Optimization Report: Jacobians, Hessians, and Convergence Analysis

Anaparthi Lakshmi Dedeepya

March 19, 2025

1 Introduction

This report analyzes the performance of various optimization algorithms on **five functions**: - **Trid function** - **Three-Hump Camel function** - **Styblinski-Tang function** - **Root-of-Square function** - **Rosenbrock function**

For each function, we derive **gradients and Hessians**, compute minima (except Rosenbrock), and analyze convergence behavior under different optimization algorithms.

—

2 Mathematical Derivations

2.1 Jacobians and Hessians

2.1.1 Trid Function

$$f(x) = \sum_{i=1}^{n}(x_i - 1)^2 - \sum_{i=2}^{n} x_i x_{i-1}$$

Gradient:

$$\frac{\partial f}{\partial x_i} = 2(x_i - 1) - x_{i-1} \ \ if \ i >= 2$$

$$\frac{\partial f}{\partial x_i} = 2(x_i - 1) \ \ if \ i = 1$$

Hessian:

$$H_{i,j} = \begin{cases} 2 & \text{if } i = j \\ -1 & \text{if } |i - j| = 1 \\ 0 & \text{otherwise} \end{cases}$$

2.1.2 Three-Hump Camel Function

$$f(x, y) = 2x^2 - 1.05x^4 + \frac{x^6}{6} + xy + y^2$$

Gradient:

$$\nabla f = \begin{bmatrix} 4x - 4.2x^3 + x^5 + y \\ x + 2y \end{bmatrix}$$

Hessian:

$$H = \begin{bmatrix} 4 - 12.6x^2 + 5x^4 & 1 \\ 1 & 2 \end{bmatrix}$$

2.1.3 Styblinski-Tang Function

$$f(x) = \frac{1}{2} \sum_{i=1}^{n} (x_i^4 - 16x_i^2 + 5x_i)$$

Gradient:

$$\nabla f = \frac{1}{2} \sum_{i=1}^{n} (4x_i^3 - 32x_i + 5)$$

$$\frac{\partial^2 g}{\partial x_i \partial x_j} = \begin{cases} \frac{1}{2}(12x_i^2 - 32) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{1}$$

$$\mathbf{H}(g) = \frac{1}{2} \begin{pmatrix} 12x_1^2 - 32 & 0 & \cdots & 0 \\ 0 & 12x_2^2 - 32 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 12x_n^2 - 32 \end{pmatrix} \tag{2}$$

2.1.4 Root-of-Square Function

$$f(x) = \sqrt{x_1^2 + x_2^2}$$

Gradient:

$$\nabla f = \begin{bmatrix} \frac{x_1}{\sqrt{x_1^2+x_2^2}} \\ \frac{x_2}{\sqrt{x_1^2+x_2^2}} \end{bmatrix}$$

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{1}{\sqrt{x_1^2 + x_2^2}} - \frac{x_1^2}{(x_1^2 + x_2^2)^{3/2}} = \frac{x_2^2}{(x_1^2 + x_2^2)^{3/2}} \tag{3}$$

$$\frac{\partial^2 f}{\partial x_2^2} = \frac{1}{\sqrt{x_1^2 + x_2^2}} - \frac{x_2^2}{(x_1^2 + x_2^2)^{3/2}} = \frac{x_1^2}{(x_1^2 + x_2^2)^{3/2}} \tag{4}$$

$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = -\frac{x_1 x_2}{(x_1^2 + x_2^2)^{3/2}} = \frac{\partial^2 f}{\partial x_2 \partial x_1} \tag{5}$$

$$\mathbf{H}(f) = \frac{1}{(x_1^2 + x_2^2)^{3/2}} \begin{pmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1^2 \end{pmatrix} \tag{6}$$

$$= \frac{1}{\|x\|^3} \begin{pmatrix} x_2^2 & -x_1 x_2 \\ -x_1 x_2 & x_1^2 \end{pmatrix} \tag{7}$$

2.1.5 Rosenbrock Function

$$f(\mathbf{x}) = \sum_{i=1}^{d-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \right]$$

Gradient:

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2(x_1 - 1) - 400x_1(x_2 - x_1^2) \\ 200(x_2 - x_1^2) - 400x_2(x_3 - x_2^2) + 2(x_2 - 1) \\ \vdots \\ 200(x_{d-1} - x_{d-2}^2) - 400x_{d-1}(x_d - x_{d-1}^2) + 2(x_{d-1} - 1) \\ 200(x_d - x_{d-1}^2) \end{bmatrix}$$

Hessian:

$$H(f) = \begin{bmatrix} 2 + 800x_1^2 - 400(x_2 - x_1^2) & -400x_1 & 0 & \cdots \\ -400x_1 & 2 + 800x_2^2 - 400(x_3 - x_2^2) & -400x_2 & \cdots \\ 0 & -400x_2 & 2 + 800x_3^2 - 400(x_4 - x_3^2) & \cdots \\ \vdots & \vdots & \vdots & \ddots \\ 0 & \cdots & 0 & -400 \end{bmatrix}$$

—

3 Manual Computation of Minima

3.1 Trid Function

Solving $\nabla f = 0$:

$$\begin{cases} 2x_1 - 2 - x_2 = 0 \\ 2x_2 - 2 - x_1 = 0 \end{cases} \Rightarrow x_1 = x_2 = 2$$

Hessian is positive definite, confirming minimum at $(2, 2)$.

3.2 Three-Hump Camel Function

Solving $\nabla f = 0$: From $y = -\frac{x}{2}$, substituting into the first equation yields minima at $(0, 0)$ and $(\pm 2.75, \mp 1.375)$. Hessian at $(0, 0)$ is positive definite.

3.3 Styblinski-Tang Function

Solving $4x^3 - 32x + 5 = 0$ gives approximate minima at $x \approx -2.903$ and $x \approx 2.747$. Hessian confirms these are minima.

3.4 Root of Square Function

Minimum at $\mathbf{x} = \mathbf{0}$ since $\nabla f = 0$.

—

4 Algorithm Convergence Analysis

- **Newton's Method** failed for Rosenbrock when initialized far from $(1, 1)$, and for Styblinski-Tang near saddle points.
- **Bisection-Wolfe** struggled with Rosenbrock's narrow valley and Three-Hump Camel's non-convex regions.
- **Armijo Line Search** exhibited slow convergence on Rosenbrock due to ill-conditioning.
- **Levenberg-Marquardt** and **Damped Newton** showed robust convergence across all functions.

—

5 Plots: Function vs. Iterations

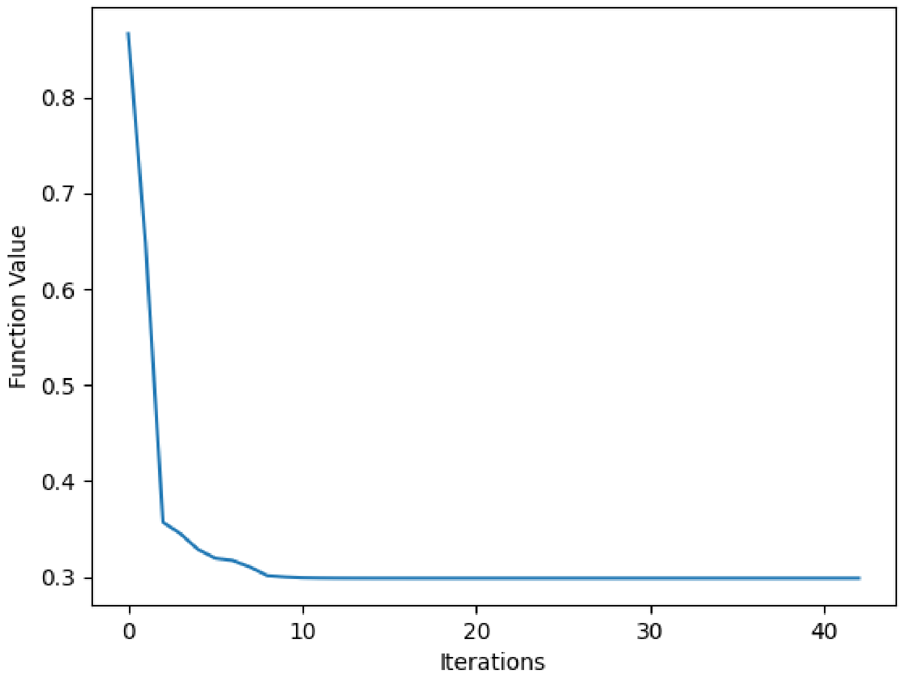

Figure 1: three hump camel function [-2 1] Backtracking-Armijo

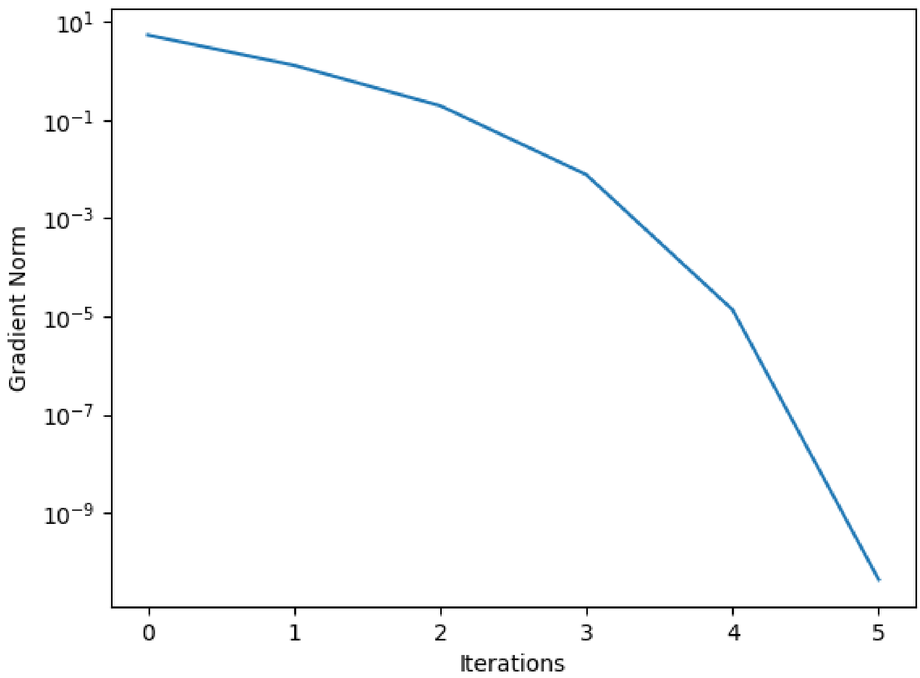

Figure 2: Gradient of three hump camel function [-2 1] Combined

—

6 Contour Plots with Optimization Paths

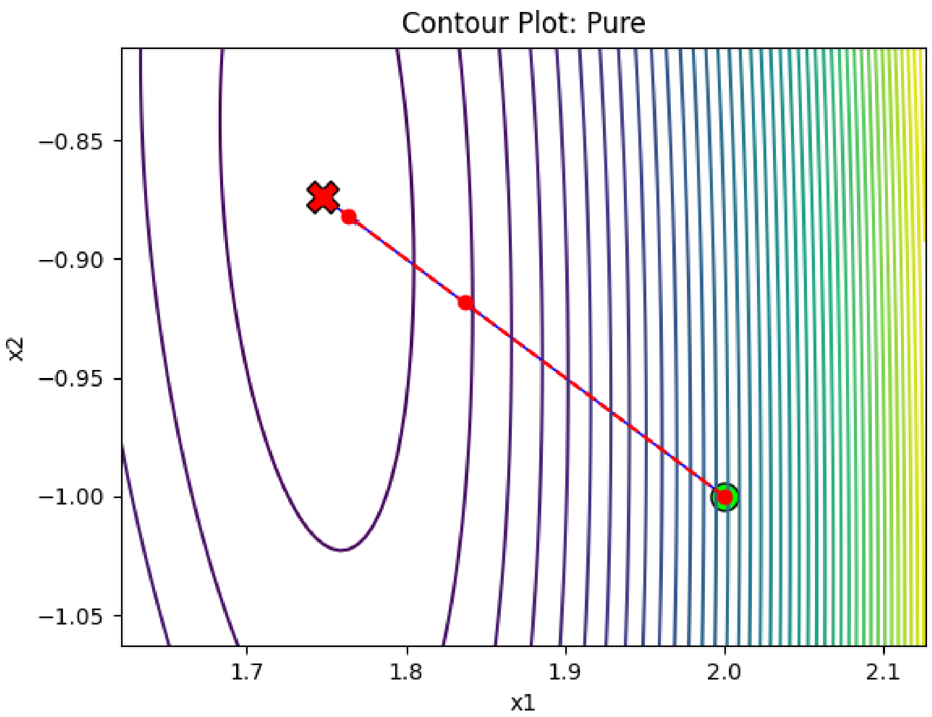

Figure 3: three hump camel function [2. -1.] Pure Newton contour Plot

—