PhD Seminar Talk 2

Stein Shrinkage for T-Product Tensor Covariance Estimation

Hemant Sharma[1]

[1]Department of Mathematics, Indian Institute of Information Technology, Design and Manufacturing Kancheepuram, Chennai, India – 600127

Abstract: Covariance estimation is a basic step in many problems where the data are third-order tensors, such as hyperspectral images, videos, fMRI arrays or multi-trial EEG recordings. The usual estimator is the sample covariance, which is also the maximum likelihood estimator. A classical result of Stein shows that this estimator is inadmissible: if we shrink it slightly towards a simple target, we can do better for every true covariance. Our aim is to understand whether this phenomenon carries over to tensors under the T-product of Kilmer and Martin, and the talk is divided into two parts.

$\star$ **Stein loss for tensors and shrinkage estimators**: We define a Stein loss L_S on the set of T-positive definite tensors and show two basic properties. It is invariant under T-congruence, and after a Fourier transform along the third mode it splits into a sum of ordinary matrix Stein losses, one for each frontal slice. This splitting is the main tool of the work, since it turns every tensor question into p independent Wishart problems of size n. One consequence is that the poor behaviour of the MLE for small samples depends on the slice size n and not on the full dimension $d = np$. Applying Haff's result to each slice, we prove that a slice-wise Haff estimator strictly dominates the sample T-covariance whenever $m \geq n+2$. We also derive a Ledoit–Wolf type estimator $\widehat{\mathcal{C}}(\rho^*) = (1-\rho^*)\mathcal{S} + \rho^*\mu\mathcal{I}$, where the optimal shrinkage intensity has a closed form in terms of $\mathrm{Ttr}(\mathcal{S})$ and $\mathrm{Ttr}(\mathcal{S}^2)$ only. These are the same quantities that appear in our earlier TDep eigenvalue bounds. An oracle inequality shows that the data-driven intensity is within $O(1/m)$ of the best possible choice.

$\star\star$ **Geometric risks and numerical results**: We then look at two other ways of measuring error on the positive definite cone, the log-spectral risk and the Bures–Wasserstein risk, and bound both in terms of the Stein risk using two elementary inequalities. For the Bures–Wasserstein case, the sharpest comparison is proved when the tensors commute and is left as a conjecture in general. On synthetic data, the results agree with the theory, and using the shrinkage estimator for tensor whitening reduces the decorrelation error by more than half when samples are few. On real image patches, shrinkage gives a much better held-out likelihood, as expected, but the plain MLE is slightly better for Wiener-filter denoising. We explain this through the spectrum: lifting the small eigenvalues helps the likelihood, but it also lets more noise pass through the Wiener filter. So whether shrinkage helps depends not only on the estimator but also on the task and the eigenvalue regime.

Keywords: T-product, Covariance estimation, Stein loss, Shrinkage estimator, Bures–Wasserstein risk, Oracle inequality, Tensor whitening.