Mathematics for Intelligent Systems
Lecture 10 Homework
(Optimization)

Marc Toussaint, Andrea Baisero

1 Convergence proof

a) Given a function $f : \mathbb{R}^n \to \mathbb{R}$ with $f_{\texttt{min}} = \min_x f(x)$. Assume that its Hessian—that is, the eigenvalues of $\nabla^2 f$—are lower bounded by $m > 0$ and upper bounded by $M > m$. Prove that for any $x \in \mathbb{R}^n$ it holds

$$f(x) - \frac{1}{2m}|\nabla f(x)|^2 \le f_{\texttt{min}} \le f(x) - \frac{1}{2M}|\nabla f(x)|^2 \ .$$

Tip: Start with bounding the 2nd-order Taylor expansion. Then consider the minima of these bounds. Note, it also follows:

$$|\nabla f(x)|^2 \ge 2m(f(x) - f_{\texttt{min}}) \ .$$

b) Consider backtracking line search with Wolfe parameter $\rho_{\texttt{ls}} \le \frac{1}{2}$, and step decrease factor ρ_α^-. First prove that line search terminates the latest when $\frac{\rho_\alpha^-}{M} \le \alpha \le \frac{1}{M}$, and then it found a new point y for which

$$f(y) \le f(x) - \frac{\rho_{\texttt{ls}}\rho_\alpha^-}{M}|\nabla f(x)|^2 \ .$$

From this, using the result from a), prove the convergence equation

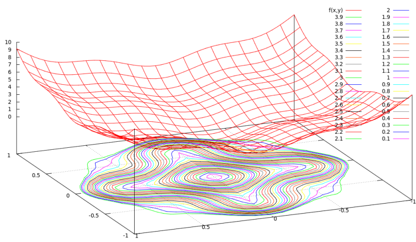
$$f(y) - f_{\texttt{min}} \le \left[1 - \frac{2m\rho_{\texttt{ls}}\rho_\alpha^-}{M}\right] (f(x) - f_{\texttt{min}}) \ .$$

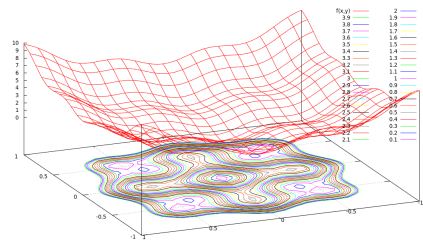
2 Misc

The Gauss-Newton method uses the "approximate Hessian" $2\nabla\phi(x)^\top\nabla\phi(x)$. First show that for any vector $v \in \mathbb{R}^n$ the matrix $vv^\top$ is symmetric and semi-positive-definite.[1] From this, how can you argue that $\nabla\phi(x)^\top\nabla\phi(x)$ is also symmetric and semi-positive-definite?

[1] A matrix $A \in \mathbb{R}^{n\times n}$ is semi-positive-definite simply when for any $x \in \mathbb{R}^n$ it holds $x^\top Ax \ge 0$. Intuitively: A might be a metric as it "measures" the norm of any x as positive. Or: If A is a Hessian, the function is (locally) convex.

3 Gauss-Newton





In $x \in \mathbb{R}^2$ consider the function

$$f(x) = \phi(x)^\top \phi(x) \,, \phi(x) = \begin{pmatrix} \sin(ax_1) \\ \sin(acx_2) \\ 2x_1 \\ 2cx_2 \end{pmatrix}$$

The function is plotted above for $a = 4$ (left) and $a = 5$ (right, having local minima), and conditioning $c = 1$. The function is non-convex.

Extend your backtracking method implemented in the last week's exercise to a Gauss-Newton method (with constant λ) to solve the unconstrained minimization problem $\min_x f(x)$ for a random start point in $x \in [-1, 1]^2$. Compare the algorithm for $a = 4$ and $a = 5$ and conditioning $c = 3$ with gradient descent.