

AMS230 – Homework #3

Zayd Hammoudeh

May 16, 2018

Name: Zayd Hammoudeh

Course Name: AMS230

Assignment Name: Homework #3

Due Date: May 16, 2018

Student Discussions: I discussed the problems with the following students. All write-ups were prepared separately and independently.

- Ben Sherman
- Bernardo Torres

Exercise #1

Consider the dog-leg path:

$$\tilde{p}(\tau) = \begin{cases} \tau p^U, & 0 \leq \tau \leq 1 \\ p^U + (\tau - 1)(p^B - p^U), & 1 \leq \tau \leq 2 \end{cases}$$

where

$$\begin{aligned} p^U &= -\frac{\|g\|^2}{g^T B g} g \\ B \cdot p^B &= -g. \end{aligned}$$

Suppose that symmetric matrix B and vector g satisfy:

1. $g^T B g > 0$
2. $(p^U)^T (p^B - p^U) > 0$

Prove that:

- i. $\|\tilde{p}(\tau)\|$ is an increasing function of τ
- ii. $m(\tilde{p}(\tau))$ is a decreasing function of τ

where $m(p) = g^T p + \frac{1}{2} p^T B p$.

Proof. This proof is to a large extent based off Lemma 4.2 from Nocedal and Wright. The proof can be divided into two cases based on the value of τ . Proving both cases separately proves the entire statement.

Case #1: $0 \leq \tau \leq 1$

Property i: $\|\tilde{p}(\tau)\| = \tau \|p^U\|$. This is clearly increasing for $\tau \in [0, 1]$ since $\|p^U\|$ is strictly positive for $p^U \neq 0$.

Property ii : This can be simplified via:

$$\begin{aligned} m(\tilde{p}(\tau)) &= \tau g^T p^U + \frac{\tau^2}{2} (p^U)^T B p^U \\ &= -\tau \frac{(\|g\|^2)^2}{g^T B g} + \frac{\tau^2}{2} \frac{(\|g\|^2)^2}{(g^T B g)^2} g^T B g \\ &= \left(-\tau + \frac{\tau^2}{2}\right) \frac{(\|g\|^2)^2}{g^T B g} \end{aligned}$$

Both $\|g\|^2$ and $g^T Bg$ are strictly positive making their ratio also strictly positive. The ratio is also fixed with respect to τ meaning it can be treated as a positive constant. For $\tau \in [0, 1]$, $-\tau + \tau^2/2$ is decreasing (as can be trivially proven by taking the derivative with respect to τ).

Case #2: $1 \leq \tau \leq 2$

Property i: If $\|\tilde{p}(\tau)\|$ is an increasing function, then so is $\tilde{p}(\tau) \cdot \tilde{p}(\tau)$. Proving the latter is an increasing function proves the former is as well.

$$\begin{aligned}\tilde{p}(\tau) \cdot \tilde{p}(\tau) &= p^U \cdot p^U + 2(\tau - 1)p^U \cdot (p^B - p^U) + (\tau - 1)^2(p^B - p^U) \cdot (p^B - p^U) \\ &= \|p^U\|^2 + 2(\tau - 1)p^U \cdot (p^B - p^U) + (\tau - 1)^2\|p^B - p^U\|^2\end{aligned}$$

Each norm above is non-negative as is $p^U \cdot (p^B - p^U)$ by supposition 2 above. Therefore, the function is increasing when $\tau \in [1, 2]$. This can again be trivially shown by taking the derivative with respect to τ and looking in the range $[1, 2]$.

Property ii: Define $\hat{h}(\alpha) = m(\tilde{p}(1 + \alpha))$. If $h'(\alpha) \leq 0$ for $\alpha \in (0, 1)$ then the property holds. Using the definition of $\tilde{p}(\tau)$ in the exercise description and the definition of trust region, we find:

$$\begin{aligned}\hat{h}'(\alpha) &= (p^B - p^U)^T(g + Bp^U) \\ &\leq (p^B - p^U)^T(g + Bp^U + B(p^B - p^U)) \\ &= (p^B - p^U)^T(g + Bp^B) = 0\end{aligned}$$

given $B \cdot p^B = -g$ as stated in the exercise description. □

Exercise #2

Code Algorithm 4.1 in Nocedal and Wrihght with:

1. Cauchy point method for the subproblem
2. Dog-leg method based on the results for Exercise 1.

Test and compare the performance of the methods on the following problem:

$$\min_{x \in \mathbb{R}^n} f(x) = \log(1 + x^T Q x)$$

where Q is a symmetric and positive definite matrix.

Q was constructed in the same way as homework #2 using QR decomposition and a diagonal matrix of eigenvalues. The gradient, g , of f is defined as:

$$g = \nabla f(x) = \frac{2Qx}{1 + x^T Q x}.$$

The Hessian can then be found via the quotient rule as shown below.

$$\begin{aligned} \nabla f^2(x) &= \frac{d}{dx} g = \frac{d}{dx} \frac{2Qx}{1 + x^T Q x} \\ &= \frac{1}{1 + x^T Q x} \frac{d}{dx} (2Qx) + 2Qx \frac{d}{dx} \left(\frac{1}{1 + x^T Q x} \right) \\ &= \frac{2Q}{1 + x^T Q x} - 2Qx \left(\frac{(2Qx)^T}{(1 + x^T Q x)^2} \right) \end{aligned}$$

For most x , the Hessian was not positive definite. To approximate the Hessian, B_k was set to Q . While this is not exactly the Hessian, it achieved good convergence. I tried an alternate approach where $B_k = Q$ *only* if the Hessian was not positive definite. However, this was generally slower to converge for both Cauchy Points and Dogleg. As such, those results are not reported in this document.

Table 1 lists the parameters used in the experiments. Note that $\mathcal{U}(a, b)$ represents a uniform random variable selected from the range $[a, b)$. Since Q is positive definite, then for all $x \neq [0]^n \implies x^T Q x > 0$. Therefore, since \log is monotonically increasing, f is minimized when $x^* = [0]^n$.

Figure 1 shows the performance of the Cauchy Point and Dogleg methods for the same x_0 . Observe that Cauchy Point took longer to converge than Dogleg. The exact number of steps to converge varied by upto 75% depending on the value of the random x_0 and Q . In addition, the difference in the converge rate of Cauchy Point and Dogleg varied with x_0 and Q .

Table 1: Parameters used in the experiments for Exercise #2

Name	Value
\log	\log_{10}
n	100
λ	$\mathcal{U}(10, 1000)$
B	Q
Δ_0	1
$\hat{\Delta}$	10,000
η	0.1
x_0	Random vector from $[0, 1)^n$
x^*	$[0]^n$

Comparison of the performance of Cauchy Points and Dogleg methods

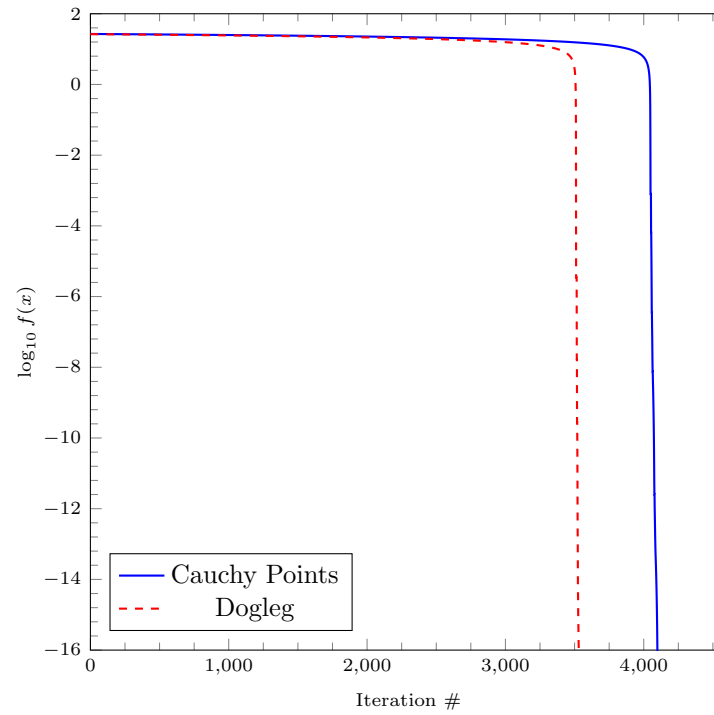


Figure 1: Comparison of the performance of Cauchy Point and Dogleg methods

Python Source Code for Exercise #2