Joanna Grace S. Saladino
Math 288: Numerical Optimization
Problem Set 1

---

1. Prove that f has a global minimum.

   (a) $f(x) = \begin{cases} (|x_1| + |x_2|)\cos\left(\frac{1}{x_1^2+x_2^2}\right), & (x_1, x_2) \neq (0,0) \\ -1 & (x_1, x_2) = (0,0) \end{cases}$

   (b) $f(x) = \frac{e^{||x||}-2}{1+e^{-||x||}}$, where $x \in \mathbb{R}^n$

   Note: If a continuous function (or lower semi continuous function) is coercive then it has a global minimizer.

   Solution of (a): Clearly, f is continuous so we only need to show that f is *coercive*. Let $\{x_k\}_{k=1}^{\infty} \in \mathbb{R}^2$ such that $||x_k|| \to +\infty$.
   First, observe that

$$
\begin{aligned}
\lim_{k\to\infty} \sqrt{x_{k1}^2 + x_{k2}^2}\, \cos\left(\frac{1}{x_{k1}^2 + x_{k2}^2}\right) &= \lim_{k\to\infty} ||x_k|| \cos\left(\frac{1}{||x_k||^2}\right) \\
&= \lim_{k\to\infty} ||x_k|| \lim_{k\to\infty} \cos\left(\frac{1}{||x_k||^2}\right) \\
&= \lim_{k\to\infty} ||x_k|| \\
&= +\infty
\end{aligned}
$$

   But since $\sqrt{x^2+y^2} \leq |x^2| + |y^2|$ for all $x, y \in \mathbb{R}$ then

$$
\lim_{k\to\infty} \sqrt{x_{k1}^2 + x_{k2}^2}\, \cos\left(\frac{1}{x_{k1}^2 + x_{k2}^2}\right) \leq \lim_{k\to\infty} (|x_{k1}| + |x_{k2}|)\cos\left(\frac{1}{x_{k1}^2 + x_{k2}^2}\right)
$$

   By transitivity, $\lim_{k\to\infty} f(x_k) = +\infty$, hence, f is coercive.
   Therfore, $f(x)$ has a global minimum.

   Solution of (1b): Clearly, f is continuous thus to prove the existence of a global minimum of (1b), it suffices to show that f is coercive.
   Let $\{x_k\}_{k=1}^{\infty} \in \mathbb{R}^n$ such that $||x_k|| \to +\infty$ then

$$
\begin{aligned}
\lim_{k\to\infty} f(\mathbf{x}_k) &= \lim_{k\to\infty} \frac{e^{||x_k||} - 2}{1 + e^{-||x_k||}} \\
&= \lim_{k\to\infty} \frac{e^{2||x_k||} - 2e^{||x_k||}}{e^{||x_k||} + 1} \\
&= \lim_{k\to\infty} \frac{2e^{2||x_k||} - 2e^{||x_k||}}{e^{||x_k||}} \\
&= \lim_{k\to\infty} 2e^{||x_k||} - 2 \\
&= +\infty
\end{aligned}
$$

   Therefore, $f(x)$ has a global minimum.

2. Consider

$$
f(x) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2
$$

   This function has three local minima. The global minimum is 0 at $(0, 0)$.

   (a) Use the method of steepest descent on f using three different initial points: $(1.5, 1.5), (0, 1.5)$, and $(-1, 1.5)$. Use the Golden Search Algorithm to estimate the step length. For the GSA, fix the initial interval to $[0, 1]$. Set the tolerance to 10^{-7} for the stopping criteria.

   (b) Plot the contour of f on $[-2, 2]^2$. Use 40 level sets. In the same figure, plot the path of the 3 sequences generated using the three different initial points in (a).

3. The function $y(t)$ satisfies the differential equation

$$\frac{dy}{dt} = 2t - 1 - \int_0^{\tan^{-1} t} e^{-x^2}\,dx$$

   (a) Prove that $y(t)$ is convex for all $t \in \mathbb{R}$.
   (b) Use the Newton method to find the minimizer of $y(t)$. Use the initial guess $t_0 = 0$ and set the tolerance to 10^{-7} for the stopping criterion. Set the step length to 1.

Proof: Suppose function $y(t)$ satisfies the differential equation

$$\begin{aligned} \frac{dy}{dt} &= 2t - 1 - \int_0^{\tan^{-1} t} e^{-x^2}\,dx \\ \frac{d^2y}{dt^2} &= 2 - e^{-\tan^{-1} t^2} \end{aligned}$$

By the second order convexity condition, a twice differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if $f''(x) \geq 0$ for all $x \in dom(f)$.

Clearly, $y : \mathbb{R} \to \mathbb{R}$ is twice differentiable. So to prove that y is convex, it suffices to show that $y''(t) > 0$ for all $t \in \mathbb{R}$. Note that the range of $\tan^{-1} t^2$ is $[0, \frac{\pi}{2}]$ so

$$\begin{aligned} & e^{-\tan^{-1} t^2} \in [e^{-\frac{\pi}{2}}, e^0] \subseteq [0.20, 1] \\ \implies & e^{-\tan^{-1} t^2} \in [0.20, 1] \\ \implies & 2 - e^{-\tan^{-1} t^2} \in [1, 1.8] \\ \implies & y''(t) \in [1, 1.8] \end{aligned}$$

Therefore, $y''(t) > 0$ for all $t \in \mathbb{R}$. Thus, $y(t)$ is convex.

4. Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable. Suppose ∇f is Lipschitz continuous with constant L. Show that

$$f(x+y) \leq f(x) + \nabla f(x)^T y + \frac{L}{2}||y||^2, \quad \forall x, y \in \mathbb{R}^n$$

Proof: Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and suppose ∇f is Lipschitz continuous with constant L, that is,

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \quad \forall x, y \in \mathbb{R}^n$$

Consider $g(t) = f(x + ty)$. With this, we get $g'(t) = \nabla f(x + ty)^T y$.

$$\begin{aligned} ||g'(t) - g'(0)|| &= ||\nabla f(x+ty)^T y - \nabla f(x)^T y|| \\ &= ||(\nabla f(x+ty) - \nabla f(x))^T y|| \\ &\leq ||(\nabla f(x+ty) - \nabla f(x))||\ ||y|| \\ &\leq L\ ||x + ty - x||\ ||y|| \\ &= L\ ||ty||\ ||y|| \\ &= t\ L\ ||y||^2 \end{aligned}$$

Clearly, $f(x+y) = g(1) = \int_0^1 g'(t)dt + g(0)$. So,

$$\begin{aligned} f(x+y) &= \int_0^1 g'(t)dt + g(0) \\ &= \int_0^1 g'(t)\ dt - g'(0) + g'(0) + g(0) \\ &= \int_0^1 (g'(t) - g'(0))\ dt + g'(0) + g(0) \\ &\leq \int_0^1 t\ L\ ||y||^2\ dt + g'(0) + g(0) \\ &= \frac{L}{2}\ ||y||^2 + g'(0) + g(0) \\ &= \frac{L}{2}\ ||y||^2 + \nabla f(x)^T y + f(x) \end{aligned}$$

Rearranging the terms, we have

$$f(x+y) \leq f(x) + \nabla f(x)^T y + \frac{L}{2}\ ||y||^2$$

for all $x, y \in \mathbb{R}^n$

5. Consider

$$u^* = \arg\min_{u \in \mathcal{A}} \frac{1}{2} \int_0^1 |u'(x)|^2 dx,$$

where $\mathcal{A} = \{u \in C^0([0,1]) : u(0) = 1, u(1) = 2\}$. Approximate u^* by a piecewise linear spline interpolation. Let u_i^* be the interpolated value of u^* at equidistant point x_i, where $x_i = \frac{i}{n}, i = 0, 1, ..., n$ that is, $u_i^* = u^*(x_i)$. Use the interpolation to convert the above minimization problem into a minimization problem in $\mathbb{R}^{n-1}$. Set the initial guess to $\mathbf{0} \in \mathbb{R}^{n-1}$. Use a diminishing step size with the initial step length set to 10. Construct the step size iterates based on the Harmonic series. Set the maximum number of iterations to 5000 and the tolerance to 10^{-7} for the stopping criterion.

6. Suppose $\{x_K\}_{k\geq 0}$ is generated using $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction and α_k satisfies the Wolfe condition. Suppose ∇f is Lipschitz with constant L.
   (a) Prove that

   $$\sum_{k=0}^{\infty} ||\nabla f(x_k)||^2 \cos^2 \theta_k,$$

   where θ_k is the angle between d_k and $-\nabla f(x_k)$ is convergent.
   (b) Use (a) to prove that the method of steepest descent using Wolfe's rule converges to a stationary point.
   (c) Suppose $d_k = -D^{-1}\nabla f(x_k)$, where D is a positive definite matrix. Prove that $\{x_k\}_{k\geq 0}$ converges to a stationary point.

Proof of (a): Suppose $\{x_k\}_{k\geq 0}$ is generated using $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction and α_k satisfies the Wolfe conditions

$$\begin{aligned} &\text{First condition:} \quad f(x_k + \alpha_k d_k) - f(x_k) \leq \sigma\ a_k \nabla f(x_k)^T d_k \\ &\text{Second condition:} \quad \nabla f(x_k + \alpha_k d_k)^T d_k \geq \beta\ \nabla f(x_k)^T d_k \end{aligned}$$

where $0 < \sigma < \beta < 1$.

Moreover, suppose that ∇f is Lipschitz with constant L, that is,

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \quad \forall x, y \in \mathbb{R}^n$$

Here, it is important to note that since θ_k is the angle between d_k and $-\nabla f(x_k)$ then

$$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{||\nabla f(x_k)||\ ||d_k||}$$

To prove the convergence of $\sum_{k=0}^{\infty} ||\nabla f(x_k)||^2 \cos^2\theta_k$, we simply need to show that it is finite, that is,

$$\sum_{k=0}^{\infty} ||\nabla f(x_k)||^2 \cos^2\theta_k < \infty$$

From Wolfe's second condition,

$$\nabla f(x_k + \alpha_k d_k)^T d_k - \nabla f(x_k)^T d_k \geq \beta\ \nabla f(x_k)^T d_k - \nabla f(x_k)^T d_k$$

which could be re-written as

$$(\nabla f(x_{k+1})^T - \nabla f(x_k))^T d_k \geq (\beta - 1)\ \nabla f(x_k)^T d_k \tag{1}$$

and from the Lipschitz condition,

$$\begin{aligned}(\nabla f(x_{k+1})^T - \nabla f(x_k))^T d_k &\leq L\ ||x_{k+1} - x_k||\ ||d_k|| \\ &\leq L\ ||x_k + \alpha_k d_k - x_k||\ ||d_k|| \\ &= L\ ||\alpha_k d_k||\ ||d_k|| \\ &= L\ ||\alpha_k d_k||\ ||d_k|| \\ &= \alpha_k\ L||d_k||^2\end{aligned}$$

Thus,

$$(\nabla f(x_{k+1})^T - \nabla f(x_k))^T d_k \leq \alpha_k\ L||d_k||^2 \tag{2}$$

Combining 1 and 2 we get the following relations,

$$\alpha_k\ L||d_k||^2 \geq (\nabla f(x_{k+1})^T - \nabla f(x_k))^T d_k \geq (\beta - 1)\ \nabla f(x_k)^T d_k$$

Thus, we get the following inequality,

$$\begin{aligned}\alpha_k\ L||d_k||^2 &\geq (\beta - 1)\ \nabla f(x_k)^T d_k \\ \implies \alpha_k &\geq \frac{(\beta - 1)}{L} \frac{\nabla f(x_k)^T d_k}{||d_k||^2} \\ \implies -\alpha_k &\leq \frac{(1 - \beta)}{L} \frac{\nabla f(x_k)^T d_k}{||d_k||^2}\end{aligned}$$

Substituting this inequality into the Wolfe's first condition,

$$\begin{aligned}f(x_k + \alpha_k d_k) &\leq f(x_k) + \sigma\ a_k \nabla f(x_k)^T d_k \\ &= f(x_k) - \sigma(-a_k)\nabla f(x_k)^T d_k \\ &\leq f(x_k) - \sigma\frac{(1 - \beta)}{L} \frac{\nabla f(x_k)^T d_k}{||d_k||^2} \nabla f(x_k)^T d_k \\ &= f(x_k) - \sigma\frac{(1 - \beta)}{L} \Big(\frac{\nabla f(x_k)^T d_k}{||d_k||}\Big)^2\end{aligned}$$

Re-writing the last inequality,

$$f(x_{k+1}) \leq f(x_k) - \sigma\frac{(1 - \beta)}{L} \cos^2\theta_k\ ||\nabla f(x_k)||^2$$

By summing this expression over all indices less than or equal to k, we obtain

$$\begin{aligned}f(x_{k+1}) &\leq f(x_0) - \lambda\sum_{j=0}^{k} \cos^2\theta_j ||\nabla f(x_j)||^2, \quad \lambda = \sigma\frac{(1 - \beta)}{L} \\ \implies f(x_0) - f(x_{k+1}) &\geq \lambda\sum_{j=0}^{k} \cos^2\theta_j ||\nabla f(x_j)||^2, \quad \lambda = \sigma\frac{(1 - \beta)}{L}\end{aligned}$$

Note: We know from previous theorem that it is necessary that f is bounded below along $\{x_k + \alpha d_k | \alpha \geq 0\}$ so that an interval of step lengths satisfying the two Wolfe's condition exists. So, f must be bounded below.

Since f is bounded below, there exists an $M < \infty$ such that $f(x_0) - f(x_{k+1}) < M$ for all k. Taking the limit,

$$\lim_{k\to\infty} \lambda \sum_{j=0}^{k} \cos^2\theta_j ||\nabla f(x_j)||^2 \le \lim_{k\to\infty} f(x_0) - f(x_{k+1}) \le M$$

Therefore,

$$\sum_{k=0}^{\infty} \cos^2\theta_k ||\nabla f(x_k)||^2 < \infty \tag{3}$$

Proof of (b): Using 3 and the standard facts about convergence of series, we have

$$\cos^2\theta_k ||\nabla f(x_k)|| \to 0, \quad \text{as } k \to \infty \tag{4}$$

For the steepest descent method, $dk = -\nabla f(x_k)^T$

$$\cos\theta_k = \frac{-\nabla f(x_k)^T d_k}{||\nabla f(x_k)||\ ||d_k||} = 1$$

It follows then that $\cos^2\theta_k ||\nabla f(x_k)|| = ||\nabla f(x_k)||$ and

$$||\nabla f(x_k)|| \to 0, \quad \text{as } k \to \infty$$

Therefore, the method of steepest descent using Wolfe's rule converges to a stationary point.

Proof of (c): Suppose $d_k = -D^{-1}\nabla f(x_k)$, where D is a positive definite matrix.

Note that if an $n \times n$ matrix A is a positive definite matrix, then

$$||A||^2 = \lambda_1(A)$$

where λ_1 is the largest eigenvalue of A. Here, $||A|| = \max_{x\neq 0} \frac{||Ax||}{||x||}$. *Consequently,*

$$||A^{-1}||^2 = \lambda_n(A)$$

With this, we then get

$$||D||\ ||D^{-1}|| \le \sqrt{\lambda_1(D)}\ (\sqrt{\lambda_n(D)})^{-1} = M$$

where $\lambda_1(D)$ and $\lambda_n(D)$ is the largest and smallest eigenvalue of D. Moreover, since D is a positive definite matrix then all the eigenvalues of D is positive and finite, thus, M must be positive and finite also. Then

$$\begin{aligned}
\cos\theta_k &= \frac{-\nabla f(x_k)^T d_k}{||\nabla f(x_k)||\ ||d_k||} \\
&= \frac{\nabla f(x_k)^T D^{-1} \nabla f(x_k)}{||\nabla f(x_k)||\ ||D^{-1}\nabla f(x_k)||} \\
&\ge \frac{1}{||\nabla f(x_k)||} \frac{||\nabla f(x_k)||^2}{||D||} \frac{1}{||D^{-1}||\ ||\nabla f(x_k)||} \\
&= \frac{1}{||D||\ ||D^{-1}||} \\
&\ge \frac{1}{M}
\end{aligned}$$

Clearly, $\cos\theta_k$ is bounded below by $\frac{1}{M}$ which is always greater than 0. So by 4, we are forced to conclude that $||\nabla f(x_k)|| \to 0$. Therefore, $\{x_k\}_{k\ge 0}$ converges to a stationary point.

7. Attached is the data set comparing the chirps/sec for the striped ground cricket (x) and the temperature in Fahrenheit (y). The linear regression $y_i \approx c_1 x_i + c_2$ on the data set $\{x_i, y_i\}_{i=1}^{1}5$ can be computed by minimizing

$$f(c_1, c_2) = \frac{1}{2} \sum_{i=1}^{15} (y_i - c_1 x_i - c_2)^2.$$

(a) Use the Conjugate Gradient method to obtain the minimizer (c_1^*, c_2^*) of f. Set the maximum iteration to 100 and the tolerance for the stopping criterion to 10^{-7}. For the step length, use the quadratic interpolation with the interval length set to $[0, 1]$. Use $(0, 0)$ as the initial guess

(b) Obtain a scatter plot of the data set $\{x_i, y_i\})i = 1^{15}$ and compare it with the plot $y = c_1^* x_i + c_2^*$, for $x \in [14, 20]$

(a) Use the regression to estimate the temperature if the striped ground cricket chirped 19 times in one second

8. Convert the system of equations

$$\begin{cases} x - xy = 1.5 \\ x - xy^2 = 2.25 \\ x - xy^3 = 2.625 \end{cases}$$

into a minimization problem in $\mathbb{R}^2$. Use the Quasi-Newton method to solve the equivalent minimization problem. Set the maximum iteration to 100 and the tolerance for the stopping criterion to 10^{-7}. For the step length, use the Armijo rule with the parameters: $\beta = 0.5,\ s = 1,$ and $\sigma = 0.2$. Use $(0, 0)$ as the initial guess.

9. Let $\mathcal{P} = \text{span}\{1, x, x^2, ..., x^1 0\}$. Define

$$f(x) = \begin{cases} -1, & -2 \le x < -1 \\ x, & -1 \le x < 1 \\ 1, & 1 \le x < 2 \end{cases}$$

Let $v^* \in \mathcal{P}$ such that

$$v^* = \arg\min_{v \in \mathcal{P}} \int_{-1}^{1} (f(x) - v(x))^2 \, dx.$$

One can interpret v^* as a polynomial interpolation of f. Use a Quasi-Newton built-in solver (fminunc for Matlab and Octave, minimize(method='BFGS') for Python) to determine v^*. Plot $f(x)$ and $v^*(x)$ for all $x \in [-2, 2]$ in one figure.