Università della Svizzera italiana

Institute of Computing CI

Optimization Methods 2024

Student: Jeferson Morales Mariciano

Discussed with: Leonardo Birindelli

Midterm **Due date:** Sunday, 12 May 2024, 11:59 PM

Exercise 1

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f = \frac{1}{2}x^T A x - b^T x$ with A symmetric positive definite. Let x_m be the minimizer of the function f. Let v be an eigenvector of A, and let λ be the associated eigenvalue. Suppose that we use Stepeest Descent (SD) method to minimize f and the starting point for the SD algorithm is $x_0 = x_m + v$.

1.

Prove that the gradient at x_0 is $\nabla f(x_0) = \lambda v$.

2.

How many iterations does the SD method take to minimize the function f if we use the optimal step length? Show the computations behind your reasoning.

Matlab scripts are provided in */code* folder. The 2 main files to run are: *GD.m, Newton.m*. They handle both computations and visualization of the Rosenbrock's function with the corresponding methods.

Exercise 2

Given a starting point $x_0 \in \mathbb{R}^n$ and a set of conjugate directions $\{p_0, p_1, \ldots, p_{n-1}\}$, we generate the sequence $\{x_k\}$ by setting

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

where

$$\alpha_k = -\frac{r_k^T p_k}{p_k^T A p_k} \tag{2}$$

and r_k is the residual, as defined in class. Consider the following theorem:

Theorem 1. *For any $x_0 \in \mathbb{R}^n$ the sequence $\{x_k\}$ generated by the conjugate gradient direction algorithm (1) and (2), converges to the solution x^* of the linear system $Ax = b$ in at most n steps.*

Prove the theorem and explain carefully every step of your reasoning.

Exercise 3

Consider the linear system $Ax = b$, where the matrix A is a symmetric positive definitive diagonal matrix constructed in three different ways:

$A = diag([1 : 10])$

$A = diag(ones(1, 10))$

$A = diag([1, 1, 1, 3, 4, 5, 5, 5, 10, 10])$

$A = diag([1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0])$

1.

How many distinct eigenvalues has each matrix?

2.

Implement the CG method (CG.m).

Matlab scripts are provided in */code* folder. The main file to run is *CG.m*. It handles both computations and visualization of the Rosenbrock's function.

3.

Construct a right-hand side $b = rand(10, 1)$ and apply the Conjugate Gradient method to solve the system for each A.

4.

Compute the logarithm energy norm of the error (i.e. $log((x - x^*)^T A(x - x^*))$) for each matrix and plot it with respect to the number of iteration.

5.

Comment on the convergence of the method for the different matrices. What can you say observing the number of iterations obtained and the number of clusters of the eigenvalues of the related matrix?

Exercise 4

Consider the Chapter 4, "Trust-region methods" of the book *Numerical Optimization*, Nocedal and Wright.

1.

Explain Cauchy point method and Dogleg method, as well as the connection between them.

2.

Write down the Trust-Region algorithm, along with Dogleg and Cauchy-point computations.

3.

Consider the following lemma (Lemma 4.2, page 75 Numerical optimization, Nocedal and Wright)

Lemma 4.2. *Let B be positive definite. Then,*

(a) $\|\tilde{p}(\tau)\|$ is an increasing function of τ, and

(b) $m(\tilde{p}(\tau))$ is a decreasing function of τ

Read carefully the proof and explain in detail how each step is obtained.

Recalling:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U & 0 \leq \tau \leq 1 \\ p^U + (\tau - 1)\left(p^B - p^U\right) & 1 \leq \tau \leq 2 \end{cases}$$

First, let's prove that $\|\tilde{p}(\tau)\|$ is an increasing function with. $\tau \in [0, 1]$.

Proof. Let's define $\phi(\tau) = \|\tilde{p}(\tau)\|$. By computing $\frac{\partial \phi}{\partial \tau}$, since $\|p^U\|$ is a positive constant, the derivative is always positive and the function is monotonically increasing.

$$\phi(\tau) = \|\tilde{p}(\tau)\| = |\tau| \cdot \|p^U\| = \tau \|p^U\|$$
$$\frac{\partial \phi}{\partial \tau} = \|p^U\| \geq 0$$

thus, $\phi(\tau)$ is an increasing function with $\tau \in [0, 1]$. $\qquad\square$

Second, let's prove that $m(\tilde{p}(\tau))$ is a decreasing function with $\tau \in [0, 1]$.

Proof. Let's expand the function $m(\tilde{p}(\tau))$. After computing its $\frac{\partial m}{\partial \tau}$, we expand p^U by its definition and simplify the expression. Remember B is SPD. At the end, since $\tau \in [0, 1]$, the expression $(\tau - 1) \leq 0$, thus the function is monotonically decreasing.

$$m(\tilde{p}(\tau)) = f + g^T \tilde{p}(\tau) + \frac{1}{2}\tilde{p}(\tau)^T B \tilde{p}(\tau)$$
$$= f + \tau g^T p^U + \frac{1}{2}\tau^2 (p^U)^T B p^U$$
$$\frac{\partial m}{\partial \tau} = g^T p^U + \tau (p^U)^T B p^U$$
$$= g^T \cdot \left(-\frac{g^T g}{g^T B g}g\right) + \tau \left(-\frac{g^T g}{g^T B g}g\right)^T B \left(-\frac{g^T g}{g^T B g}g\right)$$
$$= g^T \cdot \left(-\frac{\|g\|^2}{g^T B g}g\right) + \tau \left(-\frac{\|g\|^2}{g^T B g}g\right)^T B \left(-\frac{\|g\|^2}{g^T B g}g\right)$$
$$= -\frac{\|g\|^4}{g^T B g} + \tau \frac{\|g\|^4}{(g^T B g)^2}g^T B g$$
$$= -\frac{\|g\|^4}{g^T B g} + \tau \frac{\|g\|^4}{g^T B g}$$
$$= \underbrace{(\tau - 1)}_{\leq 0} \underbrace{\frac{\|g\|^4}{g^T B g}}_{\geq 0} \leq 0$$

thus, $m(\tilde{p}(\tau))$ is a decreasing function with $\tau \in [0, 1]$. $\qquad\square$

Third, let's prove that $\|\tilde{p}(\tau)\|$ is an increasing function with $\tau \in [1, 2]$.

Proof. Let's define $\alpha = \tau - 1$, and $\phi(\alpha) = \frac{1}{2}\|\tilde{p}(\alpha)\|^2$ with $\alpha \in [0, 1]$. Inside $\phi(\alpha)$ we defined the squared norm $\|\cdot\|^2$ to delete the square root from the calculations. Then, simplify the expression by collecting using the square of sum which is the reason why we have the scalar $\frac{1}{2}$ in $\phi(\alpha)$. Remember B is SPD. Afterwards, we compute $\frac{\partial \phi}{\partial \alpha}$ and then prove $\left(p^U\right)^T \left(p^B - p^U\right) \geq 0$ using the Cauchy-Schwarz inequality to show that ϕ is monotonically increasing.

$$\phi(\alpha) = \frac{1}{2}\|\tilde{p}(\alpha)\|^2$$
$$= \frac{1}{2}\left\|p^U + \alpha\left(p^B - p^U\right)\right\|^2$$
$$= \frac{1}{2}\left\|p^U\right\|^2 + \alpha\left(p^U\right)^T\left(p^B - p^U\right) + \frac{1}{2}\alpha^2\left\|p^B - p^U\right\|^2$$
$$\frac{\partial \phi}{\partial \alpha} = \left(p^U\right)^T\left(p^B - p^U\right) + \underbrace{\alpha\left\|p^B - p^U\right\|^2}_{\geq 0}$$
$$\geq \left(p^U\right)^T\left(p^B - p^U\right)$$
$$\left(p^U\right)^T\left(p^B - p^U\right) = -\left(p^U\right)^T\left(p^U - p^B\right)$$
$$= \frac{g^T g}{g^T B g}g^T \cdot \left[\left(-\frac{g^T g}{g^T B g}g\right) - \left(-B^{-1}g\right)\right]$$
$$= \frac{\|g\|^2}{g^T B g}g^T \cdot \left(B^{-1}g - \frac{\|g\|^2}{g^T B g}g\right)$$
$$= \frac{\|g\|^2}{g^T B g}g^T B^{-1}g \cdot \left(1 - \frac{\|g\|^2}{g^T B g B^{-1}g}g\right)$$
$$= \underbrace{\frac{\|g\|^2}{g^T B g}g^T B^{-1}g}_{\geq 0} \cdot \left(1 - \frac{\|g\|^4}{\left(g^T B g\right)\left(g^T B^{-1}g\right)}\right)$$
$$\text{let } u = B^{-1}g, v = Bg \Rightarrow \left|u^T v\right|^2 \leq \|u\|^2 \cdot \|v\|^2$$
$$\Rightarrow \left|g^T B^{-1} B g\right|^2 \leq \left\|B^{-1}g\right\|^2 \cdot \|Bg\|^2$$
$$\Rightarrow \|g\|^4 \leq \left(g^T B^{-1}g\right)\left(g^T B g\right)$$
$$\Rightarrow \frac{\|g\|^4}{\left(g^T B g\right)\left(g^T B^{-1}g\right)} \leq 1$$
$$\Rightarrow 1 - \frac{\|g\|^4}{\left(g^T B g\right)\left(g^T B^{-1}g\right)} \geq 0$$

$\square$

Lastly, let's prove that $m(\tilde{p}(\tau))$ is a decreasing function with $\tau \in [1, 2]$.

Proof. the proof

$\square$