Nonlinear Optimization Project
Chapter 6 Overview
Approximation and Fitting

Sean Woodham, Charly Stephens-Cohen, Fei Yang, Maxim Zalutskiy

04/16/ 13

1 Introduction

The purpose of this overview is to explain the basic concepts and ideas behind different types of approximation. Different examples were shown and generated through use of convex functions along with regularization methods, fittings and approximations. Here, conditons and properties of the approximations were also examined in order optimize the best fit possible for each of the optimal values.

2 Norm Approximation

2.1 Unconstrained Norm Approximation Problem

One of the simplest forms of a norm approximation problem is an unconstrained problem that can be represented in the form, minimize $\|Ax-b\|$. In this form A and b are the problem data where $A \in R^{m\times n}$ and $b \in R^m$, $\|.\|$ is the norm on R^m and $x \in R^n$ is the variable.We refer to the solution of the norm approximation problem as an approximate solution of $Ax \approx b$ in the norm $\|.\|$ and the vector $r = Ax - b$ is called the residual for the problem.

Properties of the Norm Approximation Problem are as follows:

1) It is always convex and solvable.
2) Its optimal value is zero if and only if $b \in R(A)$.
3) For the cases where $b \notin R(A)$, the columns of A are linearly independent and in the matrix A, $m > n$.

Interpretation of the solution

There are several ways to interpret the solution $\hat{x} = argmin_z \|Az - y\|$ to the norm approximation problem. These are as follows:

- ***Geometric***: The $A\hat{x}$ is the one belonging to $A = R(A)$ that is the closest to b in the norm $\|.\|$ .i.e it is the projection of b onto the linear subspace A.

- ***Regression***: If we express Ax as $Ax = x_1 a_1 + \cdots x_n a_n$, where $a_1, \cdots, a_n \in R^n$ are the columns of A then the approximation problem is to find the best fit of the vextor b by a linear combination of columns A. If we look at $a-1, \cdots, a_n$ as the regressors then we call the vector $A\hat{x}$ the regression of b.

- ***Estimation***: If we consider a linear measurement model of the form $y = Ax + v$ where $y \in R^n$ is the vector measurement and $v \in R^n$ is an unknown measurement error or a possible noise. The estimation problem is to make a guess as to what x may be given some y. Given the fact that $y = b$, the most obvious choice of x is $\hat{x}$.

- ***Optimal Design***: The components of x are input design variables to a linear system with the output $y = Ax$. Here, we can see that $\hat{x}$ gives us a result of $\hat{y} = A\hat{x}$ which approximates the result that is desired.

Examples of the Norm Approximation Problem

- *Weighted Norm Approximation*: Minimize $W(Ax - b)\|$, where $W \in R^{m\times m}$ is a weighting matrix that is used to emphasize some residuals over different components.

- *Least-squares approximation*: Minimize $\| Ax - b \|_2^2 = r_1^2 + \cdots r_m^2$. The solution $\hat{x}$ is optimal only if it satisfies the normalequations $A^T Ax = A^T b$. Since we can assume that the columns of A are independent then the least squares problem has a unique solution $\hat{x} = (A^T A)^{-1} A^T$.

2.2 Least-Norm Problem

The most basic least norm problem has the form minimize $\| x \|$ subject to $Ax = b$, where $A \in R^{m\times n}$ and $b \in R^m$ are the problem data and $x \in R^n$ is the variable.

Properties of the Least-Norm Problems are as follows:

1)The problem only becomes interesting when $m < n$.

2)With x_0 as any solution to $Ax = b$ and $Z \in R^{n\times k}$ as any matrix whose columns are a basis for the nullspace of A. It can be rewritten as $\| x_0 + Zu \|$ with the variable $u \in R^k$.

Interpretations of the solution $x^* = argmin \| x \|$:

- ***Geometric***: x^* is the point in the affine set $\{x \mid Ax = b\}$ with the minimum distance to the origin 0.
- ***Estimation***: $b = Ax$ are given measurements of $x; x^*$ is the smallest estimate consistent with the measurements.

Examples of the Least Norm Problems

Least squares solution of linear equations: Minimize $\| x \|_2^2$ subject to $Ax = b$. This can be solved analytically by $2x^* + A^T v^* = 0, Ax^* = b$ which is a pair of linear equations and can be readily solved. From the first equation we can see that $x^* = -(1/2)A^T v^*$, then substituting this into the second equation we get $-(1/2)AA^T v^* = b$, and conclude that

$$v^* = -2(AA^T)^{-}1b \text{ and } x^* = A^T(AA^T)^{-}1b.$$

Now that we have discussed Norm approximation examples and ideas we can move on to Regularized approximation.

3 Regularized approximation

3.1 Bi-criterion formulation

The basic goal of regularized approximation is to find a good approximation of $Ax \approx b$ with a small vector x. The bi-criterion problem is a vector optimization problem with two objectives, $\|Ax - b\|$ on R^m and $\|x\|$ on R^m, which are normally minimized. The first norm measures the size of the residual and the second measures the size of x. The optimal trade-off curve of $\|Ax - b\|$ versus $\|x\|$ shows how large one of the objectives must be to have the other one small. The optimal design uses a small x which makes solving this problem more efficient or using the linear model $y = Ax$ for only valid small x. The solution for this optimization is given by $x = A^{-1}b$.

3.2 Regularization

Regularization is a method used to solve the bi-criterion problem by adding an extra term or parameter associated with the norm of x. One form of regularization is:

$$\textit{minimize } \|Ax - b\| + \gamma\|x\| \text{ where } \gamma > 0.$$

Another common form is:

$$\textit{minimize } \|Ax - b\|^2 + \delta\|x\|^2 \text{ where } \delta > 0.$$

Tikhonov regularization

The most common form of regularization, used with Euclidean norms, results in a quadratic optimization problem:

$$\textit{minimize } \|Ax - b\|_2^2 + \delta\|x\|_2^2 = x^T(A^TA + \delta \mathrm{I})\mathrm{x} - 2b^TAx + b^Tb \text{ where } \delta > 0.$$

This can be solved as a least-squares problem:

$$\textit{minimize } \left\| \begin{bmatrix} A \\ \sqrt{\delta}I \end{bmatrix} x - \begin{bmatrix} b \\ 0 \end{bmatrix} \right\|_2^2,$$

which gives the solution $x = (A^TA + \delta \mathrm{I})^{-1}A^Tb$.

Smoothing regularization

This form of regularization adds the matrix D, which represents the approximate second-order differentiation operator term. So instead of using $\|x\|$, that is replaced by $\|Dx\|$ and this measures the smoothness of x.

ℓ1-norm regularization
Regularization that uses an ℓ1-norm can be instrumental for finding a sparse solution. Consider the problem:

$$minimize\ \|Ax - b\|_2 + \gamma\|x\|_1 \ where\ \gamma > 0.$$

By varying the parameter γ we use to approximate the optimal trade-off curve between $\|Ax - b\|_2$ and cardinality card(x) of the vector x, i.e., the number of nonzero elements.

3.3 Reconstruction, smoothing, and de-noising
Another regularization method is reconstruction which performs differently than the others described. A reconstruction problem starts with a signal represented by a vector $x \in R^n$. The coefficients x_i correspond to the value of some function of time and it is usually assumed that we have $x_i \approx x_{i+1}$. The signal x is corrupted by an additive noise v giving by

$$x_{cor} = x + v,$$

which is where the regularization method can be seen. The noise can be represented in a number of ways, as in it can be unknown, small, and rapidly varying. Signal reconstruction or de-noising, is the process of forming an estimate $\hat{x}$ of the original signal x and smoothing happens as an operation on x_{cor} to produce $\hat{x}$. The reconstruction problem can be formed using the bi-criterion problem between $\|\hat{x} - x_{cor}\|_2$ on R^n and $\varphi(\hat{x})$ which is the regularization function from R^n to R. This function is used to It is meant to measure the lack of smoothness of the estimate $\hat{x}$.

Quadratic smoothing
The simplest reconstruction method uses the quadratic smoothing function:

$$\varphi_{quad}(x) = \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

where $D \in R^{(n-1)\times n}$ is the bidiagonal matrix. We can then form the bi-criterion problem between $\|\hat{x} - x_{cor}\|_2$ and $\|D\hat{x}\|_2$ by minimizing $\|\hat{x} - x_{cor}\|_2^2 + \delta\|D\hat{x}\|_2^2$, where δ > 0, which gives the solution of:

$$\hat{x} = (I + \delta D^T D)^{-1} x_{cor}$$

Total variation reconstruction
The total variation method is used when the original signal is very smooth, and the noise is rapidly varying. So this method can remove much of the noise and still preserve occasional rapid variations in the original signal. The total variation of $x \in R^n$ method is based on the smoothing function:

$$\varphi_{tv}(\hat{x}) = \sum_{i=1}^{n-1} |\hat{x}_{i+1} - \hat{x}_i| = \|D\hat{x}\|_1$$

4 Robust Approximation

Recall the simplest norm approximation problem we have discussed above, it is always desirable to take the possible variation of the matrix $A \in \mathrm{R}^{m\times n}$ into consideration in practice, which is called the robust approximation.

4.1 Two kinds of Robust Approxiamtion
Stochastic Robust Approximation Problem
We can decompose the random matrix A in$\mathrm{R}^{m\times n}$ in the following way,

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

where $\bar{A} = \mathrm{E}(A)$ and U is a random matrix in $\mathrm{R}^{m\times n}$with zero mean. Based on the uncertainty, we have a nature sense to use the expected valued of $||Ax - b||$ as our objective function, as showed below, and we call this kind of problem as the *stochastic robust approximation problem*.

$$\text{minimize} \quad \mathrm{E}||Ax - b||.$$

Worst-Case Robust Approximation Problem
The 2nd way to model the uncertainty in matrix A is to use a set-based, worst-case approach. We develop a nonempty and bounded set $\mathcal{A}$, s.t. $A \in \mathcal{A} \subseteq \mathrm{R}^{m\times n}$, and use the associated *worst-case error* of a possible solution $x \in R^n$ as the following as our objective function,

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

Thus, the worst-case robust approximation problem is,

$$minimize \qquad e_{wc}(x) = sup\{||Ax - b|| \; |A \in \mathcal{A}\}.$$

4.2 Properties of SRAP and WCRAP
1) Both are always convex problems.
2) The majority of SRAP are intractable due to the difficulty in evaluating the objective and its derivative.
3) The tractability of WCRAP depends on the norm used and the uncertainty set $\mathcal{A}$.

4.3 Examples and Relations to Special Problem

- For the *least-square problem of SRAP*, if we denote $P = \mathrm{E}[U^T U] = \mathrm{E}[A^T A]$, we have,

  $$minimize \; ||\bar{A}x - b||_2^2 + ||\sqrt{P}x||_2^2 \,,$$

  which has a solution,

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

  Thus, when $\delta I = P$, the *Tikhonov regularized problem* minimizes this problem.

- Another well-known example of a tractable SRAP is *sum-of-norm problem*. In that case, A have only finite values, that is,

$$\Pr\,(A = A_i) \ = p_i\ , \quad i = 1, \cdots\cdots, k.$$

where $A_i \in \mathrm{R}$, $1^T p = 1$, $p \succcurlyeq 0$. And SRAP becomes,

$$minimize \quad p_1||A_1 x - b|| + \cdots\cdots + p_k||A_k x - b||,$$

which can be converted into a *SOCP*, if the Euclidean norm is used, and a *LP*, if l_∞-norm or l_1-norm are used.

- If $\mathcal{A} = \{\bar{A} + U|\ ||U|| \leq a\}$ is a norm ball, where $||\bullet||$ is a norm on $\mathrm{R}^{m\times n}$, the problem is called *norm bound error* and the objective function can be simplified for some particular norms, i.e., if the Euclidean norm on R^m and the maximum singular value norm on $\mathrm{R}^{m\times n}$ is used, the WCRAP becomes,

$$minimized \qquad ||\bar{A}x - b||_2 + a||x||_2,$$

In this case, we can interpret the *regularized least-squares problem* as a WCRAP.

- If the set $\mathcal{A}$ is a *finite set*, $\mathcal{A} = \{A_1, \cdots\cdots, A_k\}$ or $\mathcal{A} = \mathrm{conv}(\{A_1, \cdots\cdots, A_k\})$, the WCRAP turns out to be,

$$minimize \qquad max\, ||A_i x - b||.$$

which is equivalent to the following if the epigraph form is used,

$$\begin{array}{c} minimize \qquad t \\ \text{subject to } ||A_i x - b|| \leq t, \quad i = 1, \cdots\cdots, k. \end{array}$$

This problem turns out to be a *SOCP* if the Euclidean norm is used and a *LP* if l_∞-norm or l_1-norm are used.

5 Function fitting and interpolation

Function fitting deals with selecting the function that best fits some given set of data points. It is always convenient to combine all possible function choices into a set of *basis functions*:

$$f(u) = x_1 f_1(u) + \cdots + x_n f_n(u), \tag{1}$$

where $f_1, \dots, f_n\colon \mathbf{R}^k \to \mathbf{R}$, with common domain $\mathbf{dom}\, f_i = D$, with $x \in \mathbf{R}^n$. The family $\{f_1, .., f_n\}$ can be:

- polynomials
- trigonometric functions
- splines

etc.

There is a number of constrains that can be imposed on (1)

- ***interpolation conditions***

$$f(u_i) = y_i, \qquad i = 1, \dots, m,$$

  where $z_i \in R$ are specific points from function domain.

- ***Lipschitz constraint***

$$|f(u_j) - f(u_k)| \leq L\|u_j - u_k\|$$

- ***Nonnegativity***
- ***derivative/integral***

  etc.

The common fitting/interpolation problems include minimum norm function fitting, basis pursuit, fitting with a convex functions as well as combinations and/or variations thereof. Assume the data are given in the form:

$$(u_1, y_1), \dots, (u_m, y_m),$$

where $u_i \in D$ and $y_i \in \mathbf{R}$.

Minimum norm function fitting is a simple least-squares problem with norm of the error being minimized:

$$\text{minimize } \sum_{i=1}^{m}(f(u_i) - y_i)^2 \tag{2}$$

When dealing with a large number of basis function a technique known as *basis pursuit* can be used. It allows to limit the number of fit parameters to effectively decrease the size of the problem. The problem is formulated as:

$$\text{minimize } \sum_{i=1}^{m}(f(u_i) - y_i)^2 + \gamma\|x\|_1, \tag{3}$$

where $\gamma > 0$ is a parameter used to trade off the quality of fit to the data and the sparcity of coefficent vector. The problem (3) is solved for x initially, which is then used to solve the least-squares problem similar to (2).

The set of convex functions is an important class of functions in the field of optimization so it is sometimes needed to fit a convex function to a given set of data points. Using the least-square problem as an example, convex function fitting problem can be written as a constrained quadratic program:

$$\text{minimize } \sum_{i=1}^{m}\left(y_i - f(u_i)\right)^2$$
$$\text{subject to } f: \mathbf{R}^{\mathbf{k}} \to \mathbf{R} \text{ is convex}$$

Bibliography

1. http://www.stanford.edu/~boyd/cvxbook/bv_cvxbook.pdf, chapter 6, Apri13,2013.
2. *Robust Approximation*. Nishakova Irina, Seminar Optimierung
3. *Applications of Convex Optimization in Signal Processing and Communications*: Lecture 14. Andre Tkacenko, Signal Processing Research Group Jet Propulsion Laboratory. May 17, 2012.
4. http://en.wikipedia.org/wiki/Flat_spline, April 13,2013
5. http://www.systems.caltech.edu/dsp/ee150_acospc/lectures/EE_150_Lecture14 April 2013