6.215/6.255J/15.093J/IDS.200J: Optimization Methods

Problem Set 5

Due: December 5, 2021

Problem 1: (10 points) Consider the quadratic optimization problem:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} x^T A x + b^T x + c,$$

where A is a positive definite matrix.

(Hint: what are the gradient and minimum of this function?)

(a) From an arbitrary starting position x_0, how many steps will Newton's method take to converge. Justify your answer with a formal proof.

(b) Under what conditions on x_0 will gradient descent (steepest descent) converge to the optimal solution x^* in one step? Justify your answer with a formal proof. (Hint: You might find it helpful to use eigenvalues and eigenvectors.)

Solution

(a) Since A is symmetric (positive definiteness is only defined for symmetric matrices), the gradient of the objective function is

$$\nabla f(x) = Ax + b$$

and the Hessian is

$$\nabla^2 f(x) = A.$$

Let us characterize the minimum of this function. Since f is convex (because the Hessian is positive semi-definite for all x), a necessary and sufficient condition for a global minimum is $\nabla f(x) = 0$. Since A is positive definite, A is invertible. Thus, the only point that satisfies $\nabla f(x) = 0$ is

$$x^* = A^{-1} b$$

Take an arbitrary point x_0. We know the gradient and Hessian at x_0.

$$\nabla f(x_0) = Ax_0 + b,$$

$$\nabla^2 f(x_0) = A.$$

So using Newton's method,

$$\begin{aligned} x_1 &= x_0 - A^{-1}(Ax_0 + b) \\ &= -A^{-1} b \\ &= x^* \end{aligned}$$

Therefore, Newton's method always converges in one step.

(b) Gradient descent converges to the optimal solution x^* in one step if and only if there exists a $\lambda > 0$ such that

$$x - \lambda(Ax + b) = -A^{-1}b.$$

Rearranging terms, this happens if and only if there exists a $\lambda > 0$ such that

$$x + A^{-1}b = \lambda(Ax + b).$$

Multiplying both sides by A, this happens if and only if there exists a $\lambda > 0$ such that

$$Ax + b = \lambda A(Ax + b).$$

Since A is positive definite, all eigenvalues are positive, and this happens if and only if either $Ax + b = 0$ or $Ax + b$ is an eigenvector of A.

It is fine to stop here, but we can go even further. Let $\tilde{x}$ be an arbitrary eigenvector of A with eigenvalue λ. $Ax + b = \tilde{x}$ if and only if $x = A^{-1}\tilde{x} - A^{-1}b$. Eigenvectors of A^{-1} are eigenvectors of A with reciprocal eigenvalues. So $Ax + b$ is an eigenvector of A if and only if $x = \frac{1}{\lambda}\tilde{x} - A^{-1}b$, where $\tilde{x}$ is an eigenvector of A. But $\frac{1}{\lambda}\tilde{x}$ is just an arbitrary eigenvector of A because an eigenvector multiplied by a constant is still an eigenvector with the same eigenvalue. And $-A^{-1}b$ is just x^*. Therefore, gradient descent starting at x can converge to x^* in one step if and only if

$$x = \tilde{x} + x^*$$

where $\tilde{x}$ is an eignevector of A (or 0).

Problem 2: (10 points) Classify the following statements as true or false. All answers must be well-justified, either through a short explanation, or a counterexample. If you think a question is ambiguous or not clear, please explain your assumptions in detail.

(a) For a nonlinear optimization problem, if Newtons method converges, then it converges to a local minimum.

(b) The sequence $x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1}\nabla f(x_k)$ generated by Newtons method, when applied to the function $f(x) = x^4$, converges quadratically to zero.

Solution

(a) False. Newton's method can converge to global maxima.

(b) False. For Newton's method to converge quadratically, we need the Hessian to be nonsingular. For this example, we have $x_{k+1} = x_k - (12x_k^2)^{-1}4x_k^3 = \frac{1}{3}x_k$, so $\lim_{k\to\infty} \frac{|x_{k+1}-x^*|}{|x_k-x^*|^2} = \lim_{k\to\infty} \frac{1}{3x_k} = \infty$. Thus, the sequence doesn't converge quadratically (it converges only linearly).

Problem 3: (10 points) Consider a set of n points $\{(x_1, y_1), ..., (x_n, y_n)\}$ in the plane. We want to find a point (x, y) such that the sum of squares of Euclidean distances from this point to all the other points is minimized.

(a) Give an nonlinear optimization formulation of this problem.

(b) Is the objective function differentiable? Is this a convex optimization problem?

(c) Write the corresponding optimality conditions.

(d) Give a closed-form expression for (x, y).

Solution

(a) The objective function is given by:

$$\min_{x,y} \quad \sum_{i=1}^{n}(x_i - x)^2 + \sum_{i=1}^{n}(y_i - y)^2.$$

(b) Yes, this is a differentiable problem because the gradient exists and is equal to

$$\nabla f = [-2\sum_{i=1}^{n}(x_i - x), -2\sum_{i=1}^{n}(y_i - y)]^T$$

The problem is also convex: one way to see this is to take the Hessian and note that it is diagonal with non-negative diagonal entries, i.e., positive semidefinite.

(c) Because the problem is convex, it suffices to consider the first order stationary conditions, which are:

$$\nabla f = [-2\sum_{i=1}^{n}(x_i - x), -2\sum_{i=1}^{n}(y_i - y)]^T = [0, 0]^T$$

(d) The optimal solution is given by averaging the data co-ordinate wise, i.e., setting

$$x^* = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$y^* = \frac{1}{n}\sum_{i=1}^{n} y_i.$$

Problem 4: (10 points)

(a) Consider the optimization problem

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & x_1^{-1}x_2^2x_3^3 \\ \text{s.t.} \quad & x_1^{11}x_2^{-12}x_3^{13} \le 14 \\ & x_1^{15}x_2^{16}x_3^{-17} \le 18 \\ & x_1, x_2, x_3 \ge 1 \end{aligned} \tag{1}$$

Show that this is not a convex optimization problem as written (without any re-formulations).

(b) Explain how to use linear programming to compute both the optimal value of (1) and an optimal solution. (Hint: change variables via $x_i = e^{z_i}$, and use properties of log functions.)

(c) Reformulate the non-convex optimization problem

$$\begin{aligned} \min_{x_1,x_2,x_3} \quad & x_1^{-1}x_2^2x_3^3 + 5x_1^4x_2^5x_3^{-6} \\ \text{s.t.} \quad & x_1^{11}x_2^{-12}x_3^{13} \le 14 \\ & x_1^{15}x_2^{16}x_3^{-17} + 7x_1^{18}x_2^{-19}x_3^{20} \le 21 \\ & x_1, x_2, x_3 \ge 1 \end{aligned} \tag{2}$$

as a convex optimization problem. Clearly explain why your reformulation works. (Hint: you may use without proof that the function $y \mapsto \log(\sum_{i=1}^{k} e^{y_i})$ on $\mathbb{R}^k$ is convex.)

Solution

(a) The objective function and constraints are not convex as written. This can be seen, for instance, by restricting to lines of the form x_j = constant.

(b) After the change of variables $x_i = e^{z_i}$, (1) equals

$$\begin{aligned} \min_{z_1,z_2,z_3} \quad & e^{-z_1+2z_2+3z_3} \\ \text{s.t.} \quad & e^{11z_1-12z_2+13z_3} \leq 14 \\ & e^{15z_1+16z_2-17z_3} \leq 18 \\ & z_1, z_2, z_3 \geq 0 \end{aligned}$$

By taking logarithms in the constraints and removing the exponential from the objective, the above optimization problem has value equal to the exponential of the value of the following linear program:

$$\begin{aligned} \min_{z_1,z_2,z_3} \quad & -z_1 + 2z_2 + 3z_3 \\ \text{s.t.} \quad & 11z_1 - 12z_2 + 13z_3 \leq \log 14 \\ & 15z_1 + 16z_2 - 17z_3 \leq \log 18 \\ & z_1, z_2, z_3 \geq 0 \end{aligned}$$

Now solve this LP to obtain some optimal solution z^* with corresponding value v^*. Then x^* is an optimal solution to the original problem (1) where $x_i^* = e^{z_i^*}$, with corresponding value e^{v^*}.

(c) After the change of variables $x_i = e^{z_i}$, (2) equals

$$\begin{aligned} \min_{z_1,z_2,z_3} \quad & e^{-z_1+2z_2+3z_3} + e^{4z_1+5z_2-6z_3+\log 5} \\ \text{s.t.} \quad & e^{11z_1-12z_2+13z_3} \leq 14 \\ & e^{15z_1+16z_2-17z_3} + e^{18z_1-19z_2+20z_3+\log 7} \leq 21 \\ & z_1, z_2, z_3 \geq 0 \end{aligned}$$

By taking logs in the objective and constraints, this is equal to

$$\begin{aligned} \exp \min_{z_1,z_2,z_3} \log e^{-z_1+2z_2+3z_3} + e^{4z_1+5z_2-6z_3+\log 5} & \\ \text{s.t.} \quad 11z_1 - 12z_2 + 13z_3 \leq \log 14 & \\ \log\left(e^{15z_1+16z_2-17z_3} + e^{18z_1-19z_2+20z_3+\log 7}\right) - \log 21 \leq 0 & \\ z_1, z_2, z_3 \geq 0 & \end{aligned}$$

Note that the objective function is convex (since it is log-sum-exp). The first constraint is affine. The second constraint is a convex function less than 0. Thus, this is a convex optimization problem.