



- 1 Consider the quadratic function

$$f(x) = \frac{1}{2}x^T Ax - b^T x,$$

where $A \in \mathbb{R}^{n \times n}$ is a symmetric and positive definite matrix and $b \in \mathbb{R}^n$.

- a) Let $p \in \mathbb{R}^n$ be a direction satisfying the inequality $\nabla f(x)^T p < 0$. Compute analytically the steplength $\alpha_{x,p}$, which solves the linesearch problem $\min_{\alpha > 0} f(x + \alpha p)$
- b) Let $x, p \in \mathbb{R}^n$ and $\alpha_{x,p} > 0$ be as in the previous question. Show that the steplength $\alpha_{x,p}$ satisfies the strong Wolfe conditions if and only if $c_1 \leq 1/2$.
- c) Let $A = Q\Lambda Q^T$ be the eigenvalue decomposition of A , where Λ is a diagonal matrix with eigenvalues on the diagonal, and columns of Q are the orthonormal eigenvectors of A . In particular, $Q^T Q = I$, where $I \in \mathbb{R}^{n \times n}$ is the identity matrix.

Show that applying the steepest descent method with exact linesearch to the problem $\min_{x \in \mathbb{R}^n} 0.5x^T Ax - b^T x$ is equivalent to applying the steepest descent method with exact linesearch to $\min_{y \in \mathbb{R}^n} 0.5y^T \Lambda y$, in the following sense: if $x_0 = Qy_0 + A^{-1}b$ then the iterates generated by the two methods satisfy the same relation, $x_k = Qy_k + A^{-1}b$, $k \geq 1$.

In this sense, the behaviour of the steepest descent method is insensitive with respect to translation or orthogonal transformation of coordinates.

- 2 Let f be twice continuously differentiable in a vicinity of $x_0 \in \mathbb{R}^n$. Assume that $\nabla^2 f(x_0)$ is positive definite and consider the Newton's direction $p_x = -[\nabla^2 f(x_0)]^{-1} \nabla f(x_0)$ together with the unit Newton's step $x_1 = x_0 + p_x$.

Let us now perform an affine transformation (translation, rotation, and scaling) of coordinates $x = By + c$, where $B \in \mathbb{R}^{n \times n}$ is a non-singular matrix (not necessarily orthogonal), and $c \in \mathbb{R}^n$ is some vector. Demonstrate that Newton's method is insensitive with respect to such transformations: that is, if $g(y) = f(By + c) = f(x)$, $x_0 = By_0 + c$, and finally $y_1 = y_0 - [\nabla^2 g(y_0)]^{-1} \nabla g(y_0)$ then $x_1 = By_1 + c$.

- 3 Let $A \in \mathbb{R}^{n \times n}$ be an SPD matrix with the eigenvalue decomposition $A = Q\Lambda Q^T$, and let $b \in \mathbb{R}^n$ be an arbitrary vector. We put $x^* = A^{-1}b$ to be the optimal solution of the quadratic unconstrained minimization problem $\min_{x \in \mathbb{R}^n} 0.5x^T Ax - b^T x$. Suppose that the eigenvalues of A are sorted as $0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$. During the lecture we have discussed that for starting point of the type $x_0 = x^* + \lambda_1^{-1}q_1 + \lambda_n^{-1}q_n$, where

q_i are orthonormal eigenvectors of A (columns of Q) corresponding to eigenvalues λ_i , the steepest descent method with exact linesearch for this problem generates iterates satisfying

$$\|x_k - x^*\| = \left(\frac{\lambda_n - \lambda_1}{\lambda_n + \lambda_1} \right)^k \|x_0 - x^*\|,$$

which converges to zero linearly, and arbitrarily slowly for large condition numbers $\text{cond}(A) = \lambda_n/\lambda_1$. Approximately, the number of iterations needed to achieve some prescribed tolerance scales proportionally to the condition number of A .

- a) Implement the steepest descent method with exact linesearch for this problem and verify the estimate above numerically.

Hint: one can generate random positive definite matrices for example as follows:

```
import numpy as np
N = 10
# generate NxN random matrix
X = np.random.randn(N,N)
# generate NxN orthogonal matrix from it
Q = np.linalg.qr(X)[0]
# generate some random eigenvalues between lam_min and lam_max
lam_min = 1.0
lam_max = 100.0
lmbda = lam_min + (lam_max-lam_min)*np.sort(np.random.rand(N))
lmbda[0] = lam_min
lmbda[-1]=lam_max
Lambda = np.diag(lmbda)
A = np.matmul(Q,np.matmul(Lambda,Q.T))
# random vector
b = np.random.randn(N)
# A^{-1}b
xstar = np.linalg.solve(A,b)
# starting point
x0 = xstar + 1.0/lmbda[0]*Q[:,0] + 1.0/lmbda[-1]*Q[:, -1]
```

- b) Not everyone has given up on the steepest descent method. Consider for example the following accelerated version of the method due to Nesterov:

$$\begin{aligned} p_k &= -\nabla f(x_k), \\ y_{k+1} &= x_k + \lambda_n^{-1} p_k, \\ x_{k+1} &= s_1 y_{k+1} + s_0 y_k, \end{aligned}$$

where we put $y_0 = x_0$, $s_0 = -(\lambda_n^{1/2} - \lambda_1^{1/2})/(\lambda_n^{1/2} + \lambda_1^{1/2})$, and $s_1 = 1.0 - s_0$.

Implement this method and verify numerically, that the number of iterations needed to achieve some prescribed tolerance scales proportionally to the square root of the condition number of A , $\lambda_n^{1/2}/\lambda_1^{1/2}$.

- 4 Implement both the steepest descent method and the Newton's method with line-search satisfying Wolfe conditions (use a bisection algorithm for this).

Apply the method to minimizing the Rosenbrock function:

$$f(x, y) := 100(y - x^2)^2 + (1 - x)^2.$$

As Newton's direction is not necessarily a descent direction, we can simply use the steepest descent direction when the following inequality holds:

$$-\nabla f(x_k)^T p_k^{\text{Newton}} \leq \varepsilon \|\nabla f(x_k)\| \|p_k^{\text{Newton}}\|,$$

that is, when the angle between the Newton's direction and the steepest descent direction gets dangerously close to $\pi/2$ or exceeds this value (ε is a small positive parameter in the inequality above).

Verify numerically that the unit Newton's steps are accepted by the linesearch algorithm provided that the sufficient decrease parameter satisfies the inequality $0 < c_1 < 1/2$.