Parallel QR factorization of block-tridiagonal matrices from Generative Models

Alfredo Buttari*[1], Søren Hauberg , and Costy Kodsi

[1]Institut de recherche en informatique de Toulouse (IRIT) – CNRS : UMR5505 – 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France

Abstract

This talk presents a parallel method for the QR factorization of block-tridiagonal matrices arising from Generative Models, which are unsupervised learning techniques, where the representation space is endowed with a Riemannian metric in order to preserve the inherent structure of data [1]. In such a setting, geodesics can be employed as measures of similarity and computed through the solution of a non-linear equation where the Jacobian has a block-tridiagonal structure.
The QR factorization of such matrices can only achieve limited parallelism because the blocks along the diagonal have to be reduced sequentially. We will show how a block Nested Dissection permutation can be used to improve parallelism at the price of a higher fill-in and flop count. Achieving a favorable compromise between parallelism and complexity not only depends on an appropriate choice for such permutation but also demands for a careful scheduling of operations which effectively exploits the available concurrency while making good use of data locality and reducing the transient (temporary) fill-in occurring in the course of the factorization.

We will present preliminary results obtained with a shared-memory parallel code relying on OpenMP task parallelism.
Søren Hauberg. "Only Bayes should learn a manifold". 2018.

*Speaker