\mathrm{sgn}({x_{\rm aug}}_{i_0})\frac{Ax-b}{\|Ax-b\|_\infty}\,e_{i_0}^T & ;\text{ if } Ax-b\ne 0\\ \rho\,e_{i_0}e_{i_0}^T & ;\text{ otherwise.}\end{cases}$$

Thus we conclude that:

Theorem 7 *The unstructured worst case perturbation function can be computed as*

$$\boldsymbol{\varepsilon}_{u\infty}(x)=\|Ax-b\|_\infty+\rho\|(x,-1)\|_\infty$$

Therefore, the unstructured uncertain approximation problem can be solved as:

$$\min_x\{\|Ax-b\|_\infty+\rho\|(x,-1)\|_\infty\}, \tag{20}$$

which is a convex optimization problem which can be solved using linear programming or other specialized algorithms.

Taking $\Omega=\mathcal{S}$ in (1) yields:

$$\boldsymbol{\varepsilon}_{s\infty}(x)=\max_{\|\delta\|_\infty\le\rho}\left\|(A_0+\sum_{i=1}^L A_i\delta_i)x-(b_0+\sum_{i=1}^L b_i\delta_i)\right\|_\infty,$$

which we can write as

$$\boldsymbol{\varepsilon}_{s\infty}(x)=\max_{\|\delta\|_\infty\le\rho}\|F\,\delta-g\|_\infty \tag{21}$$

with F and g as in (9).

In order to compute $\boldsymbol{\varepsilon}_{s\infty}$ and the associated worst case perturbation we will make use of the following lemma

Lemma 2 *Let $F\in\mathbf{R}^{m\times L}$ and $g\in\mathbf{R}^m$. Let f_i^T be the rows of F, and g_i be the components of g. Then*

$$\max_{\|\delta\|_\infty\le\rho}\|F\,\delta-g\|_\infty\equiv\max_{i=1,\ldots,m}\|(\rho\,f_i,g_i)\|_1.$$

A maximizing $\delta^\star$ is given by

$$\delta^\star=-\rho\,\mathrm{sgn}(g_{i_0})\,\mathrm{sgn}(f_{i_0})\ \text{where}\ i_0=\arg\max_{i=1,\ldots,m}\|(\rho\,f_i,g_i)\|_1.$$

Proof: We first note that

$$\max_{\|\delta\|_\infty\le\rho}\|F\,\delta-g\|_\infty\equiv\max_{\|\bar\delta\|_\infty\le\rho}\|\rho\,F\,\bar\delta-g\|_\infty,$$

which is again the maximization of a real valued convex function over a compact convex set. Therefore the maximum is attained at an extreme point of the set [7]. The set of extreme points, $\mathcal{E}$, can be written as $\{\delta\ne 0\mid\delta_i\in\{-1,1\},i=1,\ldots,L\}$. Therefore:

$$\begin{aligned}\max_{\|\bar\delta\|_\infty\le 1}\|\rho\,F\,\bar\delta-g\|_\infty &= \max_{\bar\delta\in\mathcal{E}}\|\rho\,F\,\bar\delta-g\|_\infty\\ &= \max_{\bar\delta\in\mathcal{E}}\max_{i=1,\ldots,m}|\rho\,f_i^T\,\bar\delta-g_i|\\ &= \max_{i=1,\ldots,m}\max_{\bar\delta\in\mathcal{E}}|\rho\,f_i^T\,\bar\delta-g_i|.\end{aligned}$$

We complete the proof by noting that for each i, the inner maximum over $\mathcal{E}$ is $\|\rho f_i\|_1 + |g_i|$, which is achieved by the following choice of $\bar{\delta}$:

$$\bar{\delta} = -\mathrm{sgn}(g_i)\,\mathrm{sgn}(f_i).$$

■

The $\delta^\star$ above can be used to compute a worst case perturbation. Applying this lemma to (21) we have:

$$\varepsilon_{s\infty}(x) = \max_{i=1,\dots,m} \|(\rho f_i, g_i)\|_1. \tag{22}$$

Using the definitions of f_i and g_i given in (9) in (22), we obtain the following result:

Theorem 8 *The structured worst case perturbation function can be computed as*

$$\varepsilon_{s\infty}(x) = \max_{i=1,\dots,m} \|\tilde{A}_i\, x - \tilde{b}_i\|_1, \tag{23}$$

$$\tilde{A}_i^T = \left[a_{0,i}, \rho a_{1,i} \cdots \rho a_{L,i}\right], \qquad \tilde{b}_i^T = \left[b_{0,i}, \rho b_{1,i} \cdots \rho b_{L,i}\right],$$

where $a_{k,i}^T$ is the ith row of A_k, and $b_{k,i}$ is the ith component of b_k. Thus structured uncertain approximation problem can be solved as:

$$\min_x \max_{i=1,\dots,m} \|\tilde{A}_i\, x - \tilde{b}_i\|_1 \tag{24}$$

Once again, this is a convex optimization problem in x which could be solved as an LP or using other specialized methods [8].

6 System Identification Example

Consider the following simple system identification problem. Let H be a causal linear time invariant system, which obeys the linear ordinary difference equation:

$$\begin{aligned} y(k) &= -a_1 y(k-1) - \cdots - a_n y(k-n) \\ &\quad + b_1 u(k-1) + \cdots + b_n u(k-n). \end{aligned} \tag{25}$$

We would like to compute an estimate of the system's parameters $(a_1, \dots, a_n, b_1, \dots, b_n)$, in the presence of process and sensor noises, w and v, by driving it with a known input u sequence of length N and recording the output history $y = [H * (u + w)] + v$.

When the noises are white, there are several approaches for solving this problem, see [5] for example. Now suppose that all that is known about w and v is that they are independent and bounded by ρ_w and ρ_v, respectively. Then one might try the following approach: let $x = (\hat{a}_1, \dots, \hat{a}_n, \hat{b}_1, \dots, \hat{b}_n)$ denoted the parameter estimate and define

$$\begin{aligned} \hat{y}(k) &= -\hat{a}_1 y(k-1) - \cdots - \hat{a}_n y(k-n) \\ &\quad + \hat{b}_1 u(k-1) + \cdots + \hat{b}_n u(k-n). \end{aligned} \tag{26}$$

Find an x which minimizes the worst case peak absolute error, $|y(k) - \hat{y}(k)|$, over all possible perturbations of $u(k)$ and $y(k)$ of size ρ_w and ρ_v, respectively. This can be cast as a structured robust Chebyshev approximation problem: Let

$$e_i = \begin{cases} i\text{th column of } I_{N\times N} & ; \text{if } 1 \le i \le N, \\ \text{column of } N \text{ zeros} & ; \text{otherwise}, \end{cases}$$

where $I_{N\times N}$ is the $N \times N$ identity matrix. For any $v \in \mathbf{R}^N$, let $T_{N,n}(v)$ be the $N \times n$ Toeplitz matrix generated by the vector. Let $\vec{u}, \vec{y}_0$ and $\vec{y}_1$ be the vectors whose components are $u(0), \dots, u(N-1)$, $y(0), \dots, y(N-1)$, and $y(1), \dots, y(N)$, respectively. The desired estimate can now be computed by solving (24) with the following data:

$$[A_0 \mid b_0] = [-T_{N,n}(\vec{y}_0) \quad T_{N,n}(\vec{u}) \mid \vec{y}_1],$$

$$[A_i \mid b_i] = \begin{cases} \rho_w \; [\quad \mathbf{0} \qquad T_{N,n}(e_i) \mid \quad \mathbf{0} \quad] & ; 1 \le i \le N, \\ \rho_v \; [-T_{N,n}(e_{i-N}) \quad \mathbf{0} \mid e_{i-N-1}] & ; N < i \le 2N+1 \end{cases}$$

Remark 4 Traditionally, such problems are solved using prediction error methods and pseudo-linear regressions, which lead to non-convex optimization problems which are minimized using algorithms that might converge to a local minimum. In contrast, the robust approximation formulation leads to a convex optimization problem which is guaranteed to have a global minimum, that can be computed efficiently. Of course, convexity comes at the price of a tradeoff in accuracy and computational burden.

7 Conclusion

In this paper we have shown that robust linear approximation problems in $\mathbf{R}^n$ equipped with the l_1, l_2, and l_∞ norms can all be reduced to convex optimization problems, and hence can be solved efficiently in polynomial time [8]. The theory presented here is useful for desensitizing solutions to ill-contitioned problems, or for computing solutions that guarantee a certain performance in the presence of uncertainty in the data.

8 Acknowledgements

We would like to thank: Michael Saunders and Lieven Vandenberghe for their interest and useful suggestions, John Fox for his enthusiasm and support, and the controls group at Stanford Linear Accelerator Center for bringing this problem to our attention.

References

[1] A. Ben-Tal, A. Nemirovski, "Robust convex optimization," Technical report, December 1995, optimization lab, faculty of indust. eng. and management, Technion, Haifa 32000, Israel.

[2] S. Chen, D. Donoho, "Atomic decomposition by basis pursuit," technical report #479, May 1995, department of statistics, Stanford University.

[3] L. El-Ghaoui, H. Lebret, "Robust solutions to least-squares problems with uncertain data," submitted to SIAM Journal on Matrix Analysis and Applications.

[4] G. Golub, C. Van Loan, "An analysis of the total least squares problem," SIAM J. on Numerical Analysis, 17 1980, p883-893.

[5] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, 1987.

[6] M. Lobo, L. Vandenberghe, S. Boyd, H. Lebret, "Second-order cone programming: interior point methods and engineering applications," Linear Algebra and Appl., 1997. Submitted.

[7] R. Rockafellar, "Convex Analysis," Princeton Press, 1970.

[8] L. Vandenberghe, S. Boyd, *Convex Optimization*, book in preparation.

[9] G. White, T. Himel, H. Shoaee, " A Hybrid Numerical Method for Orbit Correction," IEEE Particle Accelerator Conference, 1997, Vancouver, Canada.