TMA4180
Optimization Theory
Spring 2015

Norwegian University of Science and Technology
Department of Mathematical Sciences

Exercise set 1 solution

1 **a)** The gradient and Hessian of the Rosenbrock function are given as following:

$$\nabla f = \begin{bmatrix} -400x_1(x_2 - x_1^2) - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix}$$

$$\nabla^2 f = \begin{bmatrix} 2 - 400x_2 + 1200x_1^2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix}.$$

b) We search for extreme value candidates by setting $\nabla f = 0$, and find the only viable candidate to be the point (1,1). The Hessian at (1,1) is

$$\nabla^2 f(1,1) = \begin{bmatrix} 802 & -400 \\ -400 & 200 \end{bmatrix},$$

with eigenvalues $\lambda_{\pm} = 501 \pm \sqrt{250601}$, both of which are greater than zero. Thus, $\nabla^2 f(1,1)$ is positive definite, and (1,1) is a strict local minimizer of f, by Theorem 2.4 in N&W. Since (1,1) is the only extreme point of f and since $f(x) \to \infty$ as $||x|| \to \infty$, it is also the unique global minimizer of f.

2 For any sequence x_k converging to x, we have

$$f(x) = \sup_{i \in I} f_i(x) \leq \sup_{i \in I} \liminf_{k \to \infty} f_i(x_k)$$

by lower semi-continuity of the functions f_i. Moreover, we have that for any f_i,

$$\liminf_{k \to \infty} f_i(x_k) \leq \liminf_{k \to \infty} \sup_{i \in I} f_i(x_k),$$

and especially,

$$\sup_{i \in I} \liminf_{k \to \infty} f_i(x_k) \leq \liminf_{k \to \infty} \sup_{i \in I} f_i(x_k).$$

Hence,

$$f(x) \leq \liminf_{k \to \infty} \sup_{i \in I} f_i(x_k) = \liminf_{k \to \infty} f(x_k),$$

and so f is lower semi-continuous.

3 The matter of showing that a function is lower semi-continuous becomes easier with two additional properties. The first is that a function is continuous if and only if it is lower and upper semi-continuous. Hence, any continuous function is lower semi-continuous. The second property is that if f and g are two lower semi-continuous functions, then the function $f+g$ is also lower semi-continuous, since

$$\liminf_{k\to\infty}\{f(x_k)+g(x_k)\} \geq \liminf_{k\to\infty} f(x_k) + \liminf_{k\to\infty} g(x_k) \geq f(x)+g(x).$$

(This may easily be extended to finite sums of lower semi-continuous functions.)

a) We split the function $f(x) = x^4 - 20x^3 + \sup_{k\in\mathbb{N}} \sin(kx)$ into two functions:

$$f_1(x) = x^4 - 20x^3,$$
$$f_2(x) = \sup_{k\in\mathbb{N}} \sin(kx).$$

Since f_1 is a polynomial, it is continuous and hence lower semi-continuous. Next, we see that $\sin(kx)$ is also a continuous function and thus lower semi-continuous. From exercise 2 (or the lecture notes) we know that since f_2 is of the form

$$f_2(x) = \sup_{k\in\mathbb{N}} f^k(x),$$

with each f^k lower semi-continuous, then f_2 is also lower semi-continuous. Since $f = f_1 + f_2$, with f_1 and f_2 lower semi-continuous, then f is also lower semi-continuous.
To show that f is coercive, we first observe that since $x \in \mathbb{R}$, $||x|| \to \infty$ implies $x \to \pm\infty$. Since $x^4 \to \infty$ as $x \to \pm\infty$ and dominates the two remaining terms in f, then $f(x_k) \to \infty$ for every sequence where $||x_k|| \to \infty$, and so f is coercive. By Theorem 10 in the lecture notes, f has at least one global minimizer since f is lower semi-continuous and coercive on $\mathbb{R}$, which is closed and nonempty.

b) We employ the same strategy as above, and find that since both the functions

$$g_1(x) = \mathrm{e}^x,$$
$$g_2(x) = -\frac{1}{x^2+1}$$

are continuous and thereby lower semi-continuous, then $g = g_1 + g_2$ is lower semi-continuous. However, g is not coercive, since $x \to -\infty$ implies $g(x) \to 0$. Therefore, Theorem 10 from the lecture notes does not apply, and since the domain of definition for g is $\mathbb{R}$, which is unbounded, neither does Theorem 8 from the lecture notes. However, since $f(-1) < 0$ and since f is continuous and bounded from below, there does exist a global minimum.

c) By the same reasoning as before, h is lower semi-continuous. It is not coercive, as can be seen by taking any sequence $x_k = (x_{k,1}, x_{k,2})$ with $x_{k,1} = 1$ and $x_{k,2} \to -\infty$, since $h(x_k) \to -\infty$ in that case. This also disproves the existence of a global minimizer.

4 Since f is convex, all local minima are global minima by Theorem 2.5 in N&W, i.e. f obtains the same value at all minimizers. Call this value c, and define the set of minimizers

$$S = \{x \in \mathbb{R}^n | f(x) = c\}.$$

We want to show that S is convex, i.e. that

$$\text{for all } x, y \in S \text{ and } \alpha \in [0,1], \alpha x + (1-\alpha)y \in S.$$

To this end, we choose arbitrary x and y in S, and observe that since f is convex,

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y) = c.$$

Since c is the minimum obtainable value, we conclude that

$$f(\alpha x + (1-\alpha)y) = c,$$

and so $\alpha x + (1-\alpha)y \in S$. Since x and y were chosen arbitrarily in S, S is convex.

5 We want to show that a strictly convex function f has at most one global minimizer. Let us assume the opposite, and say that x and y are distinct global minimizers of f, such that $f(x) = f(y) = c$ is the global minimum value. Then, since f is strictly convex we have for all $\alpha \in (0,1)$:

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

meaning f obtains a lower value than c at the points $\alpha x + (1-\alpha)y$, which contradicts our assumption that x and y are global minimizers. Therefore, there cannot be more than one global minimizer.

An example of a strictly convex function with no global minimizer is the exponential function $f(x) = \mathrm{e}^x$. Since $f''(x) = e^x > 0$, it is strictly convex, and since $f'(x) = e^x > 0$, it has no extreme points and thereby no global minimizer.

6 Define $g(x) = \lambda ||x||^2$. Since f and g are continuous functions, $f_\lambda = f + g$ is continuous and thus lower semi-continuous. Also, since f is bounded below and $||x||^2 \to \infty$ as $||x|| \to \infty$, then $f_\lambda(x) \to \infty$ as $||x|| \to \infty$, so f_λ is coercive. Thus, by Theorem 10 in the lecture notes, there exists at least one global minimizer of f_λ.

Next, we observe that the Hessian of g is $2\lambda I$, where I is the identity matrix. Thus, the Hessian has only one eigenvalue $2\lambda > 0$, which means it is positive definite, and so g is strictly convex. The sum of a convex function and a strictly convex function is strictly convex, and so $f_\lambda = f + g$ is strictly convex. From the previous exercise, we know that a strictly convex function has at most one global minimizer. Since there exists at least one global minimizer and there can be at most one global minimizer, we conclude that f_λ has a unique global minimizer.

7 If we allow $c_2 \geq 1$, we defeat the purpose of the second Wolfe condition, the *curvature condition*

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f(x_k)^T p_k.$$

This requirement was introduced to rule out unacceptably short steps, but if $c_2 \geq 1$, then the requirement is satisfied in a neighborhood of $\alpha_k = 0$; it is clearly satisfied for $\alpha_k = 0$, since (keep in mind that $f(x_k)^T p_k < 0$ since we are searching in a descent direction)

$$\nabla f(x_k)^T p_k \geq c_2 \nabla f(x_k)^T p_k.$$

Assuming that $\varphi'(\alpha_k) = \nabla f(x_k + \alpha_k p_k)^T p_k$ is continuous, this means that the second Wolfe condition will hold in a neighborhood of $\alpha_k = 0$, and thereby arbitrarily small steps are allowed.

If we allow $c_1 > c_2$, there may not exist a positive step length satisfying the Wolfe conditions, which is best exemplified in a figure. Figure 1 shows an example where the desired slope is not obtained before the end of the region of sufficient decrease, meaning no step length satisfies both Wolfe conditions.

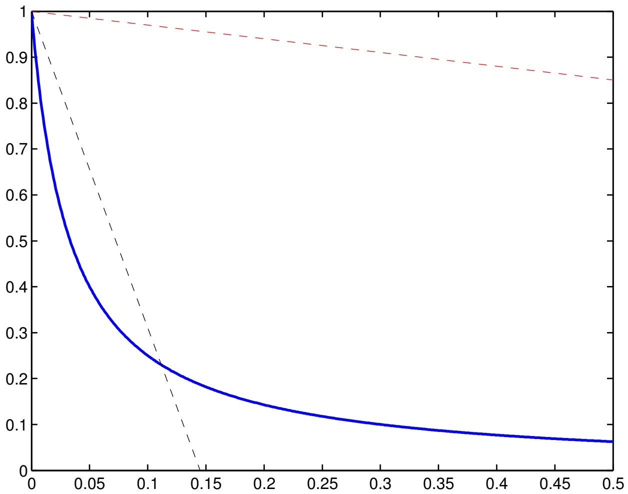

Figure 1: Blue: $\varphi(\alpha_k)$. Black dotted line: Line of sufficient decrease. Red dotted line: Desired slope.