MTH8408: Exercices

André Phu-Van Nguyen

Unconstrained optimization (slides)

Exercise 6/11

Prove the first and second order optimality conditions using

$$f(x^* + d) - f(x^*) = \nabla f(x*)^T d + \frac{1}{2} d^T \nabla^2 f(x^*) d + o(||d||^2)$$

Answer

N1 if $x*$ is a local minimum and $f \in \mathcal{C}^1$, then $\nabla f(x^*) = 0$

If $\nabla f(x^*) = 0$ then $\cancel{\frac{1}{2} d^T \nabla^2 f(x^*) d = 0}$ thus we only consider

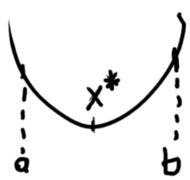
$$f(x^* + d) - f(x^*) = \nabla f(x*)^T d + o(||d||^{\cancel{2}})$$

Supposing that $o(||d||^2)$ can be neglected we can rearrange the equation into

$$\cancel{\frac{f(x^* + d) - f(x^*)}{d}} = \nabla f(x*)^T$$

!!! Division par un vecteur!!!

which is close to the definition of a limit.



For an $x \in [a, b]$ we have $f(x) \geq f(x*)$

From a to x^*

We have $x < x^*$ and $f(x) - f(x^*) \geq 0$ thus

$$\frac{f(x) - f(x^*)}{x - x^*} \leq 0$$

From b to x^*

We have $x > x^*$ and $f(x) - f(x^*) \geq 0$ thus

$$\frac{f(x) - f(x^*)}{x - x^*} \geq 0$$

Going back to the definition of a limit we have

$$\lim_{d \to \infty} \frac{f(x^* + d) - f(x)}{d} \leq 0$$

$$\lim_{d \to \infty} \frac{f(x^* + d) - f(x)}{d} \geq 0$$

N2 if x^* is a local minimum and $f \in \mathcal{C}^2$, then $\nabla f(x^*) = 0$ and $\nabla^2 f(x^*) = 0$

If $f \in \mathcal{C}^2$ then $f \in \mathcal{C}^1$ is also true, thus the previous results apply.

Incomplete

Exercise 10/71 - Unconstrained optimization with a quadratic objective

In S2, show that s^* is in fact a *global* minimum!

Answer

S2 if $Hs^* = -g$ and $H \overset{>}{\cancel{\succeq}} 0$ then s^* is a local minimum

Since $q(s)$ is a quadratic, a local minimum $s*$ of q is automatically a global minimum. C'est ce qu'il faut montrer et non brosser sous le tapis...

Exercise 46/71 - Conjugate gradient ???

Show that conjugate vectors are necessarily linearly independent.

Answer

Proof by contradiction. If p_i is linearly dependant to a p_j then we have

$$a_0 p_0 + a_1 p_1 + \ldots + a_n p_n = 0 \qquad n \in \mathbb{Z}^*$$

Where all a_n are non-zero. Multiply by A pas nécessairement mais au moins un est $\neq 0$.

$$a_i p_i^T A = 0$$

où sont les autres termes ?

Take the scalar product with p_i

$$a_i p_i^T A p_i = 0$$

We know that $A \succ 0$ which means that $p_i^T A p_i > 0$ thus $a_i = 0$ which is contradictory to a linearly dependant system.

incomplet

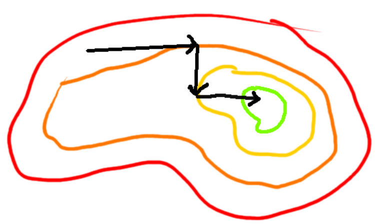
Exercise 47/71 - Conjugate gradient

Explain geometrically why $r_{k+1} \perp r_k$ i.e. $r_{k+1}^T r_k = 0$

Answer

??? préciser

When doing a linesearch we stop at a point where the direction is minimal. This also means we are tangent to the contour line (level curve), thus the next direction is orthogonal to the current one.



Exercise 47/71 - Conjugate gradient

Find a recurrence to update $q(x_k)$ in the conjugate gradient algorithm.

Answer

Should be something like

$$q(x_{k+1}) = q(x_k) + \nabla q(x_k)\alpha_k A p_k$$

So the current value of q plus the change in q and how much of that change.

Incomplete

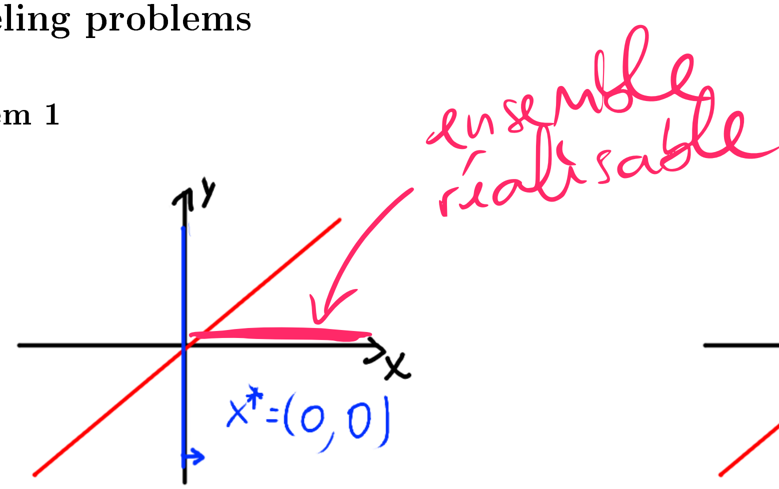
Exercice 48/71 - Conjugate gradient ???

[Prove?] $A \succ 0$ if and only if $p_k^T A p_k$ for all $k = 0,\ \ldots,\ n-1$

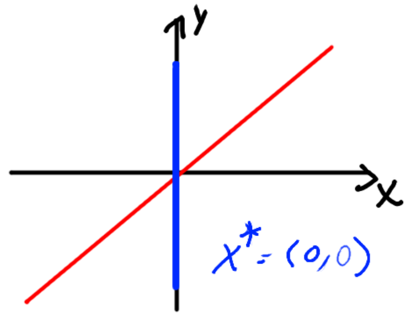
Incomplete

Modeling problems

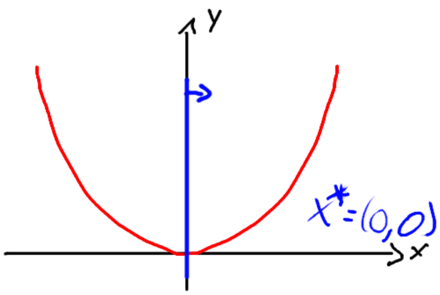
Problem 1



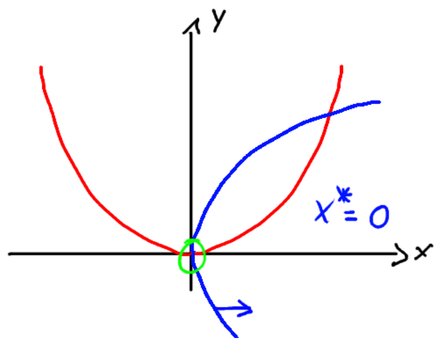
1. a)



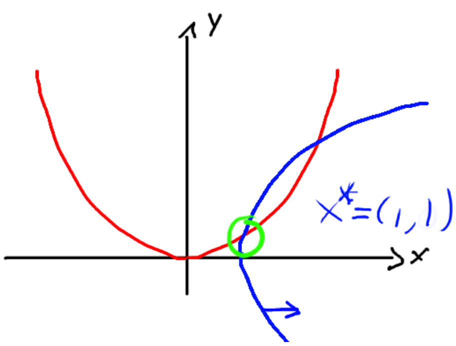
1. b)



1. c)

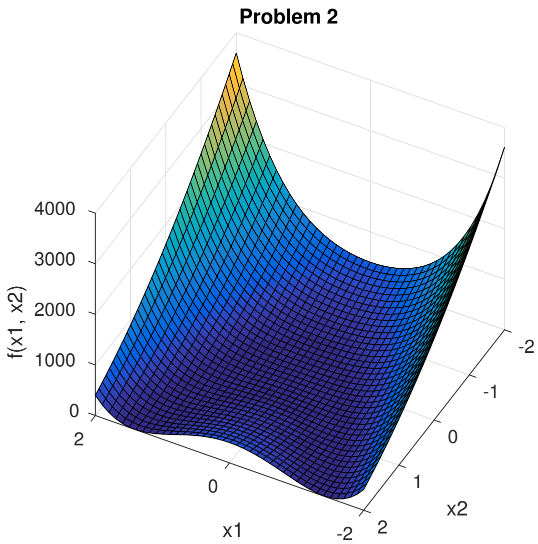


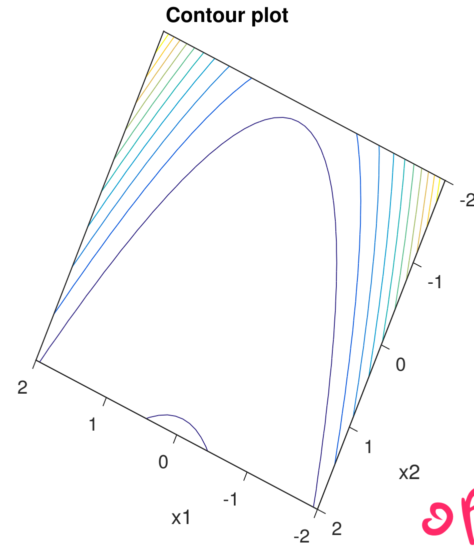
1. d)



1. e)

Problem 2





optimal objective value

2.a) We see that the two terms containing variables are square therefore the minimum is probably 0. We can get this result with $x^* = (1,1)$ it is unique because of it is the only root of $100(x_2 - x_1^2)^2 + (1 - x_1)^2$.

???

2.b)

```
model;

var x1;
var x2;

minimize function:
    100 * (x2 - x1^2)^2 + (1 - x1)^2
```

Output:

```
ampl: model modeling_pb2.mod;
ampl: solve;
MINOS 5.51: optimal solution found.
17 iterations, objective 9.247040366e-19
Nonlin evals: obj = 47, grad = 46.
ampl: display x1, x2;
x1 = 1
x2 = 1
```

Which confirms answer in **2.a)**.

2.c) Add an equality constraint which changes the solution.

```
model;

var x1;
var x2;

minimize function:
    100 * (x2 - x1^2)^2 + (1 - x1)^2;
```

```

subject to changeSolutionEqualityConstraint:
    x1 = 2;
```

Output:

```
ampl: model modeling_pb2.mod;
ampl: solve;
MINOS 5.51: optimal solution found.
2 iterations, objective 1
Nonlin evals: obj = 6, grad = 5.
ampl: display x1, x2;
x1 = 2
x2 = 4
```

2.d) Add an equality constraint which **does not** change the solution.

```
model;

var x1;
var x2;

minimize function:
    100 * (x2 - x1^2)^2 + (1 - x1)^2;

subject to dontChangeSolutionEqualityConstraint:
    x1 = 1;
```

Output:

```
ampl: model modeling_pb2.mod;
ampl: solve;
MINOS 5.51: optimal solution found.
2 iterations, objective 0
Nonlin evals: obj = 6, grad = 5.
ampl: display x1, x2;
x1 = 1
x2 = 1
```

2.e) Add an inequality constraint which changes the solution.

```
model;

var x1;
var x2;

minimize function:
    100 * (x2 - x1^2)^2 + (1 - x1)^2;

subject to changeSolutionInequalityConstraint:
    x1 >= 2;
```

Output:

```
ampl: model modeling_pb2.mod;
ampl: solve;
MINOS 5.51: optimal solution found.
2 iterations, objective 1
```

```
Nonlin evals: obj = 6, grad = 5.
ampl: display x1, x2;
x1 = 2
x2 = 4
```

2.f) Add an inequality constraint which **does not** changes the solution.

```
model;

var x1;
var x2;

minimize function:
    100 * (x2 - x1^2)^2 + (1 - x1)^2;

subject to dontChangeSolutionInequalityConstraint:
    x1 <= 2;
```

✓

Output:

```
ampl: model modeling_pb2.mod;
ampl: solve;
MINOS 5.51: optimal solution found.
17 iterations, objective 9.247040366e-19
Nonlin evals: obj = 47, grad = 46.
ampl: display x1, x2;
x1 = 1
x2 = 1
```

Problem 3

Model:

```
model;

var x1;
var x2;

minimize function:
    (x1 - 2)^2 + (x2 -1)^2;

subject to zeroConstraint:
    x1^2 - x2 <= 0;

subject to twoConstraint:
    x1 + x2 <= 2;
```

ok

Output:

```
ampl: model modeling_pb3.mod;
ampl: solve;
MINOS 5.51: optimal solution found.
12 iterations, objective 1
Nonlin evals: obj = 33, grad = 32, constrs = 33, Jac = 32.
ampl: display x1, x2;
x1 = 1
x2 = 1
```

Problem 4

4.a)

$$
\begin{aligned}
P_3(x) &= \frac{f^{(0)}(x_0)}{0!} + \frac{f^{(1)}(x_0)}{1!}(x-x_0) + \frac{f^{(2)}(x_0)}{2!}(x-x_0)^2 + \frac{f^{(3)}(x_0)}{3!}(x-x_0)^3 \\
&= \cos(1) + \frac{-\sin(1)}{1}(x-1) + \frac{-\cos(1)}{2}(x-1)^2 + \frac{\sin(1)}{6}(x-1)^3
\end{aligned}
$$

For x = 1.1 we have

$$
\begin{aligned}
P_3(1.1) &= \cos(1) + (-0.0841) + (-0.0027) + (0.0001) \\
&= 0.4536
\end{aligned}
$$

4.b)

$$
\begin{aligned}
f(x) &= \cos(x) \\
f'(x) &= \sin(\frac{1}{x})\frac{1}{x^2} \\
f''(x) &= \Big(\sin(\frac{1}{x})\Big)'\frac{1}{x^2} + \sin(\frac{1}{x})\Big(\frac{1}{x^2}\Big)' \\
&= \cos(\frac{1}{x})\frac{-1}{x^2}\frac{1}{x^2} + \sin(\frac{1}{x})\frac{-2}{x^3} \\
&= -\frac{\cos(\frac{1}{x}) + 2x\sin(\frac{1}{x})}{x^4}
\end{aligned}
$$

Thus the 2nd order Taylor expansion with $x_0 = 1$ is

$$
P_2(x) = \cos(1) + \sin(1)(x-1) - \Big(\frac{\cos(1) + 2\sin(1)}{2}\Big)(x-1)^2
$$

And for $x = 1.1$ we have

$$
\begin{aligned}
P_2(1.1) &= \cos(1) + \sin(1)(1.1-1) - \Big(\frac{\cos(1) + 2\sin(1)}{2}\Big)(1.1-1)^2 \\
&= 0.5403 + 0.0841 - 0.0111 \\
&= 0.6133
\end{aligned}
$$

4.c)

$$\begin{aligned}
f_{x_1} &= -200(x_1^2 + 2x_1 - x_2) + 2(x_1 - 1) \\
f_{x_2} &= 200(x_2 - x_1^2) \\
f_{x_1x_1} &= -200(2x_1 + 2) + 2 \\
f_{x_1x_2} &= -200(-1) = 200 \\
f_{x_2x_1} &= 200(1 - 2x_1) \\
f_{x_2x_2} &= 200
\end{aligned}$$

Thus the Taylor expansion in matrix form is

$$f(x,y) = f(0,0) + \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} -200(x_1^2 + 2x_1 - x_2) + 2(x_1 - 1) \\ 200(x_2 - x_1^2) \end{bmatrix} + \frac{1}{2!} \begin{bmatrix} x \\ y \end{bmatrix}^T \begin{bmatrix} -200(2x_1 + 2) + 2 & 200 \\ 200(1 - 2x_1) & 200 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

4.d)

Around the point $(x,y) = (a,b)$ we have

$$f(x,y) = f(a,b) + \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} a \\ b \end{bmatrix} \right)^T \begin{bmatrix} f_x(a,b) \\ f_y(a,b) \end{bmatrix} + \frac{1}{2!} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} a \\ b \end{bmatrix} \right)^T \begin{bmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{yx}(a,b) & f_{yy}(a,b) \end{bmatrix} \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} a \\ b \end{bmatrix} \right)$$

Problem 5

In the folowing problems we denote $\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \dots x_n \end{bmatrix}^T$

Function f_1

$$f_1(x) = c^T \mathbf{x}$$

Using the identity $\frac{\partial A\mathbf{x}}{\partial \mathbf{x}} = A^T$ the gradient is

$$\nabla f_1(\mathbf{x}) = \frac{\partial c^T \mathbf{x}}{\partial \mathbf{x}} = c$$

And the Hessian is given by

$$\frac{\partial c}{\partial \mathbf{x}} = 0$$

since the gradient was all constants.

Function f_2

$$f_2(x) = \frac{1}{2} x^T H x$$

To find the gradient we apply the fundamental identity [1]

[1] https://en.wikipedia.org/wiki/Matrix_calculus

$$\frac{\partial(\mathbf{u} \cdot \mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{u}^T \mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}\mathbf{v} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}\mathbf{A}^T\mathbf{u}$$

In our case $\mathbf{u} = \mathbf{v} = \mathbf{x}$

$$\begin{aligned}\frac{1}{2}\frac{\partial \mathbf{x}^T H\mathbf{x}}{\partial \mathbf{x}} &= \frac{1}{2}\left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}}H\mathbf{x} + \frac{\partial \mathbf{x}}{\partial \mathbf{x}}H^T\mathbf{x}\right) \\ &= \frac{1}{2}\left(H\mathbf{x} + H^T\mathbf{x}\right)\end{aligned}$$

And since H is a square symmetric matrix we can write

$$= H\mathbf{x}$$

✓

Thus the Hessian is

$$\frac{\partial H\mathbf{x}}{\partial \mathbf{x}} = H$$

✓

again through a fundamental identity.

Function $f_1 + f_2$

For

$$f_3(x) = f_1(x) + f_2(x) = c^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T H\mathbf{x}$$

by the sum rule, the gradient is the sum of the results derived above

$$\frac{\partial f_3(x)}{\partial \mathbf{x}} = c + H\mathbf{x}$$

✓

and the Hessian is

$$\frac{\partial(c + H\mathbf{x})}{\partial \mathbf{x}} = H$$

Problem 6 ???

Where

$$g(x) = f(Sx)$$

we have through the chain rule

$$\frac{\partial g(x)}{\partial x} = \frac{\partial x}{\partial x}\frac{\partial Sx}{\partial x}\frac{\partial f(Sx)}{\partial Sx} = S\frac{\partial f(Sx)}{\partial Sx} = S\frac{\partial f(x)}{\partial x}$$

and the Hessian is

$$\frac{\partial}{\partial x}S\frac{\partial f(x)}{\partial x} = S\frac{\partial^2 f(x)}{\partial^2 x}$$

écrire sous forme vectorielle et matricielle :
$\nabla g(x) = \ldots$
$\nabla^2 g(x) = \ldots$