Introduction to Optimization Homework (3)

Dimitri Tabatadze · Monday 11-03-2024

Problem 3.1: Directional Derivatives

(a)
- Partial: $\nabla f(x_1, x_2) = (4x_1 + 3x_2, 3x_1 + 1)$
- Directional: $\nabla f(0,0)^T d = \binom{0}{1} \cdot \binom{1}{-1} = -1$

(b)
- Partial: $\nabla f(x_1, x_2) = (\sin(x_2), x_1 \cdot \cos(x_2))$
- Directional: $\nabla f(0,0)^T d = \binom{0}{0} \cdot \binom{1}{-1} = 0$

(c)
- Partial: $\nabla f(x_1, x_2) = \begin{cases} \left(\frac{2x_1^5 x_2 + 2x_1 x_2^3 - 4x_1^5 x_2}{\left(x_1^4 + x_2^2\right)^2}, \frac{x_1^2\left(x_1^4 + x_2^2\right) - 2x_1^2 x_2^3}{\left(x_1^4 + x_2^2\right)^2}\right) & (x_1, x_2) \neq (0,0) \\ (0,0) & (x_1, x_2) = (0,0) \end{cases}$
- Directional: $\nabla f(0,0)^T d = \binom{0}{0} \cdot \binom{1}{-1} = 0$

Problem 3.2: Local Optima Along Lines

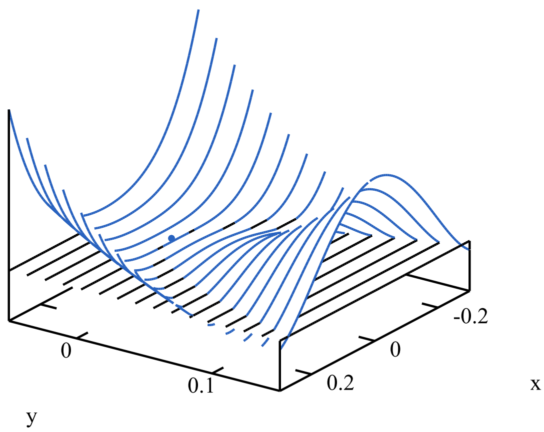


Figure 1: $f(x, y) = 2x^4 - 3x^2 y + y^2$

(a)

$$\begin{aligned} \nabla f(x, y) &= (8x^3 - 6xy, 2y - 2x^2) = 0 \\ &\Longrightarrow \begin{cases} 8x^3 - 6xy = 0 \\ 2y - 2x^2 = 0 \end{cases} \\ &\Longrightarrow \begin{cases} y = x^3 \\ 8x^3 - 6x^3 = 0 \end{cases} \\ &\Longrightarrow \begin{cases} x = 0 \\ y = 0 \end{cases} \end{aligned}$$

(b) Let x and y be constants with $d = \binom{x}{y}$ and consider

$$\begin{aligned} \varphi_d(\sigma) &= f(\sigma x, \sigma y) \\ &= 2\sigma^4 x^4 - 3\sigma^3 x^2 y^2 + \sigma^2 y^2 \\ &= \underbrace{\sigma^2}_{\geq 0}\left(\sigma^2 \cdot 2x^4 - \sigma \cdot 3x^2 y^2 + y^2\right). \end{aligned}$$

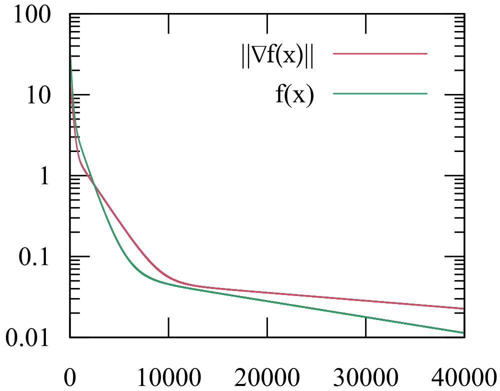
Now we just need to show that $\sigma^2 \cdot 2x^4 - \sigma \cdot 3x^2 y^2 + y^2 \geq 0$. Since

$$\begin{aligned}(3x^2y^2)^2 - 4\cdot 2x^4\cdot y^2 &= 9x^4y^4 - 4\cdot 2x^4\cdot y^2\\ &= \underbrace{x^4y^2}_{\geq 0}(9y^2-4) < 0\\ \implies 9y^2 - 4 < 0 &\implies -\frac{2}{3} < y < \frac{2}{3}\end{aligned}$$

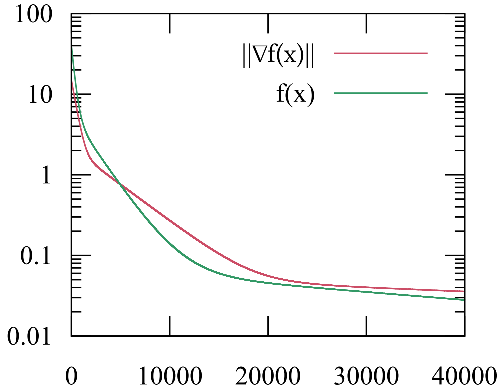
for all y near 0, $\varphi_d(\sigma) > 0$ independent of x.

(c) Since I just showed that f is bounded from below in every direction near $\binom{0}{0}$, $\binom{0}{0}$ is a local minimum.

Problem 3.3: Comparing the Efficiency of Step Size Rules

(a) 

With $\gamma = 0.5, \beta = 1, S_0 = 1$, backtracking takes a lot of iterations.

(b) 

With $\gamma = 0.1, \beta = 0.5, S_0 = 1$, it takes even more.

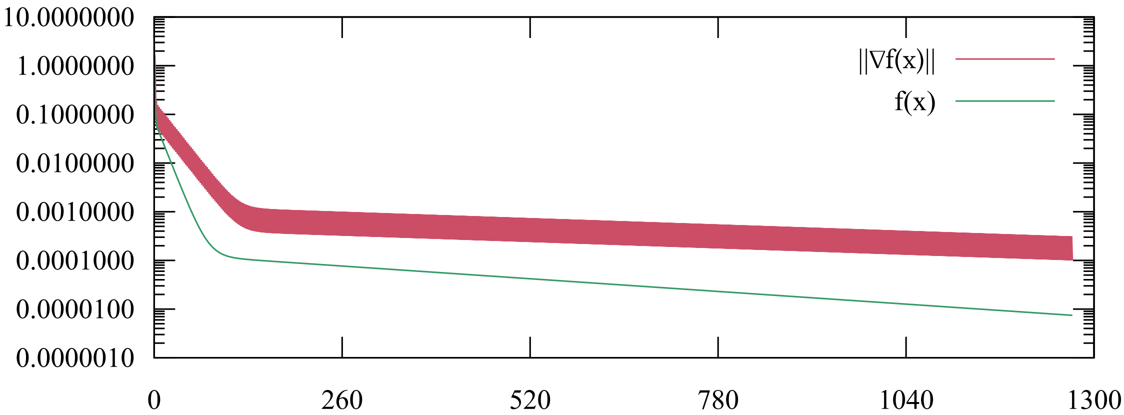
(c) To the function along $x + \sigma s$ so we do the following.

$$\begin{aligned}\varphi_{x,k}(\sigma) &= (x+\sigma s)^T A(x+\sigma s)\\ &= (x^T + \sigma s^T)A(x+\sigma s)\\ &= x^T Ax + 2\sigma s^T Ax + \sigma^2 s^T As.\end{aligned}$$

which is a simple quadratic equation with respect to σ. The extremum of the parabola can be found by

$$\sigma = \frac{-2s^T Ax}{2s^T As}.$$

This can be used for the exact line search method, which converges magnitudes faster than the backtracking method.



Problem 3.4: Quadratic Problems

$$f(x) = \frac{1}{2}x^T Hx + b^T x + c$$

(a) • $\nabla f(x) = \frac{1}{2}x^T(H + H^T) + b^T$

• $H_f(x) = \frac{1}{2}(H + H^T)$

(b) Since $x^T Hx$ is a scalar, $x^T Hx = (x^T Hx)^T = x^T H^T x$ and so the function f can be rewritten as $f(x) = \frac{1}{2}x^T H^T x + b^T x + c$.

• $\nabla f(x) = x^T H + b^T$

• $H_f(x) = H$

(c) Since in the case of H being symmetric $\nabla f(x) = x^T H + b^T$ and $H_f(x) = H$, if H also turns out to be positive definite, the sufficient conditions of $\bar{x} = H^{-1}b$ being the only local and also a global minimizer hold, thereby f is convex.

(d) • $H = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix}, b = \begin{pmatrix} -4 \\ -3 \end{pmatrix}, c = 3, f(x_1, x_2) = g(x_1, x_2)$

• H is positive definite.

• There will exist a unique minimizer.

• $\bar{x} = H^{-1}b = \frac{1}{9}\begin{pmatrix} -5 & 4 \\ 4 & -5 \end{pmatrix}\begin{pmatrix} -4 \\ -3 \end{pmatrix} = \begin{pmatrix} 8/9 \\ -1/9 \end{pmatrix}$

Problem 3.5: Source Localization Problem

First, rewrite f

$$\begin{aligned}
f(x) &= \sum_{i=1}^{m} \left(\|x - a_i\|_2 - d_i\right)^2 \\
&= \sum_{i=1}^{m} \left(\|x - a_i\|_2^2 - 2d_i\|x - a_i\|_2 + d_i^2\right) \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} (x_j - a_{ij})^2 - 2d_i\sqrt{\sum_{j=1}^{n} (x_j - a_{ij})^2} + d_i^2\right) \\
&= \sum_{i=1}^{m} \left(\sum_{j=1}^{n} (x_j^2 - 2x_j a_{ij} + a_{ij}^2) - 2d_i\sqrt{\sum_{j=1}^{n} (x_j - a_{ij})^2} + d_i^2\right).
\end{aligned}$$

Then, take partial differential of f with respect to every x_j and

$$\begin{aligned}
\frac{\partial}{\partial x_j} f(x) &= \sum_{i=1}^{m} \left((2x_j - 2a_{ij}) - d_i\frac{2x_j - 2a_{ij}}{\sqrt{\sum_{j=1}^{n} (x_j - a_{ij})^2}}\right) \\
&= \sum_{i=1}^{m} 2\left(x_j - a_{ij} - d_i\frac{x_j - a_{ij}}{\|x - a_i\|_2}\right).
\end{aligned}$$

Since every partial derivative must be zero at a stationary point, we get

$$\frac{\partial}{\partial x_j} f(x) = 0 \implies \quad mx_j = \sum_{i=1}^{m} \left(a_{ij} + d_i\frac{x_j - a_{ij}}{\|x - a_i\|_2}\right) \implies$$

$$x_j = \frac{1}{m}\left(\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i\frac{x_j - a_i}{\|x - a_i\|_2}\right) \implies x = \frac{1}{m}\left(\sum_{i=1}^{m} a_i + \sum_{i=1}^{m} d_i\frac{x - a_i}{\|x - a_i\|_2}\right).$$

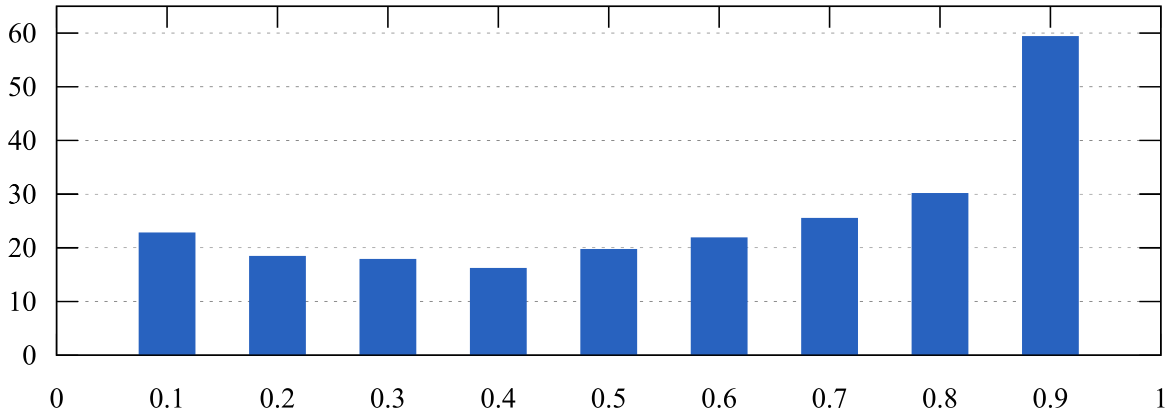
Problem 3.6: Performance of the Gradient Descent Method

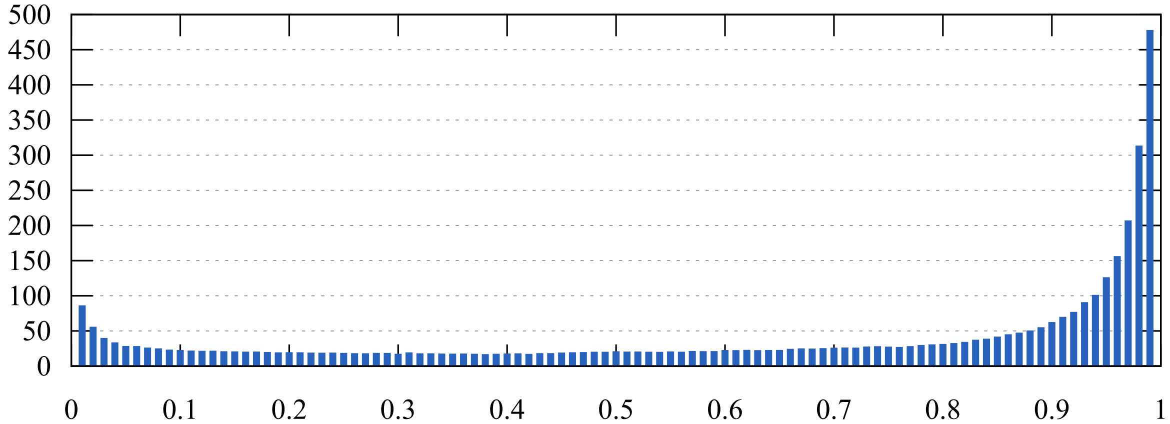
The objective function is

$$f(x_1, x_2) = e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} + e^{-x_1+0.1}$$

with the Gradient

$$\nabla f(x_1, x_2) = \begin{pmatrix} e^{x_1+3x_2-0.1} + e^{x_1-3x_2-0.1} - e^{-x_1+0.1} \\ 3e^{x_1+3x_2-0.1} - 3e^{x_1-3x_2-0.1} \end{pmatrix}$$





From the plots we see that at around 0.4 the number of iterations is the lowest.