Iterative Unfolding Optimization with the Mean Squared Error Metric

Shih-Kai Lin

Colorado State University

Finding-numu Meeting
September 27, 2019

Recap

- The standard metric in the ND group used by all analyses requiring unfolding is the average global correlation coefficient[1],

$$\rho_{avg} = \frac{1}{M}\sum_{j=1}^{M}\sqrt{1 - \frac{1}{\boldsymbol{V}_{jj}(\boldsymbol{V}^{-1})_{jj}}} \tag{1}$$

, where M is the number of bins and $\boldsymbol{V}$ is the covariance matrix in true space inferred by the unfolding algorithm.

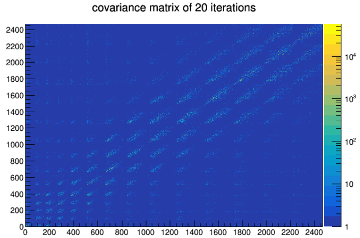
- For analyses with tens of bins, this is a convenient metric. However, for an analysis with thousands of bins, this metric turned out to be infeasible.

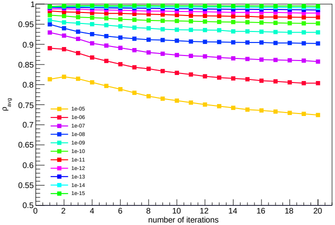
[1] Stefan Schmitt, "Data Unfolding Methods in High Energy Physics"

Infeasibility of Average Global Correlation Coefficient for Many-Bin Analyses

Inverting a covariance matrix this large turns out to be very tricky.

- The covariance matrices all have astronomical condition numbers (i.e., ill-conditioned or nearly singular).
- Numerical inversion is still possible but subject to an arbitrary, small cut-off number, or tolerance, brought into play by SVD.
- Forcefully getting the calculation through results in numerical instability, such as negative values in square root in Eq. 1. Removing unphysical values, results are shown to the right. No clear minimum is observed.





A Simple Alternative Metric: Mean Squared Error

Suppose θ is the a true parameter to be estimated, and $\hat{\theta}$ is an estimator of the parameter. The mean squared error (MSE) can be decomposed into a sum of variance and bias squared.

$$\begin{aligned} MSE &= E[(\hat{\theta}-\theta)^2] \\ &= E[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2] = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \\ &= (E[\hat{\theta}^2] - E[\hat{\theta}]^2) + (E[\hat{\theta}]^2 - 2\theta E[\hat{\theta}] + \theta^2) \\ &= V[\hat{\theta}] + b^2 \end{aligned} \tag{2}$$

, where $b = E[\hat{\theta}] - \theta$ is the bias of the estimator.
Equation 2 is called bias-variance decomposition.

This is a very common metric in statistics and machine learning as well.

Reference

Most of the contents in this document are taken from this textbook, especially Chapter 11 dedicated to unfolding.