6.4 Robust approximation

6.4.1 Stochastic robust approximation

We consider an approximation problem with basic objective $\|Ax-b\|$, but also wish to take into account some uncertainty or possible variation in the data matrix A. (The same ideas can be extended to handle the case where there is uncertainty in both A and b.) In this section we consider some statistical models for the variation in A.

We assume that A is a random variable taking values in $\mathbf{R}^{m\times n}$, with mean $\bar{A}$, so we can describe A as

$$A = \bar{A} + U,$$

where U is a random matrix with zero mean. Here, the constant matrix $\bar{A}$ gives the average value of A, and U describes its statistical variation.

It is natural to use the expected value of $\|Ax-b\|$ as the objective:

$$\text{minimize} \quad \mathbf{E}\,\|Ax-b\|. \tag{6.13}$$

We refer to this problem as the *stochastic robust approximation problem.* It is always a convex optimization problem, but usually not tractable since in most cases it is very difficult to evaluate the objective or its derivatives.

One simple case in which the stochastic robust approximation problem (6.13) can be solved occurs when A assumes only a finite number of values, *i.e.*,

$$\mathbf{prob}(A = A_i) = p_i, \quad i = 1,\ldots,k,$$

where $A_i \in \mathbf{R}^{m\times n}$, $\mathbf{1}^T p = 1$, $p \succeq 0$. In this case the problem (6.13) has the form

$$\text{minimize} \quad p_1\|A_1x-b\| + \cdots + p_k\|A_kx-b\|,$$

which is often called a *sum-of-norms problem.* It can be expressed as

$$\begin{array}{ll} \text{minimize} & p^T t \\ \text{subject to} & \|A_ix-b\| \le t_i, \quad i=1,\ldots,k, \end{array}$$

where the variables are $x \in \mathbf{R}^n$ and $t \in \mathbf{R}^k$. If the norm is the Euclidean norm, this sum-of-norms problem is an SOCP. If the norm is the ℓ_1- or ℓ_∞-norm, the sum-of-norms problem can be expressed as an LP; see exercise 6.8.

Some variations on the statistical robust approximation problem (6.13) are tractable. As an example, consider the statistical robust least-squares problem

$$\text{minimize} \quad \mathbf{E}\,\|Ax-b\|_2^2,$$

where the norm is the Euclidean norm. We can express the objective as

$$\begin{array}{rcl} \mathbf{E}\,\|Ax-b\|_2^2 & = & \mathbf{E}(\bar{A}x-b+Ux)^T(\bar{A}x-b+Ux) \\ & = & (\bar{A}x-b)^T(\bar{A}x-b) + \mathbf{E}\,x^TU^TUx \\ & = & \|\bar{A}x-b\|_2^2 + x^TPx, \end{array}$$

where $P = \mathbf{E}\, U^T U$. Therefore the statistical robust approximation problem has the form of a regularized least-squares problem

$$\text{minimize} \quad \|\bar{A}x - b\|_2^2 + \|P^{1/2}x\|_2^2,$$

with solution

$$x = (\bar{A}^T\bar{A} + P)^{-1}\bar{A}^T b.$$

This makes perfect sense: when the matrix A is subject to variation, the vector Ax will have more variation the larger x is, and Jensen's inequality tells us that variation in Ax will increase the average value of $\|Ax - b\|_2$. So we need to balance making $\bar{A}x - b$ small with the desire for a small x (to keep the variation in Ax small), which is the essential idea of regularization.

This observation gives us another interpretation of the Tikhonov regularized least-squares problem (6.10), as a robust least-squares problem, taking into account possible variation in the matrix A. The solution of the Tikhonov regularized least-squares problem (6.10) minimizes $\mathbf{E}\,\|(A + U)x - b\|^2$, where U_{ij} are zero mean, uncorrelated random variables, with variance δ (and here, A is deterministic).

6.4.2 Worst-case robust approximation

It is also possible to model the variation in the matrix A using a set-based, worst-case approach. We describe the uncertainty by a set of possible values for A:

$$A \in \mathcal{A} \subseteq \mathbf{R}^{m\times n},$$

which we assume is nonempty and bounded. We define the associated *worst-case error* of a candidate approximate solution $x \in \mathbf{R}^n$ as

$$e_{\text{wc}}(x) = \sup\{\|Ax - b\| \mid A \in \mathcal{A}\},$$

which is always a convex function of x. The (worst-case) *robust approximation problem* is to minimize the worst-case error:

$$\text{minimize} \quad e_{\text{wc}}(x) = \sup\{\|Ax - b\| \mid A \in \mathcal{A}\}, \tag{6.14}$$

where the variable is x, and the problem data are b and the set $\mathcal{A}$. When $\mathcal{A}$ is the singleton $\mathcal{A} = \{A\}$, the robust approximation problem (6.14) reduces to the basic norm approximation problem (6.1). The robust approximation problem is always a convex optimization problem, but its tractability depends on the norm used and the description of the uncertainty set $\mathcal{A}$.

Example 6.5 *Comparison of stochastic and worst-case robust approximation.* To illustrate the difference between the stochastic and worst-case formulations of the robust approximation problem, we consider the least-squares problem

$$\text{minimize} \quad \|A(u)x - b\|_2^2,$$

where $u \in \mathbf{R}$ is an uncertain parameter and $A(u) = A_0 + uA_1$. We consider a specific instance of the problem, with $A(u) \in \mathbf{R}^{20\times 10}$, $\|A_0\| = 10$, $\|A_1\| = 1$, and u

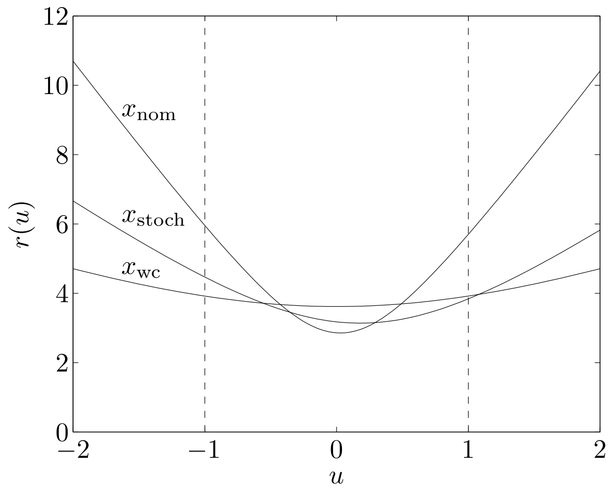

Figure 6.15 The residual $r(u) = \|A(u)x - b\|_2$ as a function of the uncertain parameter u for three approximate solutions x: (1) the nominal least-squares solution $x_{\rm nom}$; (2) the solution of the stochastic robust approximation problem $x_{\rm stoch}$ (assuming u is uniformly distributed on $[-1,1]$); and (3) the solution of the worst-case robust approximation problem $x_{\rm wc}$, assuming the parameter u lies in the interval $[-1,1]$. The nominal solution achieves the smallest residual when $u=0$, but gives much larger residuals as u approaches -1 or 1. The worst-case solution has a larger residual when $u=0$, but its residuals do not rise much as the parameter u varies over the interval $[-1,1]$.

in the interval $[-1,1]$. (So, roughly speaking, the variation in the matrix A is around $\pm 10\%$.)

We find three approximate solutions:

- *Nominal optimal.* The optimal solution $x_{\rm nom}$ is found, assuming $A(u)$ has its nominal value A_0.
- *Stochastic robust approximation.* We find $x_{\rm stoch}$, which minimizes $\mathbf{E}\,\|A(u)x - b\|_2^2$, assuming the parameter u is uniformly distributed on $[-1,1]$.
- *Worst-case robust approximation.* We find $x_{\rm wc}$, which minimizes

$$\sup_{-1\leq u\leq 1} \|A(u)x-b\|_2 = \max\{\|(A_0-A_1)x-b\|_2, \|(A_0+A_1)x-b\|_2\}.$$

For each of these three values of x, we plot the residual $r(u) = \|A(u)x-b\|_2$ as a function of the uncertain parameter u, in figure 6.15. These plots show how sensitive an approximate solution can be to variation in the parameter u. The nominal solution achieves the smallest residual when $u=0$, but is quite sensitive to parameter variation: it gives much larger residuals as u deviates from 0, and approaches -1 or 1. The worst-case solution has a larger residual when $u=0$, but its residuals do not rise much as u varies over the interval $[-1,1]$. The stochastic robust approximate solution is in between.

The robust approximation problem (6.14) arises in many contexts and applications. In an estimation setting, the set $\mathcal{A}$ gives our uncertainty in the linear relation between the vector to be estimated and our measurement vector. Sometimes the noise term v in the model $y = Ax + v$ is called *additive noise* or *additive error*, since it is added to the 'ideal' measurement Ax. In contrast, the variation in A is called *multiplicative error*, since it multiplies the variable x.

In an optimal design setting, the variation can represent uncertainty (arising in manufacture, say) of the linear equations that relate the design variables x to the results vector Ax. The robust approximation problem (6.14) is then interpreted as the robust design problem: find design variables x that minimize the worst possible mismatch between Ax and b, over all possible values of A.

Finite set

Here we have $\mathcal{A} = \{A_1, \dots, A_k\}$, and the robust approximation problem is

$$\text{minimize} \quad \max_{i=1,\dots,k} \|A_i x - b\|.$$

This problem is equivalent to the robust approximation problem with the polyhedral set $\mathcal{A} = \mathbf{conv}\{A_1, \dots, A_k\}$:

$$\text{minimize} \quad \sup\left\{\|Ax - b\| \mid A \in \mathbf{conv}\{A_1, \dots, A_k\}\right\}.$$

We can cast the problem in epigraph form as

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & \|A_i x - b\| \le t, \quad i = 1, \dots, k, \end{array}$$

which can be solved in a variety of ways, depending on the norm used. If the norm is the Euclidean norm, this is an SOCP. If the norm is the ℓ_1- or ℓ_∞-norm, we can express it as an LP.

Norm bound error

Here the uncertainty set $\mathcal{A}$ is a norm ball, $\mathcal{A} = \{\bar{A} + U \mid \|U\| \le a\}$, where $\|\cdot\|$ is a norm on $\mathbf{R}^{m\times n}$. In this case we have

$$e_{\mathrm{wc}}(x) = \sup\{\|\bar{A}x - b + Ux\| \mid \|U\| \le a\},$$

which must be carefully interpreted since the first norm appearing is on $\mathbf{R}^m$ (and is used to measure the size of the residual) and the second one appearing is on $\mathbf{R}^{m\times n}$ (used to define the norm ball $\mathcal{A}$).

This expression for $e_{\mathrm{wc}}(x)$ can be simplified in several cases. As an example, let us take the Euclidean norm on $\mathbf{R}^n$ and the associated induced norm on $\mathbf{R}^{m\times n}$, *i.e.*, the maximum singular value. If $\bar{A}x - b \ne 0$ and $x \ne 0$, the supremum in the expression for $e_{\mathrm{wc}}(x)$ is attained for $U = auv^T$, with

$$u = \frac{\bar{A}x - b}{\|\bar{A}x - b\|_2}, \qquad v = \frac{x}{\|x\|_2},$$

and the resulting worst-case error is

$$e_{\mathrm{wc}}(x) = \|\bar{A}x - b\|_2 + a\|x\|_2.$$

(It is easily verified that this expression is also valid if x or $\bar{A}x - b$ is zero.) The robust approximation problem (6.14) then becomes

$$\text{minimize} \quad \|\bar{A}x - b\|_2 + a\|x\|_2,$$

which is a regularized norm problem, solvable as the SOCP

$$\begin{array}{ll} \text{minimize} & t_1 + at_2 \\ \text{subject to} & \|\bar{A}x - b\|_2 \leq t_1, \quad \|x\|_2 \leq t_2. \end{array}$$

Since the solution of this problem is the same as the solution of the regularized least-squares problem

$$\text{minimize} \quad \|\bar{A}x - b\|_2^2 + \delta\|x\|_2^2$$

for some value of the regularization parameter δ, we have another interpretation of the regularized least-squares problem as a worst-case robust approximation problem.

Uncertainty ellipsoids

We can also describe the variation in A by giving an ellipsoid of possible values for each row:

$$\mathcal{A} = \{[a_1 \ \cdots \ a_m]^T \mid a_i \in \mathcal{E}_i, \ i = 1, \ldots, m\},$$

where

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \leq 1\}.$$

The matrix $P_i \in \mathbf{R}^{n \times n}$ describes the variation in a_i. We allow P_i to have a nontrivial nullspace, in order to model the situation when the variation in a_i is restricted to a subspace. As an extreme case, we take $P_i = 0$ if there is no uncertainty in a_i.

With this ellipsoidal uncertainty description, we can give an explicit expression for the worst-case magnitude of each residual:

$$\begin{array}{rcl} \sup_{a_i \in \mathcal{E}_i} |a_i^T x - b_i| & = & \sup\{|\bar{a}_i^T x - b_i + (P_i u)^T x| \mid \|u\|_2 \leq 1\} \\ & = & |\bar{a}_i^T x - b_i| + \|P_i^T x\|_2. \end{array}$$

Using this result we can solve several robust approximation problems. For example, the robust ℓ_2-norm approximation problem

$$\text{minimize} \quad e_{\text{wc}}(x) = \sup\{\|Ax - b\|_2 \mid a_i \in \mathcal{E}_i, \ i = 1, \ldots, m\}$$

can be reduced to an SOCP, as follows. An explicit expression for the worst-case error is given by

$$e_{\text{wc}}(x) = \left(\sum_{i=1}^m \left(\sup_{a_i \in \mathcal{E}_i} |a_i^T x - b_i| \right)^2 \right)^{1/2} = \left(\sum_{i=1}^m (|\bar{a}_i^T x - b_i| + \|P_i^T x\|_2)^2 \right)^{1/2}.$$

To minimize $e_{\text{wc}}(x)$ we can solve

$$\begin{array}{ll} \text{minimize} & \|t\|_2 \\ \text{subject to} & |\bar{a}_i^T x - b_i| + \|P_i^T x\|_2 \leq t_i, \quad i = 1, \ldots, m, \end{array}$$

where we introduced new variables $t_1, \ldots, t_m$. This problem can be formulated as

$$\begin{array}{ll} \text{minimize} & \|t\|_2 \\ \text{subject to} & \bar{a}_i^T x - b_i + \|P_i^T x\|_2 \le t_i, \quad i = 1, \ldots, m \\ & -\bar{a}_i^T x + b_i + \|P_i^T x\|_2 \le t_i, \quad i = 1, \ldots, m, \end{array}$$

which becomes an SOCP when put in epigraph form.

Norm bounded error with linear structure

As a generalization of the norm bound description $\mathcal{A} = \{\bar{A} + U \mid \|U\| \le a\}$, we can define $\mathcal{A}$ as the image of a norm ball under an affine transformation:

$$\mathcal{A} = \{\bar{A} + u_1 A_1 + u_2 A_2 + \cdots + u_p A_p \mid \|u\| \le 1\},$$

where $\|\cdot\|$ is a norm on $\mathbf{R}^p$, and the $p + 1$ matrices $\bar{A},\ A_1, \ldots, A_p \in \mathbf{R}^{m \times n}$ are given. The worst-case error can be expressed as

$$\begin{array}{rcl} e_{\text{wc}}(x) & = & \sup_{\|u\| \le 1} \|(\bar{A} + u_1 A_1 + \cdots + u_p A_p)x - b\| \\ & = & \sup_{\|u\| \le 1} \|P(x)u + q(x)\|, \end{array}$$

where P and q are defined as

$$P(x) = \left[\begin{array}{cccc} A_1 x & A_2 x & \cdots & A_p x \end{array} \right] \in \mathbf{R}^{m \times p}, \qquad q(x) = \bar{A}x - b \in \mathbf{R}^m.$$

As a first example, we consider the robust Chebyshev approximation problem

$$\text{minimize} \quad e_{\text{wc}}(x) = \sup_{\|u\|_\infty \le 1} \|(\bar{A} + u_1 A_1 + \cdots + u_p A_p)x - b\|_\infty.$$

In this case we can derive an explicit expression for the worst-case error. Let $p_i(x)^T$ denote the ith row of $P(x)$. We have

$$\begin{array}{rcl} e_{\text{wc}}(x) & = & \sup_{\|u\|_\infty \le 1} \|P(x)u + q(x)\|_\infty \\ & = & \max_{i=1,\ldots,m} \sup_{\|u\|_\infty \le 1} |p_i(x)^T u + q_i(x)| \\ & = & \max_{i=1,\ldots,m} (\|p_i(x)\|_1 + |q_i(x)|). \end{array}$$

The robust Chebyshev approximation problem can therefore be cast as an LP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -y_0 \preceq \bar{A}x - b \preceq y_0 \\ & -y_k \preceq A_k x \preceq y_k, \quad k = 1, \ldots, p \\ & y_0 + \sum_{k=1}^p y_k \preceq t\mathbf{1}, \end{array}$$

with variables $x \in \mathbf{R}^n$, $y_k \in \mathbf{R}^m$, $t \in \mathbf{R}$.

As another example, we consider the robust least-squares problem

$$\text{minimize} \quad e_{\text{wc}}(x) = \sup_{\|u\|_2 \le 1} \|(\bar{A} + u_1 A_1 + \cdots + u_p A_p)x - b\|_2.$$

Here we use Lagrange duality to evaluate $e_{\rm wc}$. The worst-case error $e_{\rm wc}(x)$ is the squareroot of the optimal value of the (nonconvex) quadratic optimization problem

$$\begin{array}{ll} \mbox{maximize} & \|P(x)u + q(x)\|_2^2 \\ \mbox{subject to} & u^T u \leq 1, \end{array}$$

with u as variable. The Lagrange dual of this problem can be expressed as the SDP

$$\begin{array}{ll} \mbox{minimize} & t + \lambda \\ \mbox{subject to} & \left[\begin{array}{ccc} I & P(x) & q(x) \\ P(x)^T & \lambda I & 0 \\ q(x)^T & 0 & t \end{array}\right] \succeq 0 \end{array} \tag{6.15}$$

with variables t, $\lambda \in \mathbf{R}$. Moreover, as mentioned in §5.2 and §B.1 (and proved in §B.4), strong duality holds for this pair of primal and dual problems. In other words, for fixed x, we can compute $e_{\rm wc}(x)^2$ by solving the SDP (6.15) with variables t and λ. Optimizing jointly over t, λ, and x is equivalent to minimizing $e_{\rm wc}(x)^2$. We conclude that the robust least-squares problem is equivalent to the SDP (6.15) with x, λ, t as variables.

Example 6.6 *Comparison of worst-case robust, Tikhonov regularized, and nominal least-squares solutions.* We consider an instance of the robust approximation problem

$$\mbox{minimize} \quad \sup_{\|u\|_2 \leq 1} \|(\bar{A} + u_1 A_1 + u_2 A_2)x - b\|_2, \tag{6.16}$$

with dimensions $m = 50$, $n = 20$. The matrix $\bar{A}$ has norm 10, and the two matrices A_1 and A_2 have norm 1, so the variation in the matrix A is, roughly speaking, around 10%. The uncertainty parameters u_1 and u_2 lie in the unit disk in $\mathbf{R}^2$.

We compute the optimal solution of the robust least-squares problem (6.16) $x_{\rm rls}$, as well as the solution of the nominal least-squares problem $x_{\rm ls}$ (*i.e.*, assuming $u = 0$), and also the Tikhonov regularized solution $x_{\rm tik}$, with $\delta = 1$.

To illustrate the sensitivity of each of these approximate solutions to the parameter u, we generate 10^5 parameter vectors, uniformly distributed on the unit disk, and evaluate the residual

$$\|(A_0 + u_1 A_1 + u_2 A_2)x - b\|_2$$

for each parameter value. The distributions of the residuals are shown in figure 6.16.

We can make several observations. First, the residuals of the nominal least-squares solution are widely spread, from a smallest value around 0.52 to a largest value around 4.9. In particular, the least-squares solution is very sensitive to parameter variation. In contrast, both the robust least-squares and Tikhonov regularized solutions exhibit far smaller variation in residual as the uncertainty parameter varies over the unit disk. The robust least-squares solution, for example, achieves a residual between 2.0 and 2.6 for all parameters in the unit disk.

6.5 Function fitting and interpolation

In function fitting problems, we select a member of a finite-dimensional subspace of functions that best fits some given data or requirements. For simplicity we