Personal notes on computational mathematics

Andriy Zhugayevych (http://zhugayevych.me)

January 16, 2023

§1. Interpolation and data fitting

1.1. Least squares approximation

Let $\{x_i\}$ be a mesh of points of a manifold M. Let $\{\phi_\alpha : M \to \mathbb{C}\}$ be a set of complex (or real) valued functions on M. Any function $f : M \to \mathbb{C}$ specified by its values f_i at the mesh points can be approximated by

$$f(x) = \sum_\alpha c_\alpha \phi_\alpha(x)$$

minimizing the sum of the absolute values of the residues, $\sum_i w_i |f(x_i) - f_i|^2$, where w_i are the weights. The vector of the optimal coefficients c_α satisfies the linear equation $Mc = V$, where

$$M_{\alpha\beta} = \sum_i w_i \overline{\phi}_\alpha(x_i)\phi_\beta(x_i), \qquad V_\alpha = \sum_i w_i \overline{\phi}_\alpha(x_i) f_i.$$

1.2. Least squares linear algebra

The least squares solution to the equation

$$\mathbf{AX} = \mathbf{B} \tag{1.1}$$

minimizes the Frobenius norm of $\mathbf{AX} - \mathbf{B}$. It is unique and is given by

$$\mathbf{X}_0 \equiv \arg\min_X ||\mathbf{AX} - \mathbf{B}||_\mathrm{F} = \left(\mathbf{A}^\top \mathbf{A}\right)^{-1} \mathbf{A}^\top \mathbf{B}.$$

If $\mathbf{X} = \boldsymbol{x}$ is a vector of size d and $\mathbf{B} = \boldsymbol{b}$ is a vector of size n denoting sampling of a random variable, then $\boldsymbol{x}_0 - \boldsymbol{x}$ is asymptotically normal with the covariance matrix $\mathbf{\Sigma} = \sigma^2 \left(\mathbf{A}^\top \mathbf{A}\right)^{-1}$, where

$$\sigma^2 = \frac{1}{n} ||\mathbf{A}\boldsymbol{X}_0 - \boldsymbol{b}||_\mathrm{F}^2$$

is the standard deviation for the residuals, implying that the estimate $\boldsymbol{b} = \boldsymbol{ax}$ asymptotically has normal distribution with the mean value $\boldsymbol{ax}_0$ and the dispersion $\boldsymbol{a}\mathbf{\Sigma}\boldsymbol{a}^\top$ (note that $\mathrm{tr}\, \mathbf{A\Sigma A}^\top = d$).

If $\mathbf{A}$ and $\mathbf{B}$ depend on some parameter then it is easy to show that

$$||\mathbf{AX}_0 - \mathbf{B}||_\mathrm{F}' = ||\mathbf{AX}_0 - \mathbf{B}||_\mathrm{F}^{-1} \,\mathrm{tr} \left[\left(\mathbf{AX}_0 - \mathbf{B}\right)^\top \left(\mathbf{A}'\mathbf{X}_0 - \mathbf{B}'\right) \right],$$

where prime means the derivative over this parameter.

In the so called orthogonal Procrustes problem the solution is constrained to orthogonal matrices only, i.e. $\mathbf{X}^\top \mathbf{X} = 1$ condition is imposed. In this case the solution is given by

$$\mathbf{X}_0 = \mathbf{UV}^\top, \text{ where } \mathbf{USV}^\top \equiv \mathbf{A}^\top \mathbf{B}$$

is the singular value decomposition. If in 3D we additionally fix rotation axis $\boldsymbol{n}$ then the angle

$$\alpha = \arctan \left[2 \sum_i \boldsymbol{n} \left(\boldsymbol{q}_i \times \boldsymbol{p}_i\right), \sum_i \left(|\boldsymbol{n} \times \boldsymbol{p}_i|^2 - |\boldsymbol{n} \times \boldsymbol{q}_i|^2 \right) \right], \text{ where } \boldsymbol{p} = \boldsymbol{r}' + \boldsymbol{r},\ \boldsymbol{q} = \boldsymbol{r}' - \boldsymbol{r},$$

gives the least squares rotation of set of points $\boldsymbol{r}$ towards the set $\boldsymbol{r}'$, provided that both sets are centered that is $\sum_i \boldsymbol{r}_i = 0$.

1.3. Extrapolation

Extrapolation of a finite sequence of vectors can improve the convergence if these vectors are generated by some slowly converging (or diverging) iteration procedure. Here we consider the minimal polynomial extrapolation (MPE) method [2, 1]. Let $x_1, \ldots, x_m$ be a sequence of vectors generated by the formula $x_{i+1} = Ax_i$. We are looking for the best estimate of the solution of the equation $\xi = A\xi$ in the form

$$x = \sum_{i=1}^{m} w_i x_i.$$

One of w_i can be chosen arbitrary, let it be w_m. The rest of w_i is chosen so as to minimize the norm $\langle x - Ax | x - Ax \rangle$, that results in the set of linear equations

$$\sum_{j=1}^{m} \langle \delta x_i | \delta x_j \rangle w_j = 0, \quad i = \overline{1, m-1}, \tag{1.2}$$

where $\delta x_j = x_{j+1} - x_j$. To fix w_m we imply the condition

$$\sum_{j=1}^{m} w_j = 1,$$

which asymptotically conserves the norm. Note that x_{m+1} is used in (1.2) through δx_m.

The minimum nontrivial extrapolation order is $m = 2$. In this case

$$w_2 = 1 - w_1 = \frac{\langle \delta x_1 | \delta x_1 \rangle}{\langle \delta x_1 | \delta x_1 - \delta x_2 \rangle},$$

reducing for a scalar sequence to Aitken's extrapolation formula.

1.4. Grid optimization (open problem)

Let $f(x, \lambda)$ be a function smooth in $x \in \mathbb{R}^d$ and "parametric" in λ. An example from computational chemistry: x are some molecular coordinates (e.g. dihedral, bond length, center of mass), f is potential energy, and λ is computational method. Different methods are compared according to the norm

$$\left(\int [f(x, \lambda_1) - f(x, \lambda_2)]^p \, \rho(x) \, \mathrm{d}x \right)^{1/p}.$$

The weight function ρ is often given by Boltzmann distribution $\mathrm{e}^{-\beta f(x, \lambda_0)}$, where λ_0 is a hypothetical "exact" method.

The problem is to find a reasonably small grid in x (let say, < 10 points) to benchmark different methods λ.

References

[1] Jbilou K, Sadok H, Vector extrapolation methods. Applications and numerical comparison, J Comp App Math 122, 149 (2000)

[2] S Cabay, L W Jackson, A polynomial extrapolation method for finding limits and antilimits of vector sequences, SIAM J Numer Anal 13, 734 (1976)