Exercises in Optimization
(ACM 40990 / ACM41030)

Dr Lennon Ó Náraigh

Exercises #2

Exercises #2 - More on Line-search Methods

1. In the notes (Chapter 6), it is shown that the Newton method satisfies:

$$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}_*\|_2 \le C\|\boldsymbol{x}_k-\boldsymbol{x}_*\|_2^2 \text{ whenever } \|\boldsymbol{x}_k-\boldsymbol{x}_*\|_2 < \delta.$$

If we choose

$$\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2 < \delta, \text{ and } \|\boldsymbol{x}_0-\boldsymbol{x}_*\| < \frac{1}{2C},$$

then

$$\frac{\|\boldsymbol{x}_1-\boldsymbol{x}_*\|_2}{\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2} \le C\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2 \le \tfrac{1}{2}.$$

The aim of this exercise is to show that these inequalities give rise to the following important result:

$$\frac{\|\boldsymbol{x}_k-\boldsymbol{x}_*\|_2}{\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2} \le \frac{1}{2^{2^k-1}}. \tag{1}$$

The proof of the inequality (1) can be obtained by the following sequence of steps:

(a) Write down inequalities for

$$\|\boldsymbol{x}_1-\boldsymbol{x}_*\|_2 \le \frac{1}{2^{\cdots}}\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2, \qquad \|\boldsymbol{x}_2-\boldsymbol{x}_*\|_2 \le \frac{1}{2^{\cdots}}\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2,$$

$$\|\boldsymbol{x}_3-\boldsymbol{x}_*\|_2 \le \frac{1}{2^{\cdots}}\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2.$$

(b) Hence, guess that the general term satisfies

$$\|\boldsymbol{x}_k-\boldsymbol{x}_*\| \le \frac{1}{2^{p_k}}\|\boldsymbol{x}_0-\boldsymbol{x}_*\|_2,$$

where

$$p_k = 2p_{k-1}+1. \tag{2}$$

(c) Equation (2) is a first-order **difference equation** with general solution $p_k = A + B\lambda^n$, where A, B, and λ are constants to be determined. Hence, show that p_k satisfies:

$$p_k = 2^k - 1, \qquad k > 1$$

with $p_1 = 1$.

(d) Conclude that

$$\|\boldsymbol{x}_k - \boldsymbol{x}_*\|_2 \le \frac{1}{2^{2^k-1}}\|\boldsymbol{x}_0 - \boldsymbol{x}_*\|_2,$$

and hence,

$$\lim_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}_*\|_2 = 0.$$

2. Show that if $0 < c_2 < c_1 < 1$, there may be no step lengths that satisfy the Strong Wolfe conditions.

   Hint: Consider the quadratic function

$$\phi(\alpha) = a + b\alpha + c\alpha^2,$$

   where $b < 0$ and $c > 0$.

3. Consider the one-dimensional function

$$\phi(\alpha) = f(\boldsymbol{x}_k + \alpha\boldsymbol{p}_k),$$

   where $\boldsymbol{p}_k$ is a descent direction – that is, $\phi'(0) < 0$ – so that our search can be confined to positive values of α. Find the value that minimizes $\phi(\alpha)$ in the case where the cost function is quadratic, specifically:

$$f(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x}\rangle + \tfrac{1}{2}\langle \boldsymbol{x}, B\boldsymbol{x}\rangle, \tag{3}$$

   where $\boldsymbol{a} \in \mathbb{R}^n$ and $B \in \mathbb{R}^{n\times n}$.

4. Consider the steepest decent method with exact line searches applied to the convex quadratic function in Equation (3).

   (a) Show that if the initial point is such that $\boldsymbol{x}_0 - \boldsymbol{x}_*$ is parallel to an eigenvector of B, then the steepest descent method will find the solution in one step.

   (b) Show that the Newton method always converges in exactly one step when the cost function is quadratic, i.e. takes the form (3).

5. Consider the optimization problem,

$$\min f(\boldsymbol{x}), \qquad f(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x}\rangle + \tfrac{1}{2}\langle \boldsymbol{x}, B\boldsymbol{x}\rangle,$$

   where now B is a specific 10×10 matrix and $\boldsymbol{a}$ is a specific 10×1 column vector. The numerical values of these arrays can be found in the spreadsheet `OP_10x10.csv`:

   - The spreadsheet contains a 10×1 array which corresponds to the vector $\boldsymbol{a}$;
   - The spreadsheet contains a 10×10 array B_0.

   The array B is obtained from B_0 by the following sequence of steps:

   (i) Symmetrize B_0:

$$B_0 \to (B_0 + B_0^T)/2;$$

(ii) Scale B_0:

$$B_0 \to B_0 / \max(|B_0|)$$

(iii) Generate a positive-definite matrix:

$$B_0 \to (B_0^T) B_0.$$

The end result of this sequence of operations is the matrix B.

Hence,

(a) Find the minimizer $\boldsymbol{x}_*$ numerically, using the steepest-descent and Newton algorithms.

(b) Why is the convergence so poor in the case of the steepest-descent algorithm?