Rutgers University
Summer Research Program

Preparation for GpyTorch Implementation

Han Liu

August 7, 2019

1 Introduction

In the process of training and predicting a gaussian process (GP), we need to get three quantities: the mean of predicted input, the log marginal likelihood function and its derivative. Their expressions are given as following:

$$u_{f|D}(\hat{x}) = u(\hat{x}) + k_{X\hat{x}}^T K^{-1} y \tag{1}$$

$$logp(y|K) = -\frac{1}{2}log((2\pi)^k|K|) - \frac{1}{2}y^T K^{-1} y \tag{2}$$

$$\frac{\partial}{\partial\theta_j} logp(y|X,\theta) = \frac{1}{2}y^T K^{-1}\frac{\partial K}{\partial\theta_j}K^{-1}y - \frac{1}{2}tr(K^{-1}\frac{\partial K}{\partial\theta_j}) \tag{3}$$

These equation have three operations in common that dominate its time complexity: $K^{-1}y$, $log|K|$, $tr(K^{-1}\frac{\partial K}{\partial\theta_j})$. Before that, I use Cholesky decomposition of K to compute all three quantities (see *GP_code.py*), but it is very computational expensive, thus we need some parallel computional algorithm.
I think the innovative part of this paper is to use matrix-matrix multipy to combine three algorithms (Conjugate Gradient, Lancos Algorithm, Pivoted Cholesky Decomposition) together. The most important formula is as following:

$$[u_0 \quad u_1 \quad ... \quad u_t] = K^{-1}[y \quad z_1 \quad ... \quad z_t] \tag{4}$$

The whole algorithm flow chart can be seen in Figure 1. In the following section, I will implement all the three algorithms individually by MATLAB.

2 The Pivoting Cholesky Decomposition

In the process of implementing the pivoting cholesky decomposition, I found the algorithm I derived last week had some problems, so I have implemented it again. Specifically, we need to find $P^TKP = R^TR$ for symmetric positive matrix $K \in R^{n\times n}$, where P is permutation matrix. Following the derivation of the pivoting cholesky decomposition in the paper, we need to use one propertiy of permutation matrix:

$$P^{-1} = P^T \tag{5}$$

The mathematical details are given in appendix. Specifically, the whole algorithm should be as following:

Figure 1: Pseudo-code for Pivoting Cholesky Decomposition