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

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Fast randomized algorithms for covariance matrix computations

Pierre Blanchard 1 - Olivier Coulaud 2 - 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 (FMM). The library can be used to approximate square roots of low-rank covariance matrices in $O(N^2)$ operations 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 $O(N \times r)$ operations using the FMM. 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 $C \in \mathbb{R}^{n \times n}$ given a prescribed numerical rank $r$ in $O(n^2 \times r)$ operations.

- Form an approximate basis $Q \in \mathbb{R}^{N \times r}$ for the range of $C$.
- Form a sketched version of $C$ using Gaussian random projection, i.e., application of $C$ to a $N \times r$ Gaussian random matrix $\Omega$.
- Then, orthogonalize $Y$ by means of a QR Decomposition
  $QR = Y$.
- Thus, we get a low-rank representation of $C$ in the form $C \approx QRQ^T$ with Frobenius/spectral error bounds that hold with high probability.

- Factorize $C_{SV}$ in SVD form: $C_{SV} = U \Sigma V^T$.
  - We start by assembling the small $r \times r$ matrix $B = Q^T C Q$.
  - Then, perform a small SVD: $B = U \Sigma V^T$.
  - Form $U = U_1 U_2$ and $\Sigma = \Sigma_1$.
- If $C$ is positive semi-definite, then $C \approx AA^T$.

The method offers many advantages:

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

However, $C$ should fulfill the following conditions:

- be low-rank ($r < N$).
- have a fast decreasing spectrum $\lambda_k(i) \in [\lambda_1^{(i)}, \lambda_r^{(i)}]$ (small $\lambda_k^{(i)}$).

EFFICIENT GENERATION OF GRF

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

Correlation kernels: A Gaussian Random Field $Y \sim \mathcal{N}(0,C)$ is a multi-variate Gaussian random variable with mean 0 and covariance $C \in \mathbb{R}^{N \times N}$. The covariance can be prescribed as a kernel matrix, i.e.,

$$C = \text{exp}(\|\mathbf{x}_i - \mathbf{x}_j\|) = \text{exp}(r \|\mathbf{x}_i - \mathbf{x}_j\|).$$

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:

$$C = A A^T$$

where the matrix factor $A \in \mathbb{R}^{N \times r}$ is often called a square root of $C$. Methods for generating Gaussian Random Fields usually differ by the way $A$ is precomputed:

- standard matrix decompositions ($O(N^3)$);
- circulant embedding ($O(N \log N)$) for equispaced grids;
- the turning bands method (approximate).

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

REFERENCES


FUNDING

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

Sources are available online as part of the open-source package FMR. They can be downloaded for free at the following address:
https://gforge.inria.fr/projects/fmr

BIOGRAPHY

Pierre Blanchard: BSc, Inria Bordeaux, 200, rue Vieille Tour 33405 Talence, France name:isambruin.fr

Eric Darve: Stanford University, CA, USA darve@stanford.edu

Olivier Coulaud: BioGeCo, Inra & PLEIADE, Inria Bordeaux, France alain.franc@inria.fr

Alain Franc: Inria Bordeaux, France alain.franc@inria.fr

Fast randomized algorithms for covariance matrix computations

Pierre Blanchard 1 - Olivier Coulaud 2 - 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 (FMM). The library can be used to approximate square roots of low-rank covariance matrices in $O(N^2 \times r)$ operations 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 $O(N \times r)$ operations using the FMM. 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 $C \in \mathbb{R}^{n \times n}$ given a prescribed numerical rank $r$ in $O(n^2 \times r)$ operations.

- Form an approximate basis $Q \in \mathbb{R}^{N \times r}$ for the range of $C$.
- Form a sketched version of $C$ using Gaussian random projection, i.e., application of $C$ to a $N \times r$ Gaussian random matrix $\Omega$.
- Then, orthogonalize $Y$ by means of a QR Decomposition
  $QR = Y$.
- Thus, we get a low-rank representation of $C$ in the form $C \approx QRQ^T$ with Frobenius/spectral error bounds that hold with high probability.

- Factorize $C_{SV}$ in SVD form: $C_{SV} = U \Sigma V^T$.
  - We start by assembling the small $r \times r$ matrix $B = Q^T C Q$.
  - Then, perform a small SVD: $B = U \Sigma V^T$.
  - Form $U = U_1 U_2$ and $\Sigma = \Sigma_1$.
- If $C$ is positive semi-definite, then $C \approx AA^T$.

The method offers many advantages:

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

However, $C$ should fulfill the following conditions:

- be low-rank ($r < N$).
- have a fast decreasing spectrum $\lambda_k(i) \in [\lambda_1^{(i)}, \lambda_r^{(i)}]$ (small $\lambda_k^{(i)}$).

EFFICIENT GENERATION OF GRF

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

Correlation kernels: A Gaussian Random Field $Y \sim \mathcal{N}(0,C)$ is a multi-variate Gaussian random variable with mean 0 and covariance $C \in \mathbb{R}^{N \times N}$. The covariance can be prescribed as a kernel matrix, i.e.,

$$C = \text{exp}(\|\mathbf{x}_i - \mathbf{x}_j\|) = \text{exp}(r \|\mathbf{x}_i - \mathbf{x}_j\|).$$

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:

$$C = A A^T$$

where the matrix factor $A \in \mathbb{R}^{N \times r}$ is often called a square root of $C$. Methods for generating Gaussian Random Fields usually differ by the way $A$ is precomputed:

- standard matrix decompositions ($O(N^3)$);
- circulant embedding ($O(N \log N)$) for equispaced grids;
- the turning bands method (approximate).

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

REFERENCES


FUNDING

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