Homework 5

Toffalini Leonardo

Exercise 1

Suppose that we want to find the root to $x^3 - 7x^2 + 8x - 3 = 0$. Is it possible to use $x_1 = 4$ as the initial point? What can you conclude about using Newton's method to approximate roots from this example?

Solution: Let's calculate the derivative of f:

$$f'(x) = 3x^2 - 14x + 8.$$

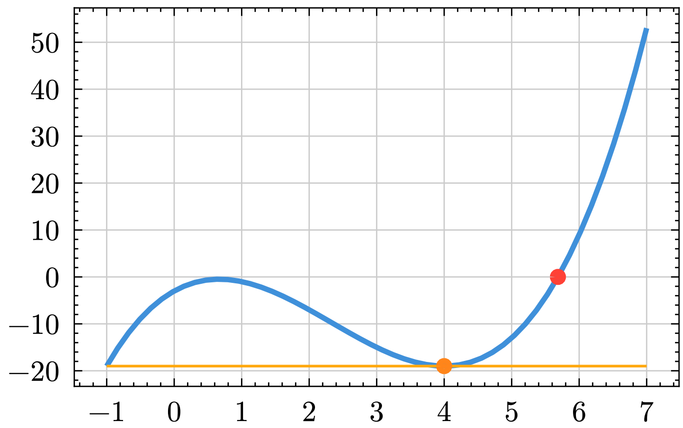
Substituting $x = 4$, we get that

$$f'(4) = 3 \cdot 4^2 - 14 \cdot 4 + 8 = 48 - 56 + 8 = 0.$$

So the derivative of f at 4 is zero, so we cannot get the next point in the Newton iteration.

This can be see easily from the below diagram, where we see $x_1 = 4$ as an orange dot and the derivative of f at 4 as an orange line, while the two roots highlighted with a red dot. Notice, that the tangent line is parallel to the x axis, so it will never intersect it, giving us no next point in the iteration.

Any other starting point $x_1 > 4$ would do just fine for finding the root of the function.



□

Exercise 2

1. Let $f(x_1, x_2) = x_1^2 + Kx_2^2$ be a quadratic function where $K > 0$. Define
$$\|(u_1, u_2)\|_{\bigcirc} := \sqrt{u_1^2 + Ku_2^2}.$$
Determine the optimum point of $\max_{\|u\|_{\bigcirc}=1} -\nabla f(x) \cdot u$.
2. Let $f(x) = x^T Ax + b^T x$ (where A is positive definite). Define $\|u\|_A := \sqrt{u^T Au}$. Determine the optimum point of $\max_{\|u\|_A=1} -\nabla f(x) \cdot u$.

Solution: **1.**
Istead of solving the constrained problem

$$\max_{\|u\|_{\bigcirc}=1} -\nabla f(x) \cdot u,$$

we will move the constraint into the objective with a lagrange multiplier, like so

$$\max_u -\nabla f(x) \cdot u + \mu(\|u\|_{\bigcirc}^2 - 1).$$

Here I also replaced $\|u\|_{\bigcirc} = 1$ with $\|u\|_{\bigcirc}^2 = 1$, as they are equivalent.

Writing out the objective function, we get

$$\begin{aligned} g(u) &= -\nabla(x_1^2 + Kx_2^2) \cdot (u_1, u_2) + \mu(u_1^2 + Ku_2^2 - 1) \\ &= -(2x_1, 2Kx_2) \cdot (u_1, u_2) + \mu(u_1^2 + Ku_2^2 - 1) \\ &= -(2x_1u_1 + 2Kx_2u_2) + \mu(u_1^2 + Ku_2^2 - 1). \end{aligned}$$

This is a nice quadratic function in u, so we can just find a root to the gradient to optimize the objective.

Thus, we just need to solve the following:

$$\nabla g(u) = \begin{bmatrix} -2x_1 + 2\mu u_1 \\ -2Kx_2 + 2\mu K u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This givs us the solutions:

$$\begin{aligned} u_1 &= \frac{x_1}{\mu} \\ u_2 &= \frac{x_2}{\mu}, \end{aligned}$$

More concisely, $u = x/\mu$. However, we still need to account for $\|u\|_{\bigcirc}^2 = 1$, so $\mu = \|x\|_{\bigcirc}$, giving us the final solution:

$$u = \frac{x}{\|x\|_{\bigcirc}}.$$

□

Solution: **2.**
Once again, we will move the constraint into the objective with a Lagrange-multiplier.

Instead of

$$\max_{\|u\|_A=1} -\nabla f(x) \cdot u,$$

we write

$$\max_u -\nabla f(x) \cdot u + \mu\big(\|u\|_A^2 - 1\big) = \max_u g(u).$$

Since A is positive definite, the above objective is convex, thus we need only find the root of the gradient.

$$\begin{aligned}
g(u) &= -\nabla f(x) \cdot u + \mu(u^T Au - 1) \\
\nabla g(u) &= -\nabla f(x) + \mu Au = 0 \\
\nabla f(x) &= \mu Au \\
u &= \frac{A^{-1}\nabla f(x)}{\mu}.
\end{aligned}$$

Since $\|u\|_A^2 = 1$, we have

$$\begin{aligned}
\mu &= \left\|A^{-1}\nabla f(x)\right\|_A \\
u &= \frac{A^{-1}\nabla f(x)}{\left\|A^{-1}\nabla f(x)\right\|_A}.
\end{aligned}$$

We can write out explicity the gradient of f as

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

and with this the explicit form of u is the following:

$$\begin{aligned}
u &= \frac{A^{-1}(Ax+b)}{\left\|A^{-1}(Ax+b)\right\|_A} \\
u &= \frac{x + A^{-1}b}{\left\|x + A^{-1}b\right\|_A}.
\end{aligned}$$

□

Exercise 3

Determine the square root of 2, accurate to four decimal places. using Newton's method.

Solution: We want to find a root to $f(x) = x^2 - 2$, which will give us the value of $\sqrt{2}$. To solve this we will, of course, use Newton's method:

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}$$
$$x_{t+1} = x_t - \frac{x_t^2 - 2}{2x_t}.$$

Let us call r the root of f, i.e. $r^2 - 2 = 0 \implies r = \sqrt{2}$.

We can verify, that

$$\begin{aligned} r^2 - 2 &= x_0^2 - 2 + 2(r - x_0)x_0 + (r - x_0)^2 \\ &= x_0^2 - 2 + 2x_0 r - 2x_0^2 + r^2 - 2x_0 r + x_0^2. \end{aligned}$$

From the update step, we have

$$x_1 = x_0 - \frac{x_0^2 - 2}{2x_0} \implies x_0^2 - 2 = 2x_0(x_0 - x_1)$$

$$\begin{aligned} 0 &= 2x_0(x_0 - x_1) + 2(r - x_0)x_0 + (r - x_0)^2 \\ 0 &= 2x_0(x_0 - x_1 + r - x_0) + (r - x_0)^2 \\ 0 &= 2x_0(r - x_1) + (r - x_0)^2 \\ x_1 - r &= \frac{(r - x_0)^2}{2x_0} \\ e_1 &= \frac{e_0^2}{2x_0} \end{aligned}$$

Which implies that the error term converges quadratically. If we choose $x_0 = 1.5$, we know that $e_0 \leq 0.5$, since $\sqrt{2} \in (1, 1.5)$, because $1.5^2 = 2.25 > 2$. If we take k steps with Newton's method, the error will be

$$e_k \leq e_0^{2^k} = 0.5^{2^k}.$$

We want that $e_k \leq 0.0001$ so $0.5^{2^k} \leq 0.0001$ which means

$$\begin{aligned} \left(\frac{1}{2}\right)^{2^k} &\leq x \\ 2^{-2^k} &\leq x \\ -2^k &\leq \log x \\ 2^k &\geq \log x \\ k &\geq \log\log x \\ k &\geq \log\log 10^{-4} \end{aligned}$$

$\square$

Exercise 4

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function. Verify that $\|\cdot\|_{H(x)^{-1}}$ is dual to the norm $\|\cdot\|_x$, where H denotes the Hessian of f.

Solution: Assuming, that $\|y\|_x := x^T H(x)x$, where $H(x)$ is the Hessian of f at x. Then, what we want to solve is the following:

$$\|z\|_{A^{-1}} = \sup_{\|y\|_A \le 1} z \cdot y.$$

Where $A = H(x)$ for some fixed x. We will prove a more general proposition, that for any A symmetric positive definite matrix the norm $\|\cdot\|_{A^{-1}}$ is dual to the norm $\|\cdot\|_A$. Since, for a convex twice differentiable function, it's Hessian is symmetric positive definite, this will in turn solve the proposed problem.

Substituting for the definitions, we have

$$\sqrt{z^T A^{-1} z} \stackrel{?}{=} \sup_{y^T A y \le 1} z \cdot y.$$

We can move the constraint into the objective function by introducing a Lagrange-multiplier to represent $y^T Ay - 1 \le 0$ and instead of finding the supremum, we will find the infimum of the negated objective to align with the lecture notes:

$$\begin{aligned}
L(z,\lambda) &= -z \cdot y + \lambda(y^T Ay - 1)\\
L'(z,\lambda) &= -z + 2\lambda Ay = 0\\
\implies y &= \frac{1}{2\lambda} A^{-1} z
\end{aligned}$$

And we want to set λ such that $y^T Ay = 1$, so after some derivation we solve for λ:

$$\begin{aligned}
\left(\frac{1}{2\lambda} A^{-1} z\right)^T A \left(\frac{1}{2\lambda} A^{-1} z\right) &= 1\\
\frac{1}{2\lambda} z^T A^{-1} A \left(\frac{1}{2\lambda} A^{-1} z\right) &= 1\\
\frac{1}{2\lambda} z^T \left(\frac{1}{2\lambda} A^{-1} z\right) &= 1\\
z^T A^{-1} z &= 4\lambda^2\\
\lambda &= \frac{1}{2}\|z\|_{A^{-1}}.
\end{aligned}$$

So we have

$$\begin{aligned}
y &= \frac{A^{-1} z}{\|z\|_{A^{-1}}}\\
z \cdot y &= \frac{z^T A^{-1} z}{\|z\|_{A^{-1}}} = \frac{\|z\|^2_{A^{-1}}}{\|z\|_{A^{-1}}} = \|z\|_{A^{-1}},
\end{aligned}$$

which is exactly what we were hoping to show.

$\square$

Exercise 5

Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a polytope. For $x \in P$, let $H(x)$ denote the Hessian of the logarithmic barrier function $F(x) = -\sum_{i=1}^{m} \log(b_i - a_i \cdot x)$ defined on the interior of P, denoted by $\mathrm{int}(P)$. The analytic center of P is the unique minimizer of $F(x)$, and is denoted by x_0^*. We define the Dinkin ellipsoid at a point $x \in P$ as

$$E_x = \{y \in \mathbb{R}^n : (y - x)^T H(x)(y - x) \leq 1\}.$$

1. Prove thet for all $x \in \mathrm{int}(P)$, $E_x \subseteq P$.
2. Assume without loss of generality (by shifting the polytope) that $x_0^* = 0$.
   Prove that $P \subseteq m E_{x_0^*}$.
3. Prove that if the set of constraints is symmetric, i.e., for every constraint of the form $a' \cdot x \leq b'$ there is a corresponding constraint $a' \cdot x \geq -b'$, then $x_0^* = 0$ and $P \subseteq \sqrt{m} E_{x_0^*}$.

Solution: Solution □