FMR: Fast randomized algorithms for covariance matrix computations
Pierre Blanchard, Olivier Coulaud, Eric Darve, Alain Franc

To cite this version:
Pierre Blanchard, Olivier Coulaud, Eric Darve, Alain Franc. FMR: Fast randomized algorithms for covariance matrix computations. Platform for Advanced Scientific Computing (PASC), Jun 2016, Lausanne, Switzerland. 2016. hal-01334747

HAL Id: hal-01334747
https://hal.archives-ouvertes.fr/hal-01334747
Submitted on 23 Jun 2016
Fast randomized algorithms for covariance matrix computations

Pierre Blanchard 1 · Olivier Coulaud 1 · Eric Darve 2 · Alain Franc 1

ABSTRACT

We present an open-source library implementing fast algorithms for covariance matrices computations, e.g., randomized low-rank approximations (LRA) and fast multipole matrix multiplication (FMMP). The library can be used to approximate square roots of low-rank covariance matrices in SVD form using randomized LRA, instead of the standard O(N^3) cost.

Low-rank covariance matrices given as kernels, e.g., Gaussian decay, evaluated on 3D grids can be decomposed in O(N) operations using the FMMP. The performance of the library is illustrated on two examples:

- Generation of Gaussian Random Fields (GRF) on large spatial grids
- Multidimensional Scaling (MDS) for the classification of species.

RANDOM PROJECTION BASED LRA

Randomized SVD is a random projection-based LRA algorithms made popular by Halko et al. [4], which returns an approximate SVD of a symmetric matrix C ∈ ℝ^{N×N} given a prescribed numerical rank r in O(N^2 × r) operations:

- Form an approximate basis Q ∈ ℝ^{N×r} for the range of C.
- Form a sketched version of C using Gaussian random projection, i.e., application of C to a N-by-Gaussian random matrix Ω:

Y = ΩC

- Then, orthogonalize Y by means of QR Decomposition

QR = Y

- Thus, we get a low-rank representation of C in the form

C ≈ C_Ω ≈ QQ^T

with Frobenius/spectral error bounds that hold with high probability.

- Factorize C_Ω in SVD form: C_Ω = UΣV^T

Thus, the method offers many advantages:

- Easily implemented and parallelized,
- Easily extended to Cholesky, Interpolative Decomposition...
- Cost dominated by matrix multiplication, i.e., O(N^2).

However, C_Ω should fulfill the following conditions:

- be low-rank (r << N),
- have a fast decreasing spectrum (σ(C)/σ(C_Ω) << N).

EFFICIENT GENERATION OF GRF

This project aims at promoting new highly efficient FMMP algorithms to perform resource demanding computations in geostatistics.

Correlation kernels A Gaussian Random Field Y ∼ N(0, C) is a multi-variate Gaussian random variable with mean 0 and covariance C ∈ ℝ^{N×N}. The covariance can be prescribed as a kernel matrix, i.e.,

C = ((h(x_i, x_j))_{i,j=1,...,N})

where r_{ij} = ||x_i - x_j||_k denotes the distances between points of an arbitrary grid and k is a correlation kernel such as

h_{ex}(r) = e^{-r^2} (Exponential decay)

h_{gauss}(r) = e^{-r^2/2} (Gaussian decay)

The length scale r characterizes the decreasing speed of the correlation.

Square-root algorithms Covariance matrices are spars by definition of correlation kernels. Hence, C admits the following representation:

C = AA^T

where the matrix factor A ∈ ℝ^N×N is often called a square root of C. Methods for generating Gaussian Random Fields usually differ by the way A is prescribed:

- standard matrix decompositions (O(N^3))
- fast algorithms (e.g., FastLA) for equispaced grids
- the turning bands method (approximate)

Most of them become computationally prohibitive for large N, i.e., N over a few thousands.

METRIC MDS

aims at reconstructing a cloud of points X in a low-dimensional feature space, e.g., X ∈ ℝ^N×D, from a given distance/dissimilarity matrix D ∈ ℝ^N×N (Smith-Watson scores of local alignment). The algorithm [2] consists in:

Assembling a similarity matrix as

C_{ij} = \frac{1}{2} D_{ij}^2 + \sum_{k \neq i,j} \frac{1}{2} D_{ik} D_{kj}

Computing the SVD of C, i.e., C = UΣU^T

Forming X = C^{1/2} / UΣ^{1/2} (LS minimizer)

REFERENCES


FUNDINGS

This work was partially supported by the associate team FastLA (Inria, Stanford University & Lawrence Berkeley National Laboratory).