Introduction to Optimization Homework (week 4)

Dimitri Tabatadze · Thursday 04-04-2024

Problem IV.1: Minimization and the Armijo Step Size Rule

Let $f(x) = \frac{1}{2}x^T Cx + c^T x$ with a symmetric and positive definite matrix $C \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}^n$. Moreover, let $s \in \mathbb{R}^n$ be a descent direction of f at a point $x \in \mathbb{R}^n$ and $\sigma^* \geq 0$ be the exact line search step size, i.e. $f(x + \sigma^*) = \min_{\sigma \geq 0} f(x + \sigma s)$.

a) Show that f is strictly convex.
b) Show that $\sigma^* > 0$ holds.
c) What form (linear, quadratic, ...) does the function $\phi(\sigma) = f(x + \sigma s)$ have? Use Taylor expansion of ϕ about $\sigma = 0$ and deduct that σ^* is well defined and uniquely determined.
d) Show that for all $\gamma \in \left(0, \frac{1}{2}\right]$ the choice $\sigma = \sigma^*$ satisfies the sufficient decrease condition

$$f(x + \sigma s) - f(x) \leq \gamma\sigma\nabla f(x)^T s,$$

though, that this is not the case for $\gamma > \frac{1}{2}$.
e) Sketch the graph of ϕ and use it to illustrate the statement discussed in d)

Solution

a) The hessian of f is $H_f(x) = C$ which is given to be symmetric positive definite implying that the function f is strictly convex.
b) Since s is a descent direction of f we know that $\nabla f(x)^T s < 0$. By continuity of f, we know that there exists $\varepsilon > 0$ such that $f(x + \varepsilon s) < f(x)$ therefore the exact line search would have found that ε, i.e $\sigma^* \geq \varepsilon > 0$.
c) The taylor expansion follows

$$\begin{aligned}\phi(\sigma) &= f(x) + \sigma\nabla f(x)^T s + \frac{1}{2}\sigma^2 s^T H_f(x)s \\ &= f(x) + \sigma\left(x^T C + c^T\right)^T s + \frac{1}{2}\sigma^2 s^T Cs\end{aligned}$$

which is a quadratic equation in terms of σ. We have to show that

$$\frac{\overbrace{-\nabla f(x)^T s}^{>0 \text{ (descent direction)}}}{\underbrace{s^T H_f(x)s}_{>0 \text{ (positive definite)}}} > 0.$$

d) We first express

$$\phi'(\sigma) = \sigma s^T Cs + \nabla f(x)^T s$$

and now we can rewrite the condition to be

$$\phi(\sigma) - \phi(0) \leq \gamma\sigma\phi'(0) \quad \Longrightarrow \quad \frac{\phi(\sigma) - \phi(0)}{\sigma} \leq \gamma\phi'(0)$$

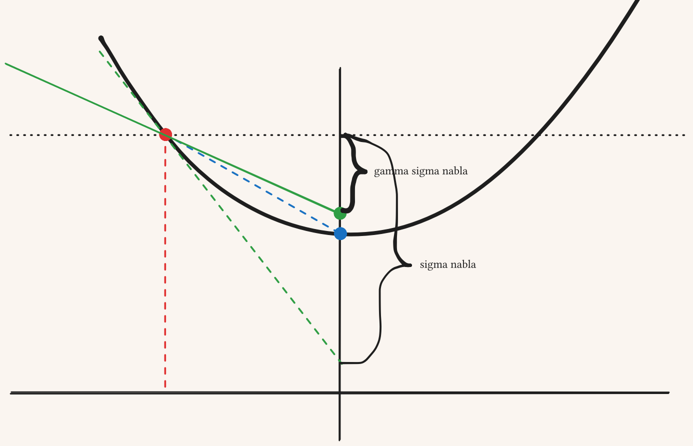
Let $g(\sigma) = \phi(\sigma^* - \sigma)$ allowing us to write

$$\frac{g(0) - g(\sigma^*)}{\sigma^*} \leq \gamma\phi'(0) \quad \Longrightarrow \quad \frac{g(\sigma^*) - g(0)}{\sigma^*} \geq \gamma g'(\sigma^*)$$

Now let $g(\sigma) = a\sigma^2 + c$ (with $g'(\sigma) = 2a\sigma$) since we know that $g(x)$ is a quadratic function and the minimum is at $\sigma = 0$ the coefficient of the linear term will be 0. So

$$a\sigma^* = \frac{a(\sigma^*)^2 + c - c}{\sigma^*} = \frac{g(\sigma^*) - g(0)}{\sigma^*} \geq \gamma g'(\sigma^*) = \gamma(2a\sigma^*)$$
$$\Downarrow$$
$$\cancel{a}\sigma^* \geq 2\gamma\cancel{a}\sigma^*$$
$$\Downarrow$$
$$\frac{1}{2} \geq \gamma$$

e)



$\blacksquare$

Problem IV.2: Application of the Wolfe-Powell Rule

a) The first Wolfe-Powell condition (sufficient decrease condition) requires
$$f(x^k + \sigma_k s^k) - f(x^k) \leq \gamma \sigma_k \nabla f(x^k)^T s^k.$$
What is the maximum step length σ_k that satisfies the condition, given that $f(x) = 5 + x_1^2 + x_2^2$, $x^k = \binom{-1}{-1}$, $s^k = \binom{1}{0}$, and $\gamma = 10^{-4}$.

b) Given a general descent algorithm with the Wolfe-Powell step size rule, provide an example to show that the set
$$\left\{ t \in \mathbb{R} : \begin{array}{l} \nabla f(x + \sigma s)^T s \geq \eta \nabla f(x)^T s \text{ and} \\ f(x + \sigma s) - f(x) \leq \gamma \sigma \nabla f(x)^T s \end{array} \right\}$$
may be empty if $0 < \gamma < \eta < 1$. I.e., that under these conditions no step size can be found. *Hint:* Think of a one-dimensional example.

Solution

a)
$$f(x^k + \sigma_k s^k) - f(x^k) \leq \gamma \sigma_k \nabla f(x^k)^T s^k$$
$$\Downarrow$$
$$f\left(\binom{-1}{-1} + \sigma_k \binom{1}{0}\right) - f\left(\binom{-1}{-1}\right) \leq 10^{-4} \sigma_k \nabla f\left(\binom{-1}{-1}\right)^T \binom{1}{0}$$

$$\Downarrow$$

$$f(\sigma_k - 1, -1) - f(-1, -1) \leq \sigma_k 10^{-4} \cdot \nabla f(-1, -1)^T \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\Downarrow$$

$$\cancel{5} + \sigma_k^2 - 2\sigma_k + \cancel{1 + 1} - \cancel{5} - \cancel{1 - 1} \leq \sigma_k(-2 \cdot 10^{-4})$$

$$\Downarrow$$

$$\sigma_k(2 - 2 \cdot 10^{-4}) - \sigma_k^2 \geq 0$$

$$\Downarrow$$

$$\sigma_k((2 - 2 \cdot 10^{-4}) - \sigma_k) \geq 0$$

$$\Downarrow$$

$$(2 - 2 \cdot 10^{-4}) \geq \sigma_k \geq 0$$

b) Consider the example $f(x) = x$, then $\nabla f(x)^T s = -1$. Take $x = 1, s = -1$. We these into the conditions to get

$$\begin{cases} 1 \cdot (-1) \geq \eta \cdot (-1) \\ 1 - \sigma - 1 \leq -\gamma\sigma \end{cases} \implies \begin{cases} 1 \leq \eta \\ 1 \geq \gamma \end{cases}$$

and since $\eta < 1$ this is a counter example.

■

Problem IV.3:

Let $f(x) = \frac{1}{2}x^T Cx + c^T x + \gamma$ with a symmetric positive definite matrix $C \in \mathbb{R}^{n \times n}$, $c \in \mathbb{R}^n$, and $\gamma \in \mathbb{R}$. Show that Gradient Descent Method with exact line search step size reaches the global minimum $\bar{x} = -C^{-1}c$ in exactly one step if the initial point x^0 is chosen such that $\nabla f(x^0)$ is an eigenvector of C. What does this imply for a strategy to choose the initial point in a diagonally scaled Gradient Descent Method?

Solution

We find

$$\nabla f(x) = Cx + c$$

and from past excercises we know that

$$\sigma^* = \frac{\|\nabla f(x^0)\|^2}{\nabla f(x^0)^T C \nabla f(x^0)} = \frac{1}{\lambda}$$

now we can write

$$\begin{aligned} x^1 &= x^0 - \frac{1}{\lambda} \nabla f(x) \\ &= x^0 - \frac{1}{\lambda} C^{-1} C \nabla f(x) \\ &= x^0 - C^{-1} \nabla f(x) \\ &= x^0 - C^{-1}(Cx^0 + c) \\ &= x^0 - x^0 - C^{-1}c \\ &= C^{-1}c. \end{aligned}$$

This result means that when using diagonal scaling it's probably best to chose initial points which are the eigenvectors of the scaling matrix. ■