Conjugate gradient method

Angelina Yaroshenko

Optimization Class Project. MIPT

Introduction

Conjugate gradient method is used for solving SPD system $\mathbf{Ax} = \mathbf{b}$ or

$$x^T Ax - x^T b \longrightarrow min$$

which is the same problem . It converges theoretically in n iterations, where n is the dimension of the problem.

Algorithm 1: $\mathrm{CG}(A, b, x_0, max_iter, tol)$

Input: positive definite symmetric matrix $A \in R^{n \times n}$; vector $b \in R^n$; x_0 initial point; max iter the number of iterations; tol - tolerance.

Output: x - solution of a system

$r_0 = b - Ax_0$

$p_0 = r_0$;

$k = 0$;

while $\frac{||r_k||}{||b||} > tol$ *and* $k < maxiter$ **do**

  $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$;

  $x_{k+1} = x_k + \alpha_k p_k$;

  $r_{k+1} = r_k - \alpha_k A p_k$;

  $\beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$;

  $p_{k+1} = r_{k+1} + \beta_k p_k$; $k = k + 1$;

return x_{k+1}

It may also be used for solving non-linear systems of equations but it requires much more properties of the input problem.

I have prepared a set of tasks on the topic of conjugate gradients. They cover such topics as conjugate gradients with preconditioning and the use of conjugate gradients for finding approximation of inverse hessian in newton method.

Preconditioned conjugate gradient

CG method can converge faster if matrix has blocks of similar eigenvalues and condition number is low. These propertis can be obtained by using preconditioner: easy-to-invert matrix, which is close to A^{-1}. The task is to apply and compare 3 different preconditioners:

- Jacobi preconditioner: $M = diag(A_{11}, A_{22} ... A_{nn})$
- Incomplete Cholesky factorization, so-called IC(0)
- Relaxation preconditioner: M $= \frac{1}{2-\omega}((\frac{1}{\omega}D + L)(\frac{1}{\omega}D)^{-1}(\frac{1}{\omega}D + U))$
  Particularly, if $\omega = 1.0$ the method is called Gauss-Seidel symmetric preconditiner.

So the problem takes the following form:

$$M^{-1}Ax = M^{-1}b$$

Algorithm 2: $\mathrm{PCG}(A, b, x_0, max_iter, tol, M)$

Input: positive definite symmetric matrix $A \in R^{n \times n}$; vector $b \in R^n$; x_0 initial point; max iter the number of iterations; tol - tolerance; M is preconditiner.

Output: x - solution of a system

$p_0 = r_0 = b - Ax_0$, $M_{inv} = M^{-1}$

$z_0 = M_{inv} r_0$, $k = 0$;

while $\frac{||r_k||}{||b||} > tol$ *and* $k < maxiter$ **do**

  $\alpha_k = \frac{r_k^T z_k}{p_k^T A p_k}$;

  $x_{k+1} = x_k + \alpha_k p_k$;

  $z_{k+1} = M_{inv} r_{k+1}$;

  $r_{k+1} = r_k - \alpha_k A p_k$;

  $\beta_k = \frac{z_{k+1}^T r_{k+1}}{z_k^T r_k}$;

  $p_{k+1} = z_{k+1} + \beta_k p_k$; $k = k + 1$;

return x_{k+1}

A - matrix is bcsstk27.mtx taken from Suite Sparse Matrix Collection, b is vector of ones.

Using CG to approximate solution of non-linear system

Newton method requires 1 Hessian inversion on eash iteration. What if we use CG to approximate this matrix? The task is to compare this modified newton method with simple newton method and nonlinear conjugate gradients:

$$\Phi(x) = \cosh(\mathbf{x}^T \mathbf{A} \mathbf{x}) - \mathbf{b}^T \mathbf{x} \to min, \qquad \mathbf{A} = \begin{pmatrix} 4 & 1 & 0 & & \dots & 0 \\ 1 & 4 & 1 & 0 & & \\ 0 & 1 & 4 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & \dots & & 0 & 1 & 4 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

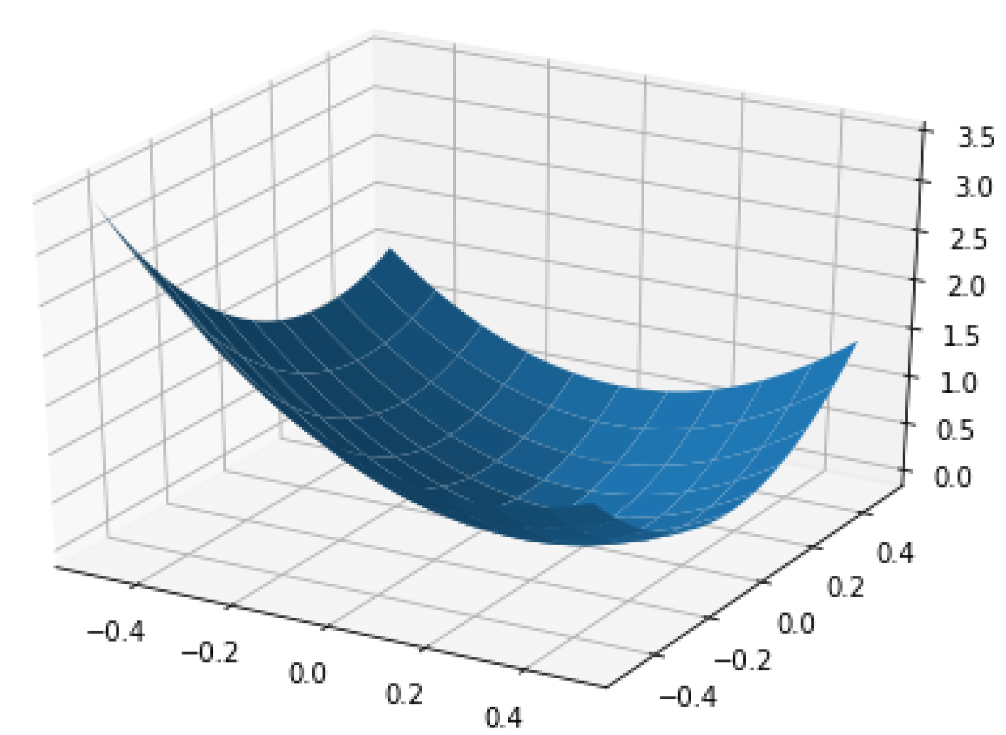

Figure 1: function in 2D.

Results

The first task

The rate of convergence of PCG can be seen on the graph.

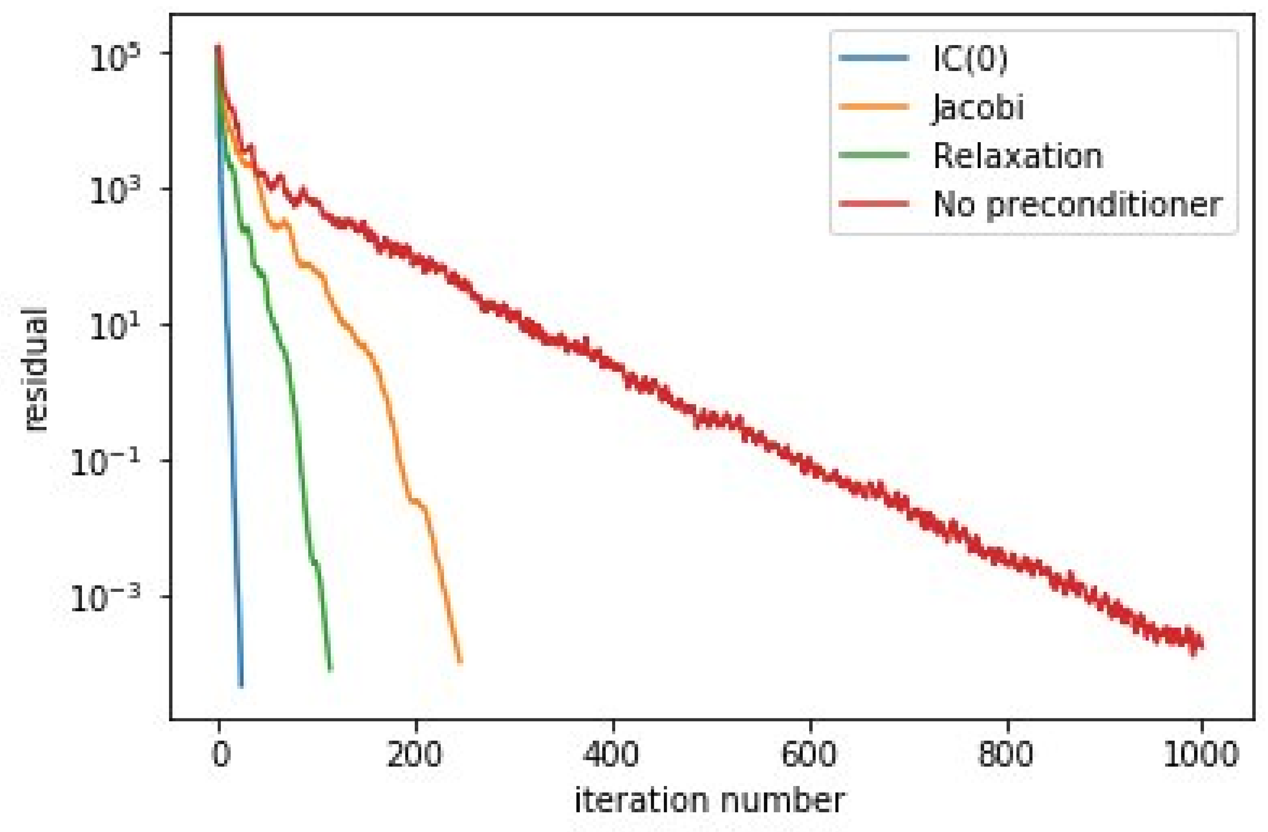

Figure 2: PCG convergence.

Here is the comparison of methods by time and number of iterations to achieve accuracy 10^{-4}:

- IC(0) : 0.26 s, 24 iterations
- Jacobi : 2.2 s, 246 iterations
- Relaxation, w = 1.0 : 1.2 s, 114 iterations
- Without preconditioner : 7.95 s 1000 iterations

The second task

The methods converge in approximately equal number of iterations, but Newton method using conjugate gradients works 6.72 seconds, while ordinary Newton method is 55.6 seconds! Moreover, during the the work on a task, a problem of machine overflow arises : starting from an arbitrary point the method breaks on the first iteration. But the region in which the modified Newton's method converges is quite wider than the region in which the non-linear conjugate gradient method converges.

References

[1] ETH Zurich. D-MATH. Homework Problem Sheet 12

[2] Preconditioning Techniques Analysis for CG Method.Huaguang Song

[3] https://colab.research.google.com/drive/1N-LVEs8mC4RTXPbFzqR700Veb2KbdQx5