PART II : Least-Squares Approximation

Basic theory

Let U be an inner product space. Let V be a subspace of U. For any $g \in U$, we look for a least-squares approximation of g in the subspace V

$$\min_{f \in V} \|f - g\|_2,$$

where $\|\cdot\|_2$ is the 2-norm induced by the inner product. Denote by $V^\perp$ the subspace of U that is orthogonal to V. By the problem statement we have the following knowledge of least-squares(LS) approximation.

> Decompose g into two orthogonal components
>
> $$g = g_1 + g_2, \quad g_1 \in V, \quad g_2 \in V^\perp.$$
>
> Then $f = g_1$ is the least-squares approximation in V to g in U, and the approximation error or *residual* is
>
> $$r \stackrel{\text{def}}{=} g - f = g_2.$$

Proof. By the decomposition of g, we have the orthogonal decomposition of $g - f$ for any $f \in V$ as well

$$g - f = (g_1 - f) + g_2, \quad g_1 - f \in V, \quad f_2 \in V^\perp.$$

Thus,

$$\|g - f\|_2^2 = \|g_1 - f\|_2^2 + \|g_2\|_2^2 \geq \|g_2\|_2^2.$$

The second equality is achieved at $f = g_1$ only. ■

In other words the LS approximation is the orthogonal projection of g into the subspace V, and the residual is the orthogonal projection of g into the subspace $V^\perp$. We note by passing that zero

tolerance on approximation errors is only possible for functions already in V.

LS SOLUTION IN A SUBSPACE OF FINITE DIMENSION

In computational practice we choose V a subspace of finite dimension. Let $\{f_k \,|\, k = 1 : n\}$ be a basis of V. Then the LS approximation problem becomes problem of coefficient determination

$$\min_a \|g - \sum_{k=1}^{n} \alpha_k f_k\|_2, \quad a = [\alpha_1, \cdots, \alpha_n]^{\mathrm{T}}.$$

We introduce a couple of approaches for the LS solution in a finite-dimension space. The first approach is traditional. By the above approximation properties, the LS solution satisfy the following equations

$$\langle f_i, g\rangle = \sum_{k=1}^{n} \alpha_k \langle f_i, f_k\rangle, \quad i = 1 : n.$$

which in matrix form is

$$Ga = b, \quad G = [\langle f_i, f_j\rangle]_{i,j=1:n}, \quad b = [\langle f_i, g\rangle]_{i=1:n}.$$

The equations are known as the *normal equations* for the LS solution. The matrix is known as the *Gram* matrix associated with the inner product and the basis functions.

Note that we have changed the minimization problem to the problem of solving the system of normal equations. The Gram matrix is determined by the chosen inner product and the basis functions, and hence may be computed once for all. In contrast, the right hand vector depends also on the function g to be approximated. It is often required that the right hand side be computed fast and accurately and that the system be solved fast and accurately.

On the solution of the normal equations, one shall notice that the Gram matrix is Hermitian and positive definite. Thus, the Cholesky factorization of the matrix exists, $G = LL^{\mathrm{H}}$, where L is lower triangular matrix. The Cholesky factorization transforms the system into two triangular systems, which can be solved easily by substitutions.The arithmetic cost for the Cholesky factorization is $c \cdot n^3$, where c is a small constant. In the special case that the basis functions are orthogonal, the Gram matrix is diagonal, and the solution a is easily obtained using at most n arithmetic operations.

We now generalize the traditional approach in two ways, which can be combined. On the representing functions f_k we often have in practice

$$\mathrm{span}\{f_k | k = 1 : q, q \geq n\} = \mathcal{F}.$$

That is, the set $\{f_k\}$ is not necessarily linearly independent but contains a basis for $\mathcal{F}$.

1. The Gram matrix G is not necessarily nonsingular. Nevertheless
$$b \in \mathrm{span}(G),$$
i.e., the LS solution exists.

2. The LS solution is unique iff $\{f_k\}$ are linearly independent. In the case of multiple LS solutions, we often selects the solution with minimum 2-norm.

In the other way, we let $\{h_i, i = 1 : p, p \geq n\}$ be another set of functions in $\mathcal{F}$ containing a basis. Then, the LS solution satisfies the following equations as well

$$\langle h_i, g \rangle = \sum_{k=1}^{q} \alpha_k \langle h_i, f_k \rangle, \quad i = 1 : p.$$

- The matrix is not necessarily Hermitian or symmetric. It is not necessarily square. It is nonsingular when $p = q = n$, and it has a nonsingular submatrix of order n in general.
- We may find h_i so that the matrix $[\langle h_i, f_k \rangle]$ is triangular and easy to solve for a. Specifically, when the matrix is upper triangular, h_i is orthogonal to f_k for $k = 1 : i - 1$.
- We may also find h_i that are *bi-orthogonal* with f_k so that the matrix is diagonal.

Approximation of continuous functions

Consider the inner-product space $U = C[a, b]$ with

$$\langle f, g \rangle = \int_a^b f(t)g(t)w(t)dt,$$

where $w(t) \geq 0$ is a weight function. LS Approximations in the following subspaces are straightforward, assuming the right hand side for the normal equations can be computed.

- Let $[a, b] = [-1, 1]$ and $V = P_n[-1, 1]$.
  1. In the case $w(t) = 1$ the Legendre polynomials are often used as an orthogonal basis.
  2. In the case $w(t) = (1 - x^2)^{-1/2}$ the Chebyshev polynomials are often used as an orthogonal basis.
- Let $[a, b] = [-\pi, \pi]$ and $V = T_n[-\pi, \pi]$. Let $g \in U$. Then g has the Fourier expansion

  $$g(t) = \frac{\alpha_0}{2} + \sum_{k=1}^{\infty} \alpha_k \cos(kx) + \beta_k \sin(kx).$$

  It is easy to verify that the LS approximation in T_n is the truncated expansion at $k = n$.

Least-Squares Data Modeling

In interpolation we ask a fitting curve to pass through all the given points (x_i, y_i). Such a request is necessary for function evaluation at non-nodal points, in which case the interpolation nodes are well designed and the interpolation at non-nodal points often employs local references only. The interpolation condition, however, may not be necessary or appropriate for data modeling for a couple of reasons. First, for a large set of data, an interpolating function may involve too many basis functions to be computationally efficient in evaluations. It means in the case of polynomial interpolation that the degree of the interpolating polynomial may be very high. Second, interpolations may vary significantly with data distributions. Third, interpolations are sensitive to errors in the data. Consequently interpolation techniques may fail characterizing (modeling) the relationship between x and y. One simple approach to alleviating the problems is to relax the interpolation condition. Instead of asking the fitting curve to pass through the given data points, we seek a fitting curve that is the closest, in some sense, to the data among a family of candidate curves. Specifically, let $f(x)$ be a member in the family $\mathcal{F}$ of candidate fitting functions. We allow fitting errors at the data

$$e_i = f(x_i) - y_i, \quad i = 1 : N,$$

and find the best fitting function that minimize the errors according to the least-square metric

$$\min_{f \in \mathcal{F}} \|e\|_2,$$

where the 2-norm is induced from an inner product. This is a special case of LS approximation in discrete vector spaces. Recall that the fitting errors at the nodal points are required to be zero in

interpolation. The LS solution is not as sensitive to the errors in the data, and the weight associated with the inner product may make the LS approximation less sensitive to data distribution. By permitting and controlling the fitting errors, the number of basis functions may be significantly reduced, Such a process is known as least-squares *regression* in statistics where the data are considered from random variables.

We elaborate on the LS data fitting problem and solution. Given data (x_i, y_i), $x_i \in [a, b]$, $i = 1 : m$. Let $\mathcal{F}$ be a family of fitting functions and a subspace of $C[a, b]$. Then, the LS curve fitting problem becomes

$$\min_{f \in \mathcal{F}} \|y - f_d\|_2,$$

with

$$y = [y_1, \cdots, y_m]^{\mathrm{T}}, \quad f_d = [f(x_1), \cdots, f(x_m)]^{\mathrm{T}} \in R^m.$$

Notice that the fitting function $f(x)$ is continuous, the LS approximation is determined via comparing $f_d = [f(x_i)] \in R^m$ against $y = [y_i = y(x_i)] \in R^m$. Let $\{f_k, k = 1 : n\}$ be a basis of $\mathcal{F}$. Then the LS approximation is in the subspace spanned by the columns

$$B = [b_{ij}] = [f_j(x_i)] \in R^{m \times n},$$

and takes the more specific form

$$\min_a \|y - B \cdot a\|_2.$$

The extreme case with zero tolerance on errors recovers the interpolation problem

$$B \cdot a = y.$$

Although f_k are linearly independent, the columns of B are not necessarily linearly independent, depending on the distribution

of nodes x_i. With the same set of representing functions f_k, the matrix changes with nodes x_i. While the interpolation solution may not exist, the LS solution always exists. We proceed with the computation for the LS curve fitting.

1. For the case $\langle u, v\rangle = u^{\mathrm{T}} v$,

   - The normal equations become

     $$B^{\mathrm{T}} B\, a = B^{\mathrm{T}} y.$$

     When B is square and nonsingular, the normal equations are viewed a symmetrized, mathematically equivalent version of $Ba = y$. But the condition number for the numerlcal solution is squared.

   - Let Q be any matrix such that

     $$\mathrm{span}(Q) = \mathrm{span}(B).$$

     Then the LS solution satisfies the generalized equations

     $$Q^{\mathrm{T}} B\, a = Q^{\mathrm{T}} y.$$

     In particular, let $B = QRP$ be a QRP factorization of B, where Q has orthogonal columns, R is upper triangular, and P is a permutation matrix. With such a matrix Q, the above system is essentially triangular.

     When B is square and nonsingular, one uses the QRP factorization to transform the system $Ba = y$ into an equivalent triangular one, and the condition number remains the same, unlike converting to the normal equations.

2. In general, $\langle u, v \rangle = u^{\mathrm{T}} W v$, where W is a weight matrix (symmetric, positive-definite), the normal equation become

$$B^{\mathrm{T}} W B\, a = B^{\mathrm{T}} W y.$$

And the generalized equations are

$$Q^{\mathrm{T}} W B\, a = Q^{\mathrm{T}} W y,$$

where $\mathrm{span}(Q) = \mathrm{span}(B)$. There exists Q so that $Q^{\mathrm{T}} W B$ is triangular.