Randomized Iterative Algorithms for Solving Elliptic PDEs

Jialun Zhang[1]

[1] ICME, Stanford University

Abstract

We consider the classical problem of solving an elliptic partial differential equation on a circular disk in $\mathbb{R}^2$. These equations can be solved by various deterministic algorithms such as finite differences and finite elements. The finite element method in particular projects the equation onto a finite-dimensional subspace by discretizing the domain and considering piece-wise linear functions. This usually gives us a linear system of the form $Au = b$, where A is a large sparse matrix that is positive semi-definite. For these kind of linear systems, iterative algorithms such as the conjugate gradient method has been widely used. In this poster, we chiefly consider two *randomized* iterative methods, namely the **randomize coordinate descent** method and **Gaussian sampling**.

Iterative Randomized Least Squares

Many randomized iterative algorithms can be described as a specific case of the following algorithm: for some positive definite $B \in \mathbb{R}^{n\times n}$ and random matrix $S \in \mathbb{R}^{m\times q}$, choose an initial $x^{(0)}$ and apply updates of the form

$$x^{k+1} = \arg\min_{x\in\mathbb{R}^n} \|x - x^k\|_B^2 \quad \text{subject to} \quad S^T Ax = S^T b,$$

where $\|x\|_B = \sqrt{x^T Bx}$. Analytically, this can be written as

$$x^{k+1} = x^k - B^{-1}A^T S \left(S^T AB^{-1}A^T S\right)^\dagger S^T \left(Ax^k - b\right).$$

Randomized Coordinate Descent

Assume that A is positive definite, so we can choose $B = A$ and $S = e^i$. This yields the following update

$$x^{k+1} = x^k - \frac{(A_{i:})^T x^k - b_i}{A_{ii}} e^i. \tag{1}$$

The convergence can be written as

$$\mathbf{E}\left[\|x^k - x^*\|_A^2\right] \le \left(1 - \frac{\lambda_{\min}(A)}{\mathrm{Tr}(A)}\right)^k \|x^0 - x^*\|_A^2. \tag{2}$$

The algorithm has an iteration complexity of

$$O\left(\mathrm{Tr}(A)/\lambda_{\min}(A)\right),$$

while CG has an iteration complexity of

$$O(\sqrt{\lambda_{\max}(A)/\lambda_{\min}(A)}).$$

The advantage is that each iteration costs $O(n)$ instead of $O(n^2)$.

Gaussian Sampling

Assume that A is positive definite, so we can choose $B = A$. Let $S = \zeta \sim N(0, \Sigma)$. This yields the update

$$x^{k+1} = x^k - \frac{\eta^T\left(Ax^k - b\right)}{\|\eta\|_A^2}\eta. \tag{3}$$

The convergence can be written as

$$\mathbf{E}\left[\|x^k - x^*\|_A^2\right] \le \rho^k \|x^0 - x^*\|_A^2, \tag{4}$$

where

$$1 - \frac{1}{n} \le \rho \le 1 - \frac{2}{\pi}\frac{\lambda_{\min}(\Omega)}{\mathrm{Tr}(\Omega)}, \tag{5}$$

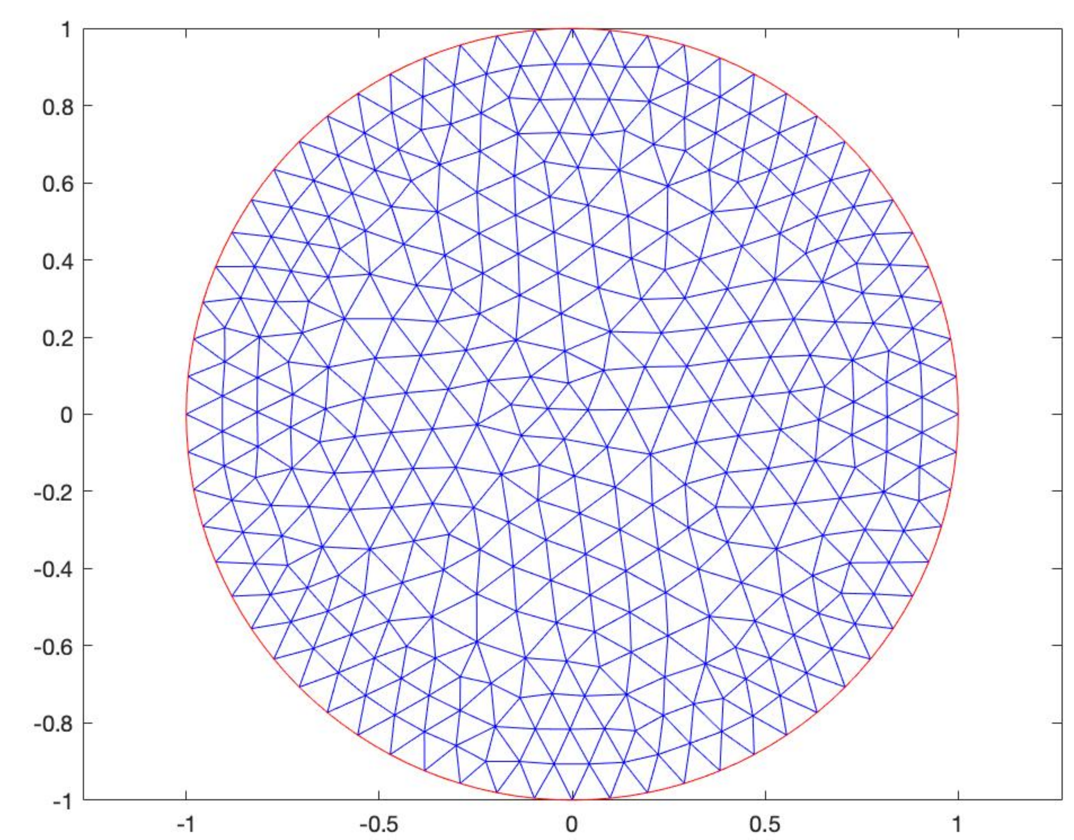
where $\Omega \overset{\text{def}}{=} B^{-1/2}A^\top \Sigma A B^{-1/2}$.

Elliptic PDEs and Finite Elements

We consider a partial differential equation of the form

$$\mathcal{A}u := -\nabla\cdot(a\nabla u) = f \quad \text{in } \Omega, \quad \text{with } u = 0 \quad \text{on } \Gamma.$$

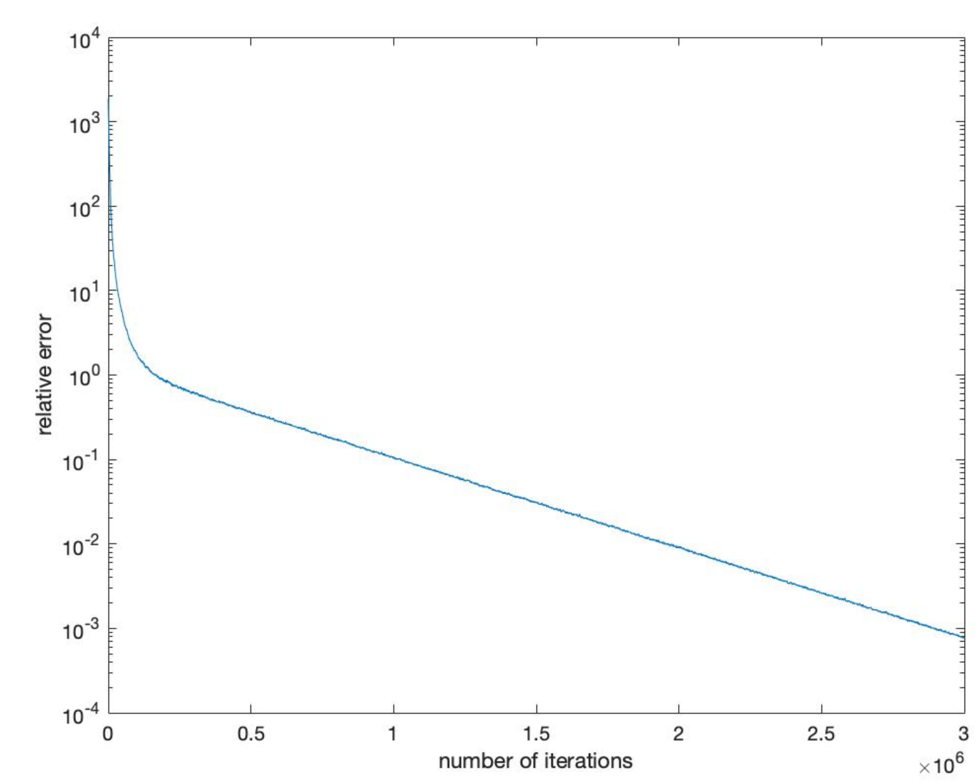
We discretize domain as shown in Figure 1.



By considering piecewise linear functions defined on the triangles, we attain an equation of the form $Au = b$, where A is sparse and SPD.

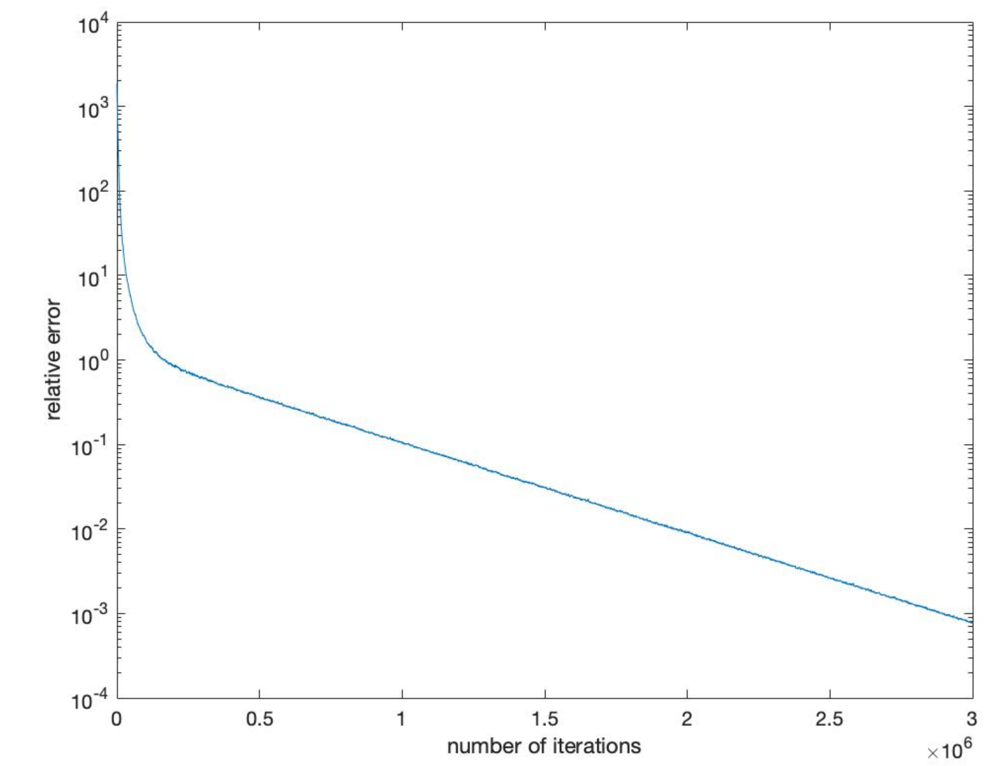
Convergence

Convergence for Randomized Coordinate Descent.



Convergence (continued)

Convergence for Gaussian Sampling.



Discussion

Due to limitations in computation, we chose a mesh size of $h = 0.1$ which results in a SPD matrix of size $O(10^3)$. On these scales, the randomized methods perform much slower compared to CG. The number of iterations it takes to converge is almost squared. However, it is clear that these methods do eventually converge to the optimal solution, and can be useful when A is sufficiently large.

Reference