Iterative numerical methods for sampling from high dimensional Gaussian distribution

Présented by :

ONANA ASSOUGA Florent

Under the supervision of:

Issa KARAMBAL & Franca HOFFMANN

UNIVERSITY OF YAOUNDÉ I

DÉPARTEMENT OF MATHÉMATICS

26 juillet 2023

Introduction and Overview

In the Bayesian inference (Machine learning) setting, quantifying uncertainties implies the computation of the log determinant of the precision matrix. The method of choice depends on the problem as well as the structure of the covariance. In low dimension, the method of choice is based on Cholesky factorization. However in high dimensions, computing the Cholesky factorization may be prohibitive due to memory limitation. Developing efficient methods that are comparable in term of efficiency to Cholesky factorization in low dimension for sampling Gaussian distributions and ease the computation of the log determinant of the precision matrix is currently a hot research topic. We aim to present three main iterative methods for sampling high dimensional Gaussian processes. All of them are based on Krylov method (Saad 2003) and rational approximations of specific matrix function.

Objectives

- Present the direct sampling method of Gaussian processes and the the algorithm of Krylov
- Present three iterative methods for sampling Gaussian processes : Lanczos method (see Saad 2003), the quadrature representation formula (see Hale et al 2008) and the continuous deformation method based on system of ODEs (see Allen et al 2000)
- Give some orientatins of our future tasks (Leja points techniques (matrix multiplication and estimation of eigenvalues bounds of the precision matrix))

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

Cholesky factorization

We set by $\sum$ the covariance matrix of aour Gaussian process and denote by $\mathcal{Q}$ its precision matrix (means $\mathcal{Q} = \sum^{-1}$). Since $\mathcal{Q}$ is a symmetric positive matrix, we can write $\mathcal{Q}$ as $\mathcal{Q} = LL^T$, where L is lower triangular Cholesky factor.
We have the following lemma.

Lemme 1

Suppose that the precision matrix has the following factorization, $\mathcal{Q} = LL^T$. If we want a sample $x \sim \mathcal{N}(0, \mathcal{Q}^{-1})$, it is enough to compute

$$x = L^{-T} z \quad \text{for} \quad z \sim \mathcal{N}(0, I). \tag{1}$$

Démonstration.

We have

$$Cov(L^{-T} z) = L^{-T} \underbrace{Cov(z)}_{=I} L^{-1} = (LL^T)^{-1}.$$

Cholesky factorization

On the other hand, we can factorize $\mathcal{Q}$ as $\mathcal{Q} = VDV^T$, where V is the orthogonal eigenvector matrix and D has the eigenvalues of $\mathcal{Q}$. Consequently, we have the following lemma :

Lemme 2

Suppose that the precision matrix has the following factorization, $\mathcal{Q} = VDV^T$. If we want a sample $x \sim \mathcal{N}(0, \mathcal{Q}^{-1})$, it is enough to compute

$$x = Q^{-1/2} z \quad \text{for} \quad z \sim \mathcal{N}(0, I). \tag{2}$$

Démonstration.

We have

$$\begin{aligned} Cov(Q^{-1/2} z) &= Cov(VD^{-1/2} V^T z) = VD^{-1/2} \underbrace{V^T V}_{=I} D^{-1/2} V^T \\ &= VD^{-1} V^T = \mathcal{Q}^{-1}. \end{aligned}$$

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

Krylov algorithm

Here we want to solve iteratively the equation

$$\mathcal{Q}x = z, \tag{3}$$

where $\mathcal{Q}$ and z are the input and the ouput is $x \approx \mathcal{Q}^{-1}z$.
The Krylov method is the conjuguate gradient method for solving (3).For iteration m, we have the folowing definition

Définition 3

The subspace

$$\mathcal{K}_m(\mathcal{Q}, r_0) = span\left\{r_0, \mathcal{Q}r_0, \mathcal{Q}^2 r_0, \cdots, \mathcal{Q}^m r_0\right\}, \tag{4}$$

where $r_0 = z - \mathcal{Q}x_0$ $(x_0 = 0)$,is a Krylov space.

Let define the functional

$$f(x) = \frac{1}{2}x^T \mathcal{Q}x - x^T z.$$

Krylov Algorithm

Since the matrix precision is symmetric positve define, the minimum of the functional f exists and is unique on the Krylov space $\mathcal{K}_m(\mathcal{Q}, r_0)$. The CG method finds the best possible solution to $\mathcal{Q}x = z$ in the subspace $\mathcal{K}_m(\mathcal{Q}, r_0)$.

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

The Lanczos method

It is a building block for self-adjoint Krylov methods and is perhaps the easiest ways to forming an orthonormal basis for $\mathcal{K}_m(\mathcal{Q}, r_0)$. The Lanczos method buids an orthonormal basis for the Krylovspace, namely $V = (v_1, v_2, \cdots, v_m)$ as well as the tridiagonal Hessenberg matrix $T_m = V^T \mathcal{Q} V$. Lastly, we project out the closest solution to $\mathcal{Q}^{-1} z$ in $\mathcal{K}_m(\mathcal{Q}, r_0)$ by

$$x = x_0 + \beta_0 V T_m^{-1/2} e_1. \tag{6}$$

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

Main idea

We need a rational approximation $f_N(\cdot)$ of the matrix inverse square root $\mathcal{Q}^{-1/2}$, means that, we design a rational function $f_N(\cdot)$ such that $f_N(\mathcal{Q})z \approx \mathcal{Q}^{-1/2}z$ on the spectrum of $\mathcal{Q}$. We set

$$x = \mathcal{Q}^{-1/2}z \approx f_N(\mathcal{Q})z = \sum_{j=1}^{N} \alpha_j (\mathcal{Q} - \sigma_j I)^{-1} z. \tag{7}$$

One achieve attractive solution through numerical quadrature on a contour integral. Generally, for analytic function in a domain containing the spectrum of $\mathcal{Q}$, it is possible to compute $f(\mathcal{Q})z$ by Cauchy's integral formula

$$f(\mathcal{Q})z = \frac{1}{2\pi i} \oint_{\Gamma} f(\zeta)(\zeta I - \mathcal{Q})^{-1} z d\zeta, \tag{8}$$

where Γ is a curve which encloses the spectrum of $\mathcal{Q}$.

Cauchy's integral formula

In our case, we have

$$\mathcal{Q}^{-1/2}z = \frac{1}{2\pi i}\zeta^{-1/2}(\zeta I - \mathcal{Q})^{-1}z d\zeta. \tag{9}$$

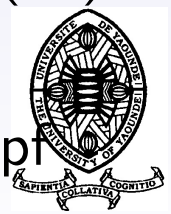
We have to make two important choices when approximaing this integral :

- Which curve Γ to use,
- what type of quadrature to employ.

If wee use the function $f(\mathcal{Q}) = \mathcal{Q}^{1/2}$, we have

$$\begin{aligned} \mathcal{Q}^{-1/2}z &= \mathcal{Q}^{-1}f(\mathcal{Q})z = \frac{1}{2\pi i}\oint_{\Gamma} f(\zeta)\zeta^{-1}(\zeta I - \mathcal{Q})^{-1}z d\zeta \\ &= \frac{1}{\pi i}\oint_{\Gamma_\omega}(\omega^2 - \mathcal{Q})^{-1}z d\omega, \end{aligned} \tag{10}$$

where we set $\omega^2 = \zeta$ and Γ_ω is a resulting curve from the change of variable $\zeta \mapsto \omega$.



There are several works on the choices of the curve to use and the type of quadrature. Hale et al 2008 suggested a conformal or angle preserving maping of the Jacobi elliptic function $\omega = \lambda_{min}^{1/2} sn(\frac{t}{k^2})$. Then we can have (see Zolotarev 1877 and Akhiezer 1990)/

$$\begin{aligned} \mathcal{Q}^{-1/2} z &= -\frac{1}{\pi i} \oint_{\mathbb{R}_-} (\mathcal{Q} - \omega^2)^{-1} z d\omega \\ &\approx \sum_{j=1}^{N} \alpha_j (\mathcal{Q} - \sigma_j I)^{-1} z. \end{aligned} \tag{11}$$

plan de l'exposé

1. Cholesky factorization and algorithm of Krylov
   - Direct sampling method of Gaussian processes
   - Algorithm of Krylov
2. three iterative methods for sampling Gaussian processes
   - The Lanczos method
   - Optimal rational approxilations with linear solves
   - Continuous deformation method
3. Remarks and Open problems

Continuous deformation

This method is based on solving the following ODE (see Allen et al 2000)

$$\frac{dz}{dt} = r(\mathcal{Q} - I)\left[t(\mathcal{Q} - I) + I\right]^{-1} z, \quad t \in [0, 1] \tag{12}$$

where $z(0) \sim \mathcal{N}(0, I)$ abd $r = 1/2$. The solution at the end point, $z(1) = x$ is a sample from $\mathcal{N}(0, \mathcal{Q}^{-1})$.

plan de l'exposé

A word on Leja points, Newton polynomial interpolation and interpolation at Leja points

A word on truncated Taylor series expansions

Bibliographie

- Erlend Aune, Jo Eidsvik, Yvo Pokern, Iterative numerical methods for sampling from high dimensional Gaussian distributions, Stat Comput 2012.
- Golub, G.H, Van Loan, Matrix computations. The Johns Hopkins University Press, 3rd edition october 1996.
- Rasmussen, C.E, Williams, Gaussian Processes for Machine learning, MIT 2006.
- Wainwright, Joran, Logdeterminant relaxation for approximate inference in descrete markov random fields.
- Fitzsimons, Cutazor, Osborne, Roberts, Filippone, Bayesian inference of log determinants 2017.

THANK YOU

FOR YOUR KIND

ATTENTION !