Big Data analysis by polyspline non-parametric regressions

Ognyan Kounchev

Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

kounchev@math.bas.bg

**Абстракт:**

The analysis of large amounts of data (Big Data) either of one dimensional or of multidimensional origin, requires models which admit fast algorithms for their calculation.

One of the most amazing features of the interpolation and smoothing splines of odd degree is that their computation reduces to computations with tridiagonal matrices, which implies linear order of operations, [5]. In the multivariate case, the Thin Plate Splines (TPS) are usually considered to be a generalization of the cubic splines which is appropriate for the analysis of large volumes of data. TPS are a special case of a more general family of functions, the so-called Radial Basis Functions (RBFs). The TPS are used to solve interpolation and smoothing problems for two-dimensional data, cf. [2], [3], [4], [5]. However, there are no such algorithms which are fast, and having the same rate of convergence which would correspond to the one-dimensional rate.

We report about a newly developed multivariate parametric regression based on the theory of polysplines [1], which represent an alternative concept to the Radial Basis Functions. We study fast algorithms for interpolation and smoothing by means of one-dimensional L-splines, to which the computation of the polysplines is reduced in special domains, [1]. For these L-splines, we provide a generalization of the fast Reinsch algorithm which is famous from the computation of smoothing cubic splines, cf. [5].

This is a joint research with Tsvetomir Tsachev, sponsored by projects DH02-13 and I 02/19 with Bulgarian NSF.

References

[1] O. Kounchev, Multivariate Polysplines. Applications to Numerical and Wavelet Analysis, Academic Press (Elsevier), 2001.

[2] Wahba, G., Spline models for observational data, SIAM, Philadelphia, Pennsylvania, 1990.

[3] Hastie, T., Tibshirani, R., Friedman, J., The elements of statistical learning: Data Mining, Inference, and Prediction, 2009.

[4] Green, P., B. Silverman, Nonparametric regression and generalized linear models, Chapman and Hall, 1994.

[5] Stoer, J., Bulirsch, R., Introduction to Numerical Analysis, Springer, 2003.