

Nonlinear Optimization

Alberto Quaini

Exercise 1

Let

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto a^T x + b,$$

for some $a \in \mathbb{R}^n, b \in \mathbb{R}$. Suppose that f has a minimizer at $x^* \in \mathbb{R}^n$. Since any point in \mathbb{R}^n is an interior point, by the FONC, $Df^T(x) = a$ must be zero, which implies that f is constant. Suppose now that a is nonzero. Then, by the same condition we know that there is no minimizer.

Exercise 2

Suppose $b \in \mathbb{R}^m, A \in M_{m \times n}(\mathbb{R})$ is a nonzero matrix and

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, x \mapsto \|Ax - b\|_2.$$

Notice that minimizing f is the same as minimizing f^2 . Also,

$$\begin{aligned} f^2(x) &= \|Ax - b\|_2^2 \\ &= (Ax - b)^T (Ax - b) \\ &= x^T A^T A x - 2b^T A x + b^T b. \end{aligned}$$

Since $b^T b$ is constant, minimizing f^2 is equivalent to minimizing $g(x) = x^T A^T A x - 2b^T A x$.

Notice that g is continuous and continuously differentiable, $Dg(x) = 2x^T A^T A - 2b^T A$ and $D^2g(x) = 2A^T A$, which is positive definite. First, suppose that x^* is a minimizer of g . Then FONC implies $A^T A x^* = A^T b$. Now suppose that x^* is such that $A^T A x^* = A^T b$. Then $Dg(x^*)$ is zero and by the SOSC we have that x^* is a minimizer of g .

Exercise 3

In this chapter, the book covers basic iterative methods to solve unconstrained optimization problems with a nonlinear objective function, say f , mapping \mathbb{R}^n to the real line. All of them require an initial value, say x_0 , and produce a finite sequence of approximate solutions, $(x_1, \dots, x_i, \dots, x_N)$, which hopefully get closer and closer to the true solution, say x , until a certain criterion is met. There can be multiple criteria stopping the iteration. One can be to stop when the difference between two successive iterations gets lower than a certain small tolerance with respect to some norm. Another can be the same idea applied to the difference between function values computed at two successive iterations. Some other criteria are possible. When the chosen criterion (virtually all of them) is not guaranteed to be met in a finite number of iterations, a stopping criterion imposing some maximum number of iterations is employed.

Gradient descent methods

The basic idea of gradient descent methods is to compute at each iteration the direction of steepest descent given by the negative gradient, $-Df^T(x_i)$, and move the next iteration along this direction, $x_{i+1} = x_i - \alpha Df^T(x_i)$, where α is some number measuring the step. α is usually set via line search methods, trying to pick the optimal step, or via cheaper methods. These methods are useful to get closer to x when the initial guess is far away from it, but in some cases they can converge very slowly. Finally, these methods require that the gradient exists, therefore are applicable just to differentiable objective functions.

Newton and Quasi-Newton methods

Newton's method performs a local quadratic approximation of the objective function at each iteration and moves towards the minimizer of the quadratic approximation. It can be very fast for small dimensional problems, where the Hessian $D^2f(x_i)$ is positive definite, or when $(D^2f(x_i))^{-1}Df^T(x_i)$ can be computed cheaply and reliably. If the Hessian is not positive definite and/or the dimension of the problem is large, Then Newton's method becomes computationally expensive and variants such as Quasi-Newton methods or BFGS can be employed instead.

Conjugate gradient methods

Gradient methods are computationally cheap but have slow convergence, whereas Newton's methods may converge fast but require an expensive computation of the Hessian at each iteration. Conjugate gradient methods fall between these two methods. At each iteration, these methods move along so called Q -conjugate directions, which are cheaply computed and do not require much information from previous steps. Conjugate gradient methods are especially suited for solving large quadratic problems of the form $x^T Q x - b^T x + c$, where Q is symmetric, positive definite and possibly sparse.

Nonlinear least squares methods

When the objective function can be written as the sum of squares of a set of residuals, then Gauss-Newton method, an adaptation of Newton's method, can be employed.

Exercise 4

We first prove the "only if" direction. Suppose x_0 is such that $Df^T(x_0) = Qx_0 - b$ is an eigenvector of Q , that is, $Q(Qx_0 - b) = \lambda(Qx_0 - b)$ for some real λ . Then

$$\begin{aligned} x_1 &= x_0 - \alpha_0 Q(Qx_0 - b) \\ &= x_0 - \alpha_0 \lambda (Qx_0 - b). \end{aligned}$$

and

$$\alpha_0 = \arg \min_a f(x_1).$$

If $\alpha = 1/\lambda^2$, then:

$$\begin{aligned} Qx_1 &= Q(x_0 - \alpha_0 \lambda (Qx_0 - b)) \\ &= Qx_0 - \alpha_0 \lambda^2 (Qx_0 - b) \\ &= Qx_0 - Qx_0 + b \\ &= b. \end{aligned}$$

Since $Qx_1 = b$, x_1 is the minimum of the function, α_0 was chosen righteously and the algorithm converged in one step.

We now prove the "if" direction. Suppose the algorithm converges in one step. Then $Qx_1 = b$, or

$$Q(x_0 - \alpha_0 Q(Qx_0 - b)) = b.$$

Now notice that:

$$\begin{aligned} (I - \alpha_0 Q)(Qx_0 - b) &= Qx_0 - b - \alpha_0 Q(Qx_0 - b) \\ &= Q(x_0 - \alpha_0 Q(Qx_0 - b)) - b \\ &= Qx_1 - b \\ &= 0. \end{aligned}$$

Thus, $Qx_0 - b$ is an eigenvector of Q with eigenvalue α_0 .

Exercise 5

Suppose $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is C^1 . The steepest descent iteration is $x_{i+1} = x_i - \alpha_{i+1} Df^T(x_i)$, where

$$\alpha_{i+1} = \arg \min_a f(x_i - a Df^T(x_i)),$$

and so $Df(x_i - \alpha_{i+1} Df^T(x_i)) Df^T(x_i) = 0$. Then, notice that

$$\begin{aligned} (x_{i+2} - x_{i+1})^T (x_{i+1} - x_i) &= \alpha_{i+1} \alpha_i Df(x_{i+1}) Df^T(x_i) \\ &= \alpha_{i+1} \alpha_i Df(x_i - \alpha_{i+1} Df^T(x_i)) Df^T(x_i) \\ &= 0. \end{aligned}$$

Exercise 10

Suppose $f(x) = x^T Q x / 2 - b^T x$, where $Q \in M_n(\mathbb{R})$ is symmetric and positive definite, $b \in \mathbb{R}^n$. Notice that $Df^T(x) = Qx - b$ and $D^2 f(x) = Q$. By the SONC we know that $x^* = Q^{-1}b$ is a strict local minimizer. Take an initial guess $x_0 \in \mathbb{R}^n$. Then the Newton's method implies

$$\begin{aligned} x_1 &= x_0 + (D^2 f(x_0))^{-1} Df^T(x_0) \\ &= x_0 + Q^{-1}(Qx_0 - b) \\ &= Q^{-1}b, \end{aligned}$$

which is the expression for the minimizer of f , x^* .

Exercise 12

Suppose $A \in M_n(\mathbb{F})$ has eigenvalues $\lambda_1, \dots, \lambda_n$, with corresponding eigenvectors x_1, \dots, x_n . Suppose also that $B = A + \mu I$. Then $B \in M_n(\mathbb{F})$ and for $i = 1, \dots, n$, we have

$$\begin{aligned} Bx_i &= Ax_i + \mu Ix_i \\ &= \lambda_i x_i + \mu x_i \\ &= (\lambda_i + \mu)x_i, \end{aligned}$$

so that B has the same eigenvectors of A , with corresponding eigenvalues $\lambda_1 + \mu, \dots, \lambda_n + \mu$.

Exercise 15

First consider these identities:

$$\begin{aligned} U + UCV A^{-1}U &= UC(C^{-1} + VA^{-1}U) \\ &= (A + UCV)A^{-1}U \end{aligned}$$

and

$$(A + UCV)^{-1}UC = A^{-1}U(C^{-1} + VA^{-1}U)^{-1}.$$

Then,

$$\begin{aligned} A^{-1} &= (A + UCV)^{-1}(A + UCV)A^{-1} \\ &= (A + UCV)^{-1}(I + UCV A^{-1}) \\ &= (A + UCV)^{-1} + (A + UCV)^{-1}UCVA^{-1} \\ &= (A + UCV)^{-1} + A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}, \end{aligned}$$

and the proof is concluded.

Exercise 18

Let $Q \in M_n(\mathbb{R})$ satisfy $Q > 0$, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$ and let

$$f : \mathbb{R}^n \rightarrow \mathbb{R}, \quad x \mapsto x^T Q x - b^T x + c.$$

Given an initial guess x_0 and Q -conjugate directions d_0, d_1, \dots, d_{n-1} in \mathbb{R}^n , the optimal line search solution for $x_{k+1} = x_k + \alpha_k d_k$ is defined as

$$\begin{aligned} \alpha_k &= \arg \min_a f(x_k + ad_k) \\ &= \arg \min_a \tilde{x}_k(a)^T Q \tilde{x}_k(a) / 2 - b^T \tilde{x}_k(a) + c, \end{aligned}$$

where $\tilde{x}_k : \mathbb{R} \rightarrow \mathbb{R}^n$, $a \mapsto x_k + ad_k$. Notice that the FONC for finding α_k implies:

$$\begin{aligned} d_k^T Q \tilde{x}_k(\alpha_k) &= d_k^T b \iff \\ d_k^T Q(x_k + \alpha_k d_k) &= d_k^T b_k \iff \\ \alpha_k &= \frac{d_k^T r_k}{d_k^T Q d_k}, \end{aligned}$$

where $r_k = b - Qx_k$.