Numerical Optimization – Tutorial on the gradient method

L. Desbat & F. Iutzeler

A. Differentiability, Minima, and Convexity

Exercise 1 (Quadratic functions)**.**

a. In $\mathbb{R}^n$, compute the gradient of the squared Euclidean norm $\|\cdot\|_2^2$ at a generic point $x \in \mathbb{R}^n$.
b. Let A be an $m \times n$ real matrix and b a size-m real vector. We define $f(x) = \|Ax - b\|_2^2$. For a generic vector $a \in \mathbb{R}^n$, compute the gradient $\nabla f(a)$ and Hessian $H_f(a)$.
c. Let C be an $n\times n$ real matrix, d a size-n real vector, and $e \in \mathbb{R}$. We define $g(x) = x^{\mathrm{T}}Cx + d^{\mathrm{T}}x + e$. For a generic vector $a \in \mathbb{R}^n$, compute the gradient $\nabla g(a)$ and Hessian $H_g(a)$.
d. Can all functions of the form of f and be written in the form of g? And conversely?

Exercise 2 (Basic Differential calculus)**.** Use the composition lemma to compute the gradients of:

a. $f_1(x) = \|Ax - b\|_2^2$.
b. $f_2(x) = \|x\|_2$.

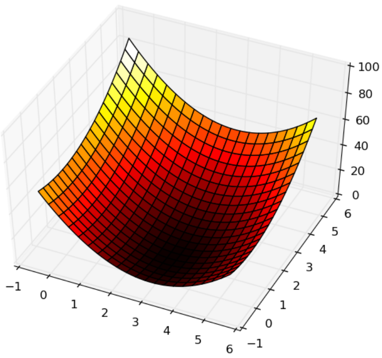
Exercise 3 (Preparing the Lab)**.** In the first lab, we will consider the following toy functions:

$$
\begin{array}{rrcl}
f: & \mathbb{R}^2 & \to & \mathbb{R} \\
& (x_1, x_2) & \mapsto & 4(x_1 - 3)^2 + 2(x_2 - 1)^2 \\
g: & \mathbb{R}^2 & \to & \mathbb{R} \\
& (x_1, x_2) & \mapsto & \log(1 + \exp(4(x_1 - 3)^2) + \exp(2(x_2 - 1)^2)) - \log(3) \\
r: & \mathbb{R}^2 & \to & \mathbb{R} \\
& (x_1, x_2) & \mapsto & (1 - x_1)^2 + 100(x_2 - x_1^2)^2 \\
t: & \mathbb{R}^2 & \to & \mathbb{R} \\
& (x_1, x_2) & \mapsto & (0.6x_1 + 0.2x_2)^2 \left((0.6x_1 + 0.2x_2)^2 - 4(0.6x_1 + 0.2x_2) + 4\right) + (-0.2x_1 + 0.6x_2)^2 \\
p: & \mathbb{R}^2 & \to & \mathbb{R} \\
& (x_1, x_2) & \mapsto & |x_1 - 3| + 2\,|x_2 - 1|\,.
\end{array}
$$

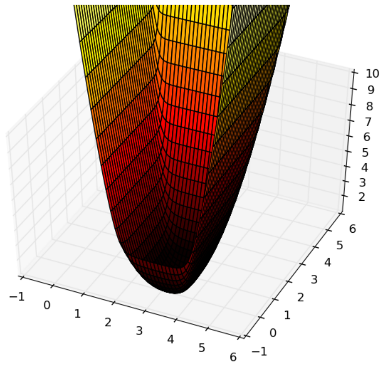
a. From the 3D plots of A.1, which functions are visibly non-convex.
b. For all five functions, show that they are convex or give an argument for their non-convexity.
c. For functions f, g, r, t, compute their gradient.
d. For functions f, g, compute their Hessian.

Exercise 4 (Fundamentals of convexity)**.**

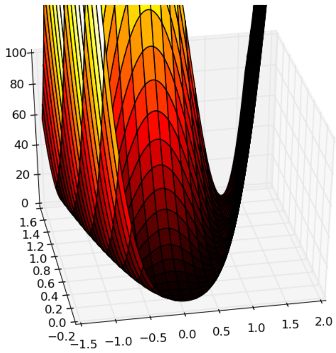
a. Let f and g be two convex functions. Show that $m(x) = \max(f(x), g(x))$ is convex.
b. Show that $f_1(x) = \max(x^2 - 1, 0)$ is convex.
c. Let f be a convex function and g be a convex, non-decreasing function. Show that $c(x) = g(f(x))$ is convex.
d. Show that $f_2(x) = \exp(x^2)$ is convex. What about $f_3(x) = \exp(-x^2)$
e. Justify why the 1-norm, the 2 norm, and the squared 2-norm are convex.



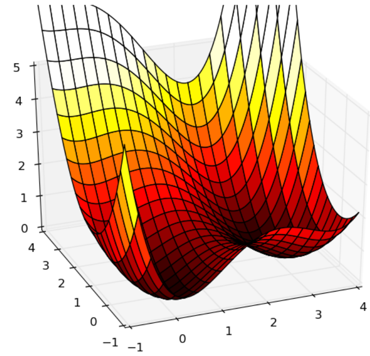
(A) a *simple* function: f



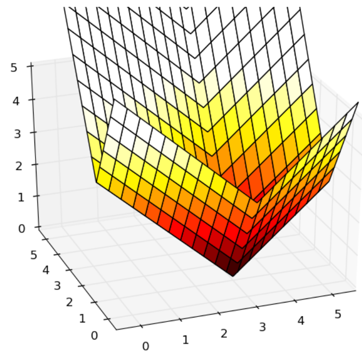
(B) some *harder* function: g



(C) *Rosenbrock*'s function: r



(D) *two pits* function: t



(E) *polyhedral* function: p

FIGURE A.1. 3D plots of the considered functions

Exercise 5 (Strict and strong convexity)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said

- *strictly convex* if for any $x \neq y \in \mathbb{R}^n$ and any $\alpha \in]0, 1[$

$$f(\alpha x + (1 - \alpha)y) < \alpha f(x) + (1 - \alpha)f(y)$$

- *strongly convex* if there exists $\beta > 0$ such that $f - \frac{\beta}{2}\|\cdot\|_2^2$ is convex.

a. For a strictly convex function f, show that the problem

$$\left\{ \begin{array}{l} \min f(x) \\ x \in C \end{array} \right.$$

where C is a convex set admits at most one solution.

b. Show that a strongly convex function is also strictly convex.
(hint: use the identity $\|\alpha x + (1-\alpha)y\|^2 = \alpha\|x\|^2 + (1-\alpha)\|y\|^2 - \alpha(1-\alpha)\|x-y\|^2$*.)*

Exercise 6 (Optimality conditions)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function and $\bar{x} \in \mathbb{R}^n$. We suppose that f admits a local minimum at $\bar{x}$ that is $f(x) \geq f(\bar{x})$ for all x in a neighborhood[1] of $\bar{x}$.

a. For any direction $u \in \mathbb{R}^n$, we define the $\mathbb{R} \to \mathbb{R}$ function $q(t) = f(\bar{x} + tu)$. Compute $q'(t)$.
b. By using the first order Taylor expansion of q at 0, show that $\nabla f(\bar{x}) = 0$.
c. Compute $q''(t)$. By using the second order Taylor expansion of q at 0, show that $\nabla^2 f(\bar{x})$ is positive semi-definite.

B. the Gradient Algorithm

Exercise 7 (Descent lemma)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be L-smooth if it is differentiable and its gradient ∇f is L-Lipchitz continuous, that is

$$\forall x, y \in \mathbb{R}^n, \quad \|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

The goal of the exercise is to prove that if $f : \mathbb{R}^n \to \mathbb{R}$ is L-smooth, then for all $x, y \in \mathbb{R}^n$,

$$f(x) \leq f(y) + (x-y)^{\mathrm{T}}\nabla f(y) + \frac{L}{2}\|x-y\|^2$$

a. Starting from fundamental theorem of calculus stating that for all $x, y \in \mathbb{R}^n$,

$$f(x) - f(y) = \int_0^1 (x-y)^{\mathrm{T}}\nabla f(y + t(x-y))\mathrm{d}t$$

prove the descent lemma.

b. Give a function for which the inequality is tight and one for which it is not.

Exercise 8 (Smooth functions)**.** Consider the constant stepsize gradient algorithm $x_{k+1} = x_k - \gamma\nabla f(x_k)$ on an L-smooth function f with some minimizer (i.e. some $x^\star$ such that $f(x) \geq f(x^\star)$ for all x).

a. Use the *descent lemma* to prove convergence of the sequence $(f(x_k))_k$ when $\gamma \leq 2/L$.
b. Did you use at some point that the function was convex? Conclude about the convergence of the gradient algorithm on smooth non-convex functions.

[1] Formally, one would write $\forall x \in \mathbb{R}^n$ such that $\|x - \bar{x}\| \leq \varepsilon$ for $\varepsilon > 0$ and some norm $\|\cdot\|$.