

Optimisation methods – Homework 1

Claudio Maggioni

March 25, 2021

1 Exercise 1

1.1 Gradient and Hessian

The gradient and the Hessian for f are the following:

$$\nabla f = \begin{bmatrix} 2x_1 + x_2 \cdot \cos(x_1) \\ 9x_2^2 + \sin(x_1) \end{bmatrix}$$

$$H_f = \begin{bmatrix} 2 - x_2 \cdot \sin(x_1) & \cos(x_1) \\ \cos(x_1) & 18x_2 \end{bmatrix}$$

1.2 Taylor expansion

$$f(h) = 0 + \left\langle \begin{bmatrix} 0+0 \\ 0+0 \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \right\rangle + \frac{1}{2} \left\langle \begin{bmatrix} 2-0 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \right\rangle + O(\|h\|^3)$$

$$f(h) = \frac{1}{2} \left\langle \begin{bmatrix} 2h_1 + h_2 \\ h_1 \end{bmatrix}, \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} \right\rangle + O(\|h\|^3)$$

$$f(h) = \frac{1}{2} (2h_1^2 + 2h_1h_2) + O(\|h\|^3)$$

$$f(h) = h_1^2 + h_1h_2 + O(\|h\|^3)$$

2 Exercise 2

2.1 Gradient and Hessian

For A symmetric, we have:

$$\frac{d}{dx} \langle b, x \rangle = \langle b, \cdot \rangle = b$$

$$\frac{d}{dx} \langle Ax, x \rangle = 2 \langle Ax, \cdot \rangle = 2Ax$$

Then:

$$\begin{aligned}\nabla J &= Ax - b \\ H_J &= \frac{d}{dx} \nabla J = A\end{aligned}$$

2.2 First order necessary condition

It is a necessary condition for a minimizer x^* of J that:

$$\nabla J(x^*) = 0 \Leftrightarrow Ax^* = b$$

2.3 Second order necessary condition

It is a necessary condition for a minimizer x^* of J that the first order necessary condition holds and:

$$\nabla^2 J(x^*) \geq 0 \Leftrightarrow A \text{ is positive semi-definite}$$

2.4 Sufficient conditions

It is a sufficient condition for x^* to be a minimizer of J that the first necessary condition is true and that:

$$\nabla^2 J(x^*) > 0 \Leftrightarrow A \text{ is positive definite}$$

2.5 Does $\min_{x \in R^n} J(x)$ have a unique solution?

Not in general. If for example we consider A and b to be only zeros, then $J(x) = 0$ for all $x \in R^n$ and thus J would have an infinite number of minimizers.

However, if A is guaranteed to be s.p.d, the minimizer would be unique because the first order necessary condition would hold only for one value x^* . This is because the linear system $Ax^* = b$ would have one and only one solution (due to A being full rank and that solution being the minimizer since the hessian would be always convex).

3 Exercise 3

3.1 Quadratic form

$f(x, y)$ can be written in quadratic form in the following way:

$$f(v) = v^T \begin{bmatrix} 1 & 0 \\ 0 & \mu \end{bmatrix} v + \begin{bmatrix} 0 \\ 0 \end{bmatrix}^T v$$

where:

$$v = \begin{bmatrix} x \\ y \end{bmatrix}$$

3.2 Matlab plotting with surf and contour plots

The graphs generated by MATLAB are shown below:

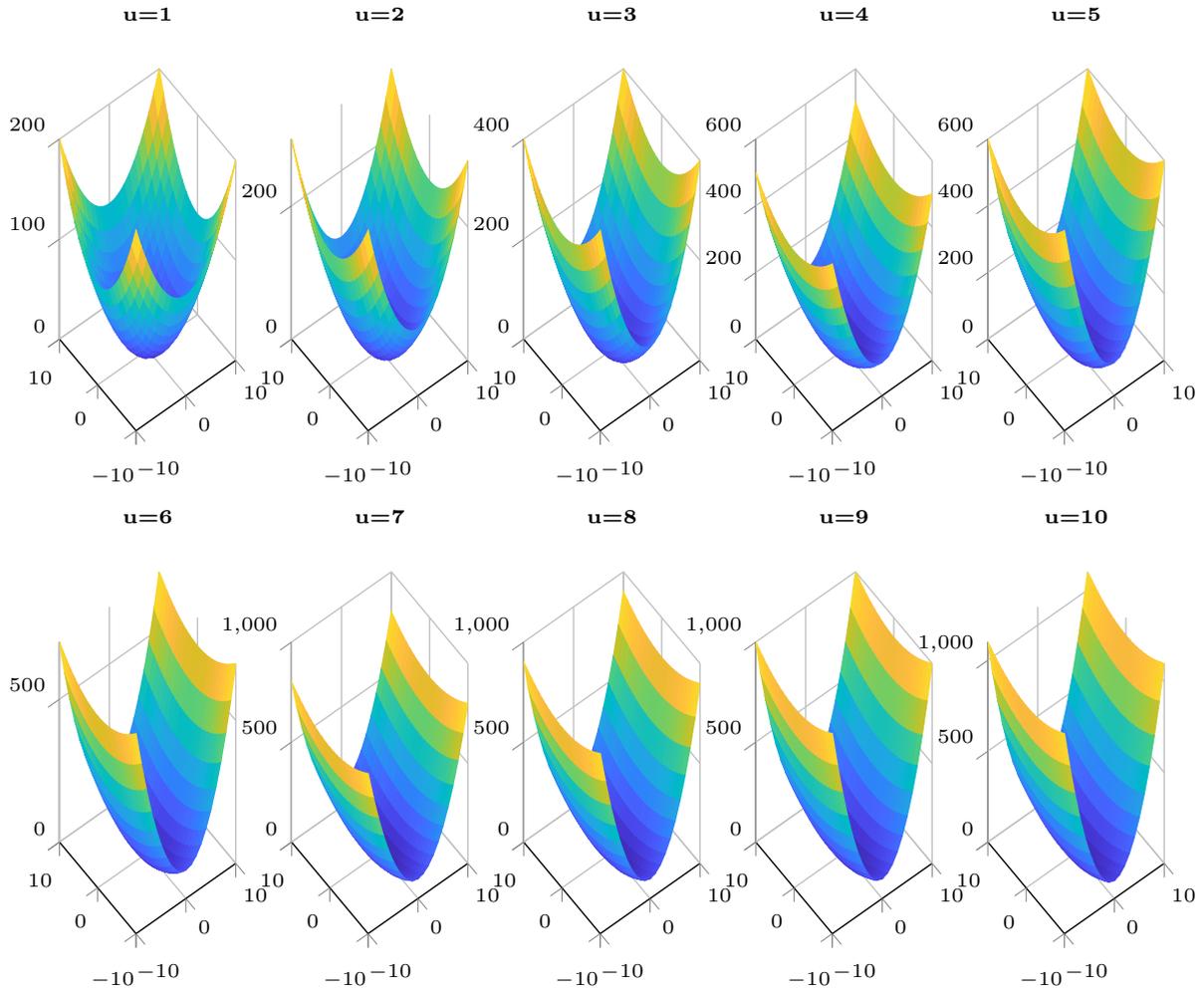


Figure 1: Surf plots for different values of μ

Isolines (found in Figure ??) get stretched along the y axis as μ increases. For $\mu \neq 1$, points well far away from the axes are a problem since picking search directions and steps using the gradient method iterations will zig-zag to the minimizer reaching it slowly.

From the surf plots (Figure ??), we can see that the behaviour of isolines is justified by a "stretching" of sorts of the function that causes the y axis to be steeper as μ increases.

3.3 Finding the optimal step length α

Considering p , our search direction, as the negative of the gradient (as dictated by the gradient method), we can rewrite the problem of finding an optimal step size α as the problem of minimizing the objective function along the line where p belongs. This can be written as minimizing a function $l(\alpha)$, where:

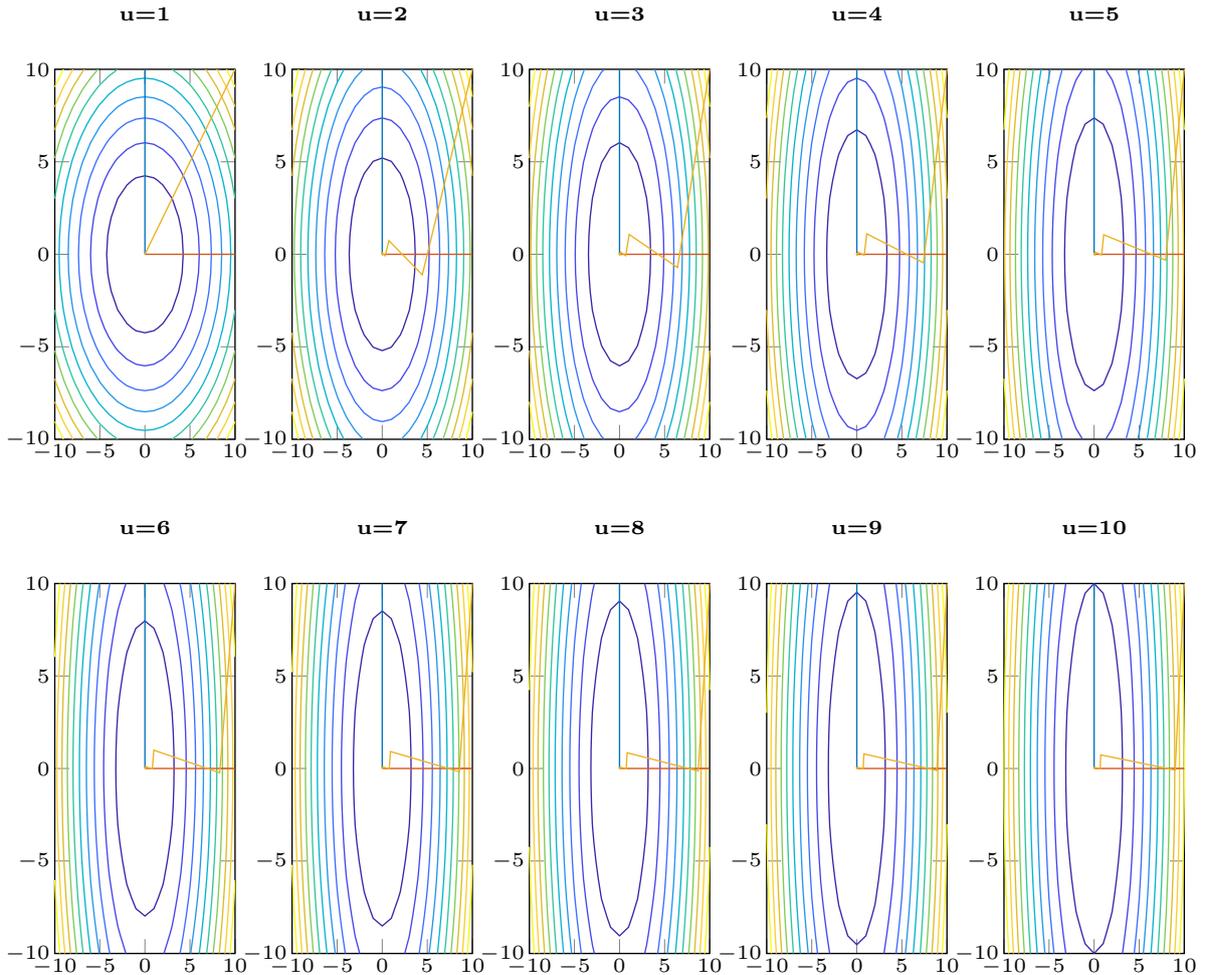


Figure 2: Contour plots and iteration steps. Red has $x_0 = [10 \ 0]^T$, yellow has $x_0 = [10 \ 10]^T$, and blue has $x_0 = [0 \ 10]^T$

$$l(\alpha) = \langle A(x + \alpha p), x + \alpha p \rangle$$

To minimize we compute the gradient of $l(\alpha)$ and fix it to zero to find a stationary point, finding a value for α in function of A , x and p .

$$l'(\alpha) = 2 \cdot \langle A(x + \alpha p), p \rangle = 2 \cdot (\langle Ax, p \rangle + \alpha \langle Ap, p \rangle)$$

$$l'(\alpha) = 0 \Leftrightarrow \alpha = -\frac{\langle Ax, p \rangle}{\langle Ap, p \rangle}$$

Since A is s.p.d. by definition the hessian of function $l(\alpha)$ will always be positive, the stationary point found above is a minimizer of $l(\alpha)$ and thus the definition of α given above gives the optimal search step for the gradient method.

3.4 Matlab code for the gradient method and convergence results

The main MATLAB file to run to execute the gradient method is `ex3.m`. Convergence results and number of iterations are shown below, where the verbatim program output is written:

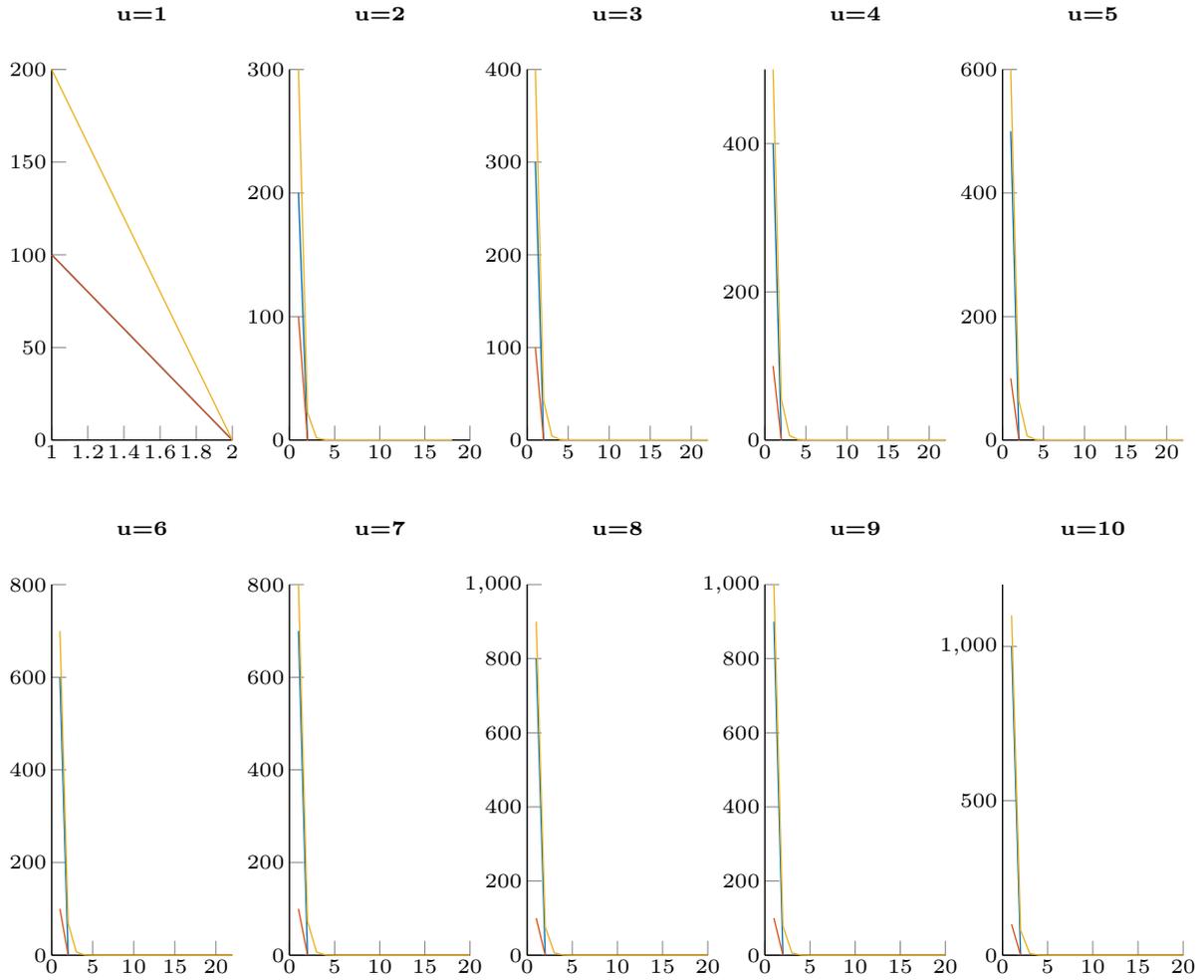


Figure 3: Iterations over values of the objective function. Red has $x_0 = [10 \ 0]^T$, yellow has $x_0 = [10 \ 10]^T$, and blue has $x_0 = [0 \ 10]^T$.

```

u= 1 x0=[ 0,10] it= 2 x=[0,0]
u= 1 x0=[10, 0] it= 2 x=[0,0]
u= 1 x0=[10,10] it= 2 x=[0,0]
u= 2 x0=[ 0,10] it= 2 x=[0,0]
u= 2 x0=[10, 0] it= 2 x=[0,0]
u= 2 x0=[10,10] it=18 x=[4.028537e-09,-1.007134e-09]
u= 3 x0=[ 0,10] it= 2 x=[0,0]
u= 3 x0=[10, 0] it= 2 x=[0,0]
u= 3 x0=[10,10] it=22 x=[1.281583e-09,-1.423981e-10]
u= 4 x0=[ 0,10] it= 2 x=[0,0]
u= 4 x0=[10, 0] it= 2 x=[0,0]
u= 4 x0=[10,10] it=22 x=[2.053616e-09,-1.283510e-10]
u= 5 x0=[ 0,10] it= 2 x=[0,0]
u= 5 x0=[10, 0] it= 2 x=[0,0]
u= 5 x0=[10,10] it=22 x=[1.397370e-09,-5.589479e-11]
u= 6 x0=[ 0,10] it= 2 x=[0,0]
u= 6 x0=[10, 0] it= 2 x=[0,0]
u= 6 x0=[10,10] it=22 x=[7.313877e-10,-2.031632e-11]
u= 7 x0=[ 0,10] it= 2 x=[0,0]

```

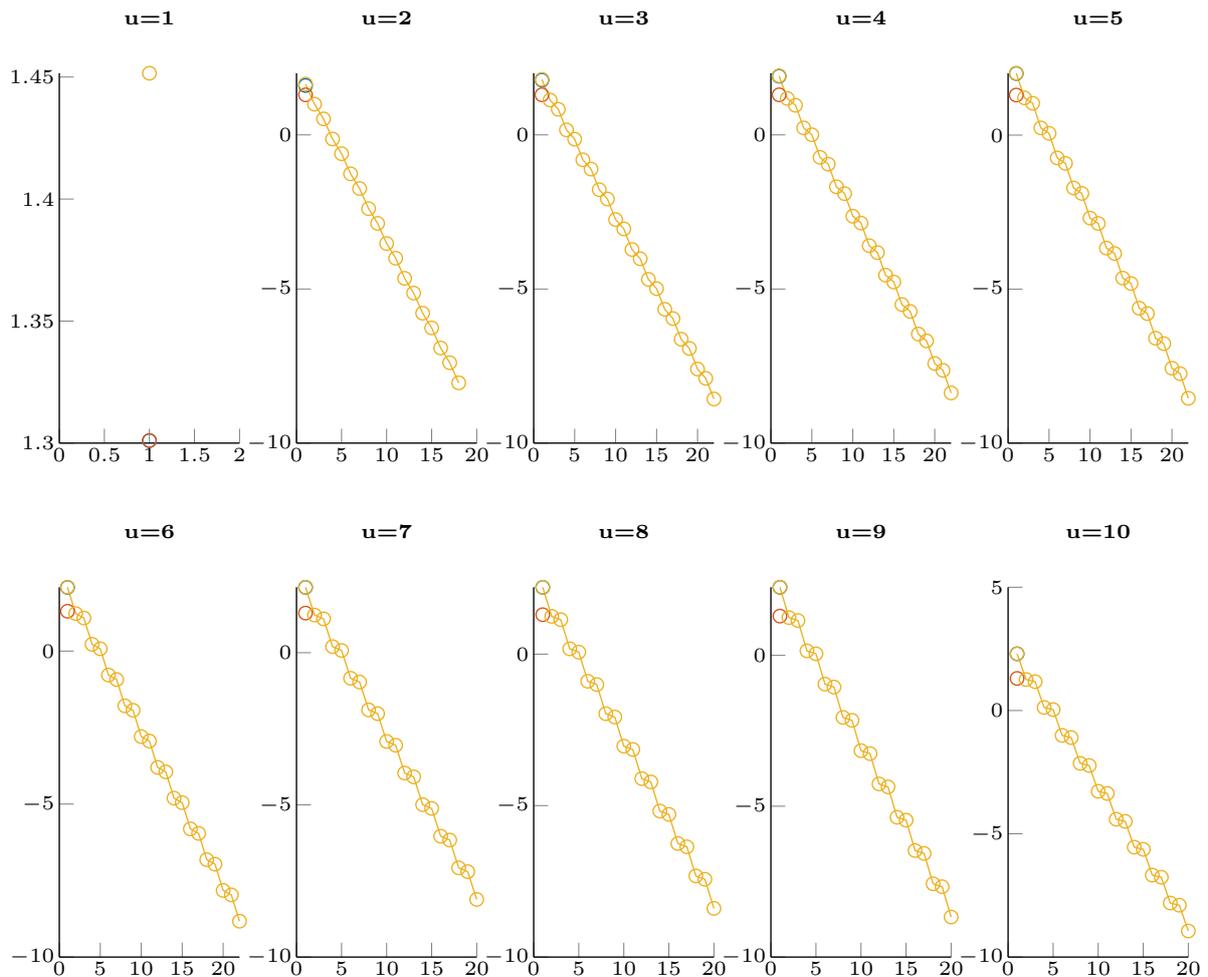


Figure 4: Iterations over base 10 logarithm of gradient norms. Note that for $\mu = 1$ the search immediately converges to the exact minimizer no matter the value of x_0 , so no gradient norm other than the very first one is recorded. Again, Red has $x_0 = [10 \ 0]^T$, yellow has $x_0 = [10 \ 10]^T$, and blue has $x_0 = [0 \ 10]^T$.

```

u= 7 x0=[10, 0] it= 2 x=[0,0]
u= 7 x0=[10,10] it=20 x=[3.868636e-09,-7.895176e-11]
u= 8 x0=[ 0,10] it= 2 x=[0,0]
u= 8 x0=[10, 0] it= 2 x=[0,0]
u= 8 x0=[10,10] it=20 x=[2.002149e-09,-3.128358e-11]
u= 9 x0=[ 0,10] it= 2 x=[0,0]
u= 9 x0=[10, 0] it= 2 x=[0,0]
u= 9 x0=[10,10] it=20 x=[1.052322e-09,-1.299163e-11]
u=10 x0=[ 0,10] it= 2 x=[0,0]
u=10 x0=[10, 0] it= 2 x=[0,0]
u=10 x0=[10,10] it=20 x=[5.671954e-10,-5.671954e-12]

```

3.5 Comments on the various plots

The objective function plots and the gradient norm plots can be found respectively in figures ?? and ??.

What has been said before about the convergence of the gradient method is additionally showed in the last two sets of plots. From the objective function plot we can see that iterations starting from $[10 \ 10]^T$ (depicted in yellow) take the highest number of iterations to reach the minimizer (or an acceptable approximation of it). The zig-zag behaviour described before can be also observed in the contour plots, showing the iteration steps taken for each μ and starting from each x_0 .

Finally, in the gradient norm plots a phenomena that creates increasingly flatter plateaus as μ increases can be observed.