Introduction to Optimization
Exercise 1

Marc Toussaint
Machine Learning & Robotics lab, U Stuttgart
Universitätsstraße 38, 70569 Stuttgart, Germany

May 1, 2014

As I was ill last week, we can first rediscuss open questions on the previous exercises.

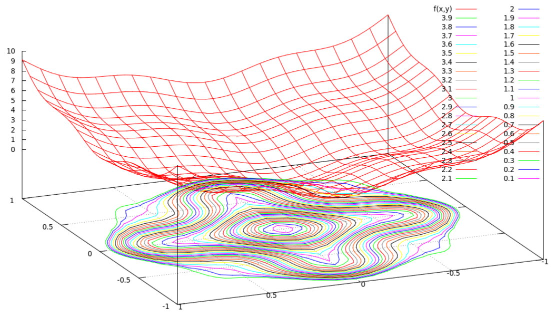
1 Misc

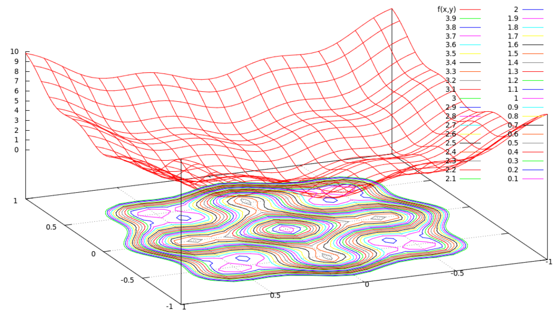
a) How do you have to choose the "damping" λ depending on $\nabla^2 f(x)$ in line 3 of the Newton method (slide 02-16) to ensure that the d is always well defined (i.e., finite)?

b) 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 $\phi(x)^\top\nabla\phi(x)$ is also symmetric and semi-positive-definite?

c) In the context of BFGS, convince yourself that choosing $H^{-1} = \frac{\delta\delta^\top}{\delta^\top y}$ indeed fulfills the desired relation $\delta = H^{-1}y$, where δ and y are defined as on slide 02-23. Are there other choices of H^{-1} that fulfill the relation? Which?

2 Gauss-Newton





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

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

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.

a) 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.

b) If you work in Octave/Matlab or alike, optimize the function also using the `fminunc` routine from Octave. (Typically this uses BFGS internally.)

[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 $x^\top Ax \geq 0$. 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.