ROBUST APPROXIMATION

Seminar Optimierung

Nishakova Irina

CONTENTS

Robust approximation

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 or 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:

$$\text{minimize } \mathbf{E}\,\|Ax - b\|. \tag{1.1}$$

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.

Sum-of-norms problem

One simple case in which the stochastic robust approximation problem (1.1) 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 $A_i \in \mathbf{R}^{m\times n}, \mathbf{1}^T p = 1, p \geq 0$. In this case the problem (1.1) has the form

$$\text{minimize } p_1\|A_1 x - b\| + \cdots + p_k\|A_k x - b\|,$$

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

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

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.

Statistical robust least-squares problem

Some variations on the statistical robust approximation problem (1.1) are tractable. As an example, consider the statistical robust least-squares problem

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

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

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

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

$$\text{minimize } \|\bar{A}x-b\|_2^2+\left\|P^{1/2}x\right\|_2^2, \qquad (1.2)$$

with solution

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

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.

Interpretation of the Tikhonov regularization problem

The Tikhonov regularized least-squares problem

$$\text{minimize } \|Ax-b\|_2^2+\delta\|x\|_2^2 \qquad (1.3)$$

has the analytical solution

$$x=(A^TA+\delta I)^{-1}A^Tb.$$

The statistical robust least-squares problem in the form of a regularized least-squares problem (1.2) gives us another interpretation of the Tikhonov regularization problem (1.3) as a robust least-squares problem, taking into account possible variation in the matrix A.

Consideration of $P=\mathbf{E}\,U^TU$ (where U is a random matrix with zero mean, so, $\mathbf{E}\,u_{ij}=0$) gives us:

$$\begin{aligned}\mathbf{E}\,(U^TU)_{kl} &= \mathbf{E}\,\Big(\sum_{i=1}^{m} {u_{ki}}^T\,u_{il}\Big)=\mathbf{E}\,\Big(\sum_{i=1}^{m} u_{ik}\,u_{il}\Big)\\ &= \sum_{i=1}^{m}\mathbf{E}\,(u_{ik}\,u_{il})=\delta_{kl}\sum_{i=1}^{m}\mathbf{E}\,(u_{ik})^2.\end{aligned}$$

If the variance of u_{ij} is equal to δ/m, $i.e.$

$$\mathbf{V}\, u_{ij} = \mathbf{E}\,(u_{ij})^2 - (\mathbf{E}\, u_{ij})^2 = \mathbf{E}\,(u_{ij})^2 = \delta/m,$$

we get:

$$\mathbf{E}\,(U^T U)_{kl} = \delta_{kl} \sum_{i=1}^{m} \mathbf{E}\,(u_{ik})^2 = \delta_{kl} m\, \mathbf{E}\,(u_{ij})^2 = (\delta I)_{kl}.$$

Thus, the solution of the Tikhonov regularized least-squares problem (1.3) minimizes $\mathbf{E}\, \|(\bar{A} + U)x - b\|_2^2$, where u_{ij} are zero mean, uncorrelated random variables, with variance δ/m (and here, A is deterministic).

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

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

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

$$\text{minimize } e_{\mathrm{wc}}(x) = \sup\{\|Ax - b\| \mid A \in \mathcal{A}\}, \qquad (2.1)$$

where the variable is x, and the problem data are band the set $\mathcal{A}$. When $\mathcal{A}$ is the singleton $\mathcal{A} = \{A\}$, the robust approximation problem (2.1) reduces to the basic norm approximation problem

$$\text{minimize } \|Ax - b\|.$$

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: 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

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

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

We find three approximate solutions:

- *Nominal optimal.* The optimal solution x_{nom} is found, assuming $A(u)$ has its nominal value A_0:

$$x_{nom} = \left({A_0}^T A_0\right)^{-1} {A_0}^T b.$$

- *Stochastic robust approximation.* We find x_{stoch}, which minimizes $\mathbf{E}\, \|A(u)x - b\|_2^2$, assuming the parameter u is uniformly distributed on $[-1, 1]$. Consider $P = \mathbf{E}\ \mathrm{U^T U}$, taking into account that $U = uA_1$:

$$\begin{aligned} P = \mathbf{E}\ \mathrm{U^T U} = {} & \mathbf{E}\ (uA_1)^{\mathrm{T}} uA_1 \\ & = \mathbf{E}\ (u^2 {A_1}^T A_1) = \mathbf{E}\ (u^2) {A_1}^T A_1 \\ & = \frac{1}{3} {A_1}^T A_1. \end{aligned}$$

The solution is:

$$x_{stoch} = \left({A_0}^T A_0 + \tfrac{1}{3} {A_1}^T A_1\right)^{-1} {A_0}^T b.$$

- *Worst-case robust approximation.* The solution x_{wc} minimizes

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

This is an SOCP since it is equivalent to:

$$\begin{aligned} & \text{minimize } t \\ & \text{subject to } \|(A_0 - A_1)x - b\|_2 \le t, \\ & \qquad\qquad\quad \|(A_0 + A_1)x - b\|_2 \le t. \end{aligned}$$

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 1. 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.

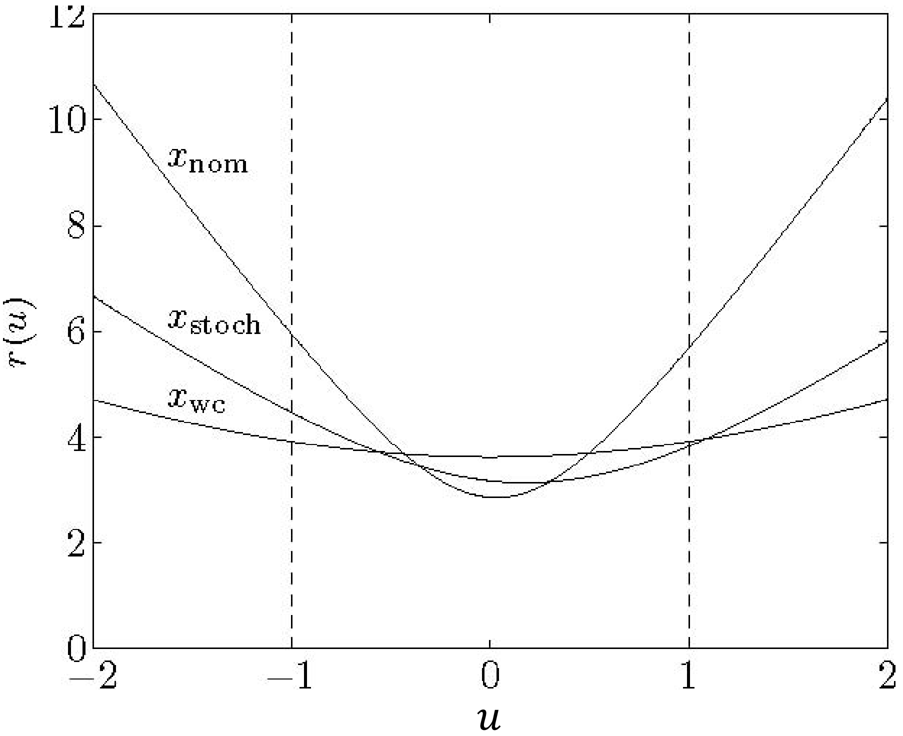

Figure 1. The residual $r(u) = ||A(u)x - b||_2$ as a function of the uncertain parameter u for three approximate solutions x:

(1) the nominal least-squares solution x_{nom};

(2) the solution of the stochastic robust approximation problem x_{stoch} (assuming u is uniformly distributed on $[-1, 1]$);

(3) the solution of the worst-case robust approximation problem x_{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 $[-1, 1]$.

The robust approximation problem (1.1) 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 (1.1) is then interpreted as the robust design problem: find design variables x that minimize the worst possible mismatch between Ax and b, overall possible values of A.

Finite set
Here we have $\mathcal{A} = \{\mathrm{A}_1, \ldots, \mathrm{A}_k\}$, and the robust approximation problem is

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

This problem is equivalent to the robust approximation problem with the polyhedral set $\mathcal{A} = \mathbf{conv}\{\mathrm{A}_1, \ldots, \mathrm{A}_k\}$:

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

We can cast the problem in epigraph form as

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

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, the sum-of-norms problem can be expressed 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

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

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 $\mathcal{A}$).

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

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

where

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

The matrix $P_i \in \mathbf{R}^{n\times n}$ describes the variation in a_i.

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.