Optimization and Algorithms 2023/24

Instituto Superior Técnico

Quiz – October 9, 2023

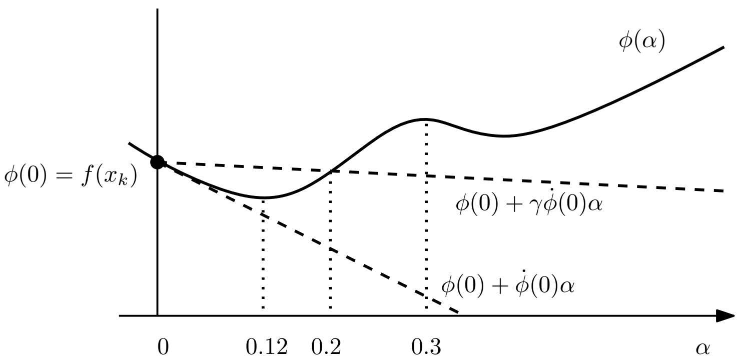
Student ID: ____________ **Name:** ____

- Quiz duration 45 minutes
- Mark the answers to multiple-choice questions in the table below. For a question worth C points with n options to choose from, you will lose $C/(n-1)$ points if you answer it incorrectly. There is no penalty for leaving it blank
- At the end of the quiz you should turn in the table of answers to multiple-choice questions, making sure that you fill out your **name and student ID**

Answers to multiple-choice questions

	a	b	c	d	e	f	g	h	i
Q1									
Q2									
Q3									
Q4									
Q5									
Q6									

Q 1. [3 val] The following picture shows a slice of a cost function $f(x)$ taken along a descent direction d_k in a line search algorithm, i.e., from the current iterate x_k it shows the function $\phi(\alpha) = f(x_k + \alpha d_k)$, with $\alpha \geq 0$. A backtracking subroutine using initial step $\hat{\alpha} = 1$, slope scaling factor $\gamma = 10^{-4}$, and step contraction factor $\beta = 0.6$ is invoked to select a suitable step α based on ϕ. Starting with $\hat{\alpha} = 1$ in iteration 1, in which iteration will a valid α be obtained?



a) 2 **b)** 3 **c)** 4 **d)** 5 **e)** 6 **f)** 7

Q 2. [3.5 val] The copy center at IST needs to buy new photocopiers and is considering two options. Option H is a high-speed model that can handle 20k copies a day and costs 6 k€. Option M is a medium-speed model that handles up to 10k copies a day and costs only 4 k€. At least one H and four other H or M machines are needed for parallel operation. In total, they should be able to handle at least 75k copies a day. If any M copiers are bought, the maintenance contract for *all* of them costs 15 k€ over their lifetime. Maintenance costs for H are similarly flat, but can be neglected here because they will be included in any valid configuration, as specified above. Choose a formulation to find the minimum-cost mix of copiers.

a)
$$\begin{aligned} \underset{h,m,y}{\text{minimize}} \quad & 6m + 4h + 15y \\ \text{subject to} \quad & 20h + 10m \geq 75, \quad h + m \geq 5, \quad h \geq 1, \quad y \leq m + 2, \\ & h, m \in \mathbb{Z}, \quad y \in \{0,\, 1\} \end{aligned} \tag{1}$$

b)
$$\begin{aligned} \underset{h,m}{\text{minimize}} \quad & 6h + 4m + 15 \\ \text{subject to} \quad & 20h + 10m \geq 75, \quad h + m \geq 5, \quad h \geq 1, \quad m \geq 0, \\ & h, m \in \mathbb{Z} \end{aligned} \tag{2}$$

c)
$$\begin{aligned} \underset{h,m,y}{\text{minimize}} \quad & 20h + 10m \\ \text{subject to} \quad & 6h + 4m + 15y \geq 75, \quad (h+m)y \geq 5, \quad h, m \geq 0, \\ & h, m, y \in \mathbb{Z} \end{aligned} \tag{3}$$

d)
$$\begin{aligned} \underset{h,m,y}{\text{minimize}} \quad & 6h + 4m + 15y \\ \text{subject to} \quad & 20h + 10m \geq 75, \quad h + m \geq 5, \quad h \geq 1, \quad m \geq 0, \quad y \geq \frac{m}{10}, \\ & h, m \in \mathbb{Z}, \quad y \in \{0,\, 1\} \end{aligned} \tag{4}$$

Q 3. [3.5 val] In robust regression problems we adopt an l_1 norm for the data fidelity term in the cost function instead of the usual Euclidean (l_2) norm to limit the impact of outliers that may be present in the observations. Given a set of observations in y, consider the specific regularized cost function for minimization over x given by

$$\|Ax - y\|_1 + \lambda \|x\|_1$$

with

$$A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \lambda = 4$$

$x_1 - 2x_2 - 1 + 4(x_1 + x_2)$
$+ x_2 + 1$

Find A' and p when this is written as a single norm $\|A'x - y'\|_p$.

a) $A' = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 4 & 0 \\ 0 & 4 \end{bmatrix}, p = 1$ **b)** $A' = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 3 & 0 \\ 0 & 3 \end{bmatrix}, p = 2$ **c)** $A' = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 9 & 0 \\ 0 & 9 \end{bmatrix}, p = 1$

d) $A' = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 9 & 9 \end{bmatrix}, p = 1$ **e)** $A' = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 4 & 4 \end{bmatrix}, p = 1$ **f)** $A' = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 2 & 0 \\ 0 & 2 \end{bmatrix}, p = 2$

Q 4. [3.5 val] We observe a set of T consecutive time-domain samples of a signal stored in vector $y \in \mathbb{R}^T$. These observations are modeled as $y = x + n$, where $x \in \mathbb{R}^T$ is the signal of interest and $n \in \mathbb{R}^T$ is noise. We know that x is a piecewise linear signal. Also, the magnitude of its increments from one time instant to the next one usually does not exceed 1. Finally, we know that the noise vector n is near zero. Choose a suitable formulation to estimate x from y.

a)

$$\underset{x}{\text{minimize}} \quad \|y-x\|_2^2 + \lambda \sum_{t=3}^{T} |x_t - 2x_{t-1} + x_{t-2}| + \mu \sum_{t=2}^{T} (-(x_t - x_{t-1}) + 1)_+ \tag{5}$$

b)

$$\underset{x}{\text{minimize}} \quad \|x\|_2^2 + \lambda \sum_{t=2}^{T} |x_t - x_{t-1}| + \mu \sum_{t=2}^{T} (|x_t - x_{t-1}| - 1)^2 \tag{6}$$

c)

$$\underset{x}{\text{minimize}} \quad \|y-x\|_2^2 + \lambda \sum_{t=3}^{T} |x_t - 2x_{t-1} + x_{t-2}| + \mu \sum_{t=2}^{T} (|x_t - x_{t-1}| - 1)_+ \tag{7}$$

d)

$$\underset{x}{\text{minimize}} \quad \|y-x\|_2^2 + \lambda \sum_{t=3}^{T} |x_t - 2x_{t-1} + x_{t-2}| + \mu \sum_{t=2}^{T} (-|x_t - x_{t-1}| + 1)_+ \tag{8}$$

Q 5. [3.5 val] At the k-th iteration of the Levenberg-Marquardt algorithm the residual $r(x_k)$ and its Jacobian $J_r(x_k)$ are given by

$$r(x_k) = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \qquad J_r(x_k) = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$$

Determine the line search direction d_k to find x_{k+1} for regularization factor $\lambda_k = 0$.

a) $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ **b)** $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ **c)** $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ **d)** $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ **e)** $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ **f)** $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Q 6. [3 val] We wish to select between the gradient method or the Newton method for unconstrained minimization of a cost function. The initial point x_0 and the corresponding function value, gradient, and Hessian are given below

$$x_0 = \begin{bmatrix} 3 \\ \frac{1}{2} \end{bmatrix}, \quad f(x_0) = \sqrt{2}, \quad \nabla f(x_0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}, \quad \nabla^2 f(x_0) = Q \Lambda Q^T, \quad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

Assuming that the outlook at x_0 is representative of how the minimization will evolve in subsequent iterations, choose among the different values for diagonal matrix Λ given below the one for which you expect the Newton method to provide the greatest advantage in convergence speed over the gradient method.

a) $\begin{bmatrix} 0.1 & 0 \\ 0 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 0 \\ 0 & 10^3 \end{bmatrix}$ **c)** $\begin{bmatrix} 0.1 & 0 \\ 0 & 10 \end{bmatrix}$

d) $\begin{bmatrix} \sqrt{10} & 0 \\ 0 & \frac{1}{\sqrt{10}} \end{bmatrix}$ **e)** $\begin{bmatrix} 10^3 & 0 \\ 0 & 10^3 \end{bmatrix}$ **f)** $\begin{bmatrix} 10^{-4} & 0 \\ 0 & 1 \end{bmatrix}$

Optimization and Algorithms 2023/24

Instituto Superior Técnico

Answers to the quiz – October 9, 2023

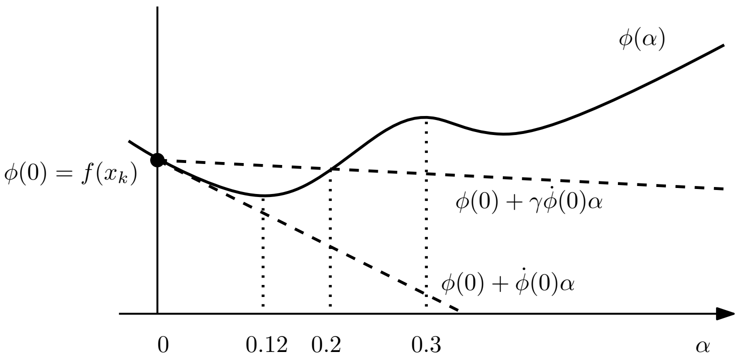
- Answers are only schematically presented, as they are meant to complement the students' own solutions to the problems

---

Answers to multiple-choice questions

	a	b	c	d	e	f	g	h	i
Q1				✓					
Q2				✓					
Q3	✓								
Q4			✓						
Q5					✓				
Q6						✓			

Q 1. The following picture shows a slice of a cost function $f(x)$ taken along a descent direction d_k in a line search algorithm, i.e., from the current iterate x_k it shows the function $\phi(\alpha) = f(x_k + \alpha d_k)$, with $\alpha \geq 0$. A backtracking subroutine using initial step $\hat{\alpha} = 1$, slope scaling factor $\gamma = 10^{-4}$, and step contraction factor $\beta = 0.6$ is invoked to select a suitable step α based on ϕ. Starting with $\hat{\alpha} = 1$ in iteration 1, in which iteration will a valid α be obtained?



a) 2 **b)** 3 **c)** 4 **d)** 5 **e)** 6 **f)** 7

Note: Start with $\alpha = 1$ in iteration 1, $\alpha = \beta$ in iteration 2, $\alpha = \beta^2$ in iteration 3, and so on. Backtracking will stop when $\phi(\alpha) < \phi(0) + \gamma\dot{\phi}(0)\alpha$, i.e., $\alpha < 0.2$, which occurs in iteration 5.

Q 2. The copy center at IST needs to buy new photocopiers and is considering two options. Option H is a high-speed model that can handle 20k copies a day and costs 6 k€. Option M is a medium-speed model that handles up to 10k copies a day and costs only 4 k€. At least one H and four other H or M machines are needed for parallel operation. In total, they should be able to handle at least 75k copies a day. If any M copiers are bought, the maintenance contract for *all* of them costs 15 k€ over their lifetime. Maintenance costs for H are similarly flat, but can be neglected here because they will be included in any valid configuration, as specified above. Choose a formulation to find the minimum-cost mix of copiers.

a)

$$\begin{aligned} \underset{h,m,y}{\text{minimize}} \quad & 6m + 4h + 15y \\ \text{subject to} \quad & 20h + 10m \geq 75, \quad h + m \geq 5, \quad h \geq 1, \quad y \leq m + 2, \\ & h, m \in \mathbb{Z}, \quad y \in \{0, 1\} \end{aligned} \tag{1}$$

b)

$$\begin{aligned} \underset{h,m}{\text{minimize}} \quad & 6h + 4m + 15 \\ \text{subject to} \quad & 20h + 10m \geq 75, \quad h + m \geq 5, \quad h \geq 1, \quad m \geq 0, \\ & h, m \in \mathbb{Z} \end{aligned} \tag{2}$$

c)

$$\begin{aligned} \underset{h,m,y}{\text{minimize}} \quad & 20h + 10m \\ \text{subject to} \quad & 6h + 4m + 15y \geq 75, \quad (h + m)y \geq 5, \quad h, m \geq 0, \\ & h, m, y \in \mathbb{Z} \end{aligned} \tag{3}$$

d) ✓

$$\begin{aligned} \underset{h,m,y}{\text{minimize}} \quad & 6h + 4m + 15y \\ \text{subject to} \quad & 20h + 10m \geq 75, \quad h + m \geq 5, \quad h \geq 1, \quad m \geq 0, \quad y \geq \frac{m}{10}, \\ & h, m \in \mathbb{Z}, \quad y \in \{0, 1\} \end{aligned} \tag{4}$$

Note: For the specified daily number of copies a quick calculation shows that a single H (mandatory) and six M could handle the load. An optimal solution might have fewer M and more H than this, so we will surely have $m \leq 6$. When any M are bought we will then have $1 \leq m \leq 6$, and so the constraint $y \geq \frac{m}{10}$ for binary y equivalently means $y = 1$. This "activates" the term in the cost function that accounts for the maintenance contract of M. When $m = 0$ we can have $y = 0$, which inactivates that term.

Q 3. In robust regression problems we adopt an l_1 norm for the data fidelity term in the cost function instead of the usual Euclidean (l_2) norm to limit the impact of outliers that may be present in the observations. Given a set of observations in y, consider the specific regularized cost function for minimization over x given by

$$\|Ax - y\|_1 + \lambda \|x\|_1$$

with

$$A = \begin{bmatrix} 1 & -2 \\ 0 & 1 \end{bmatrix}, \qquad y = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \qquad \lambda = 4$$

Find A' and p when this is written as a single norm $\|A'x - y'\|_p$.

a) $A' = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 4 & 0 \\ 0 & 4 \end{bmatrix}, p = 1$ b) $A' = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 3 & 0 \\ 0 & 3 \end{bmatrix}, p = 2$ c) $A' = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 9 & 0 \\ 0 & 9 \end{bmatrix}, p = 1$

d) $A' = \begin{bmatrix} -2 & 1 \\ 1 & 0 \\ 9 & 9 \end{bmatrix}, p = 1$ e) $A' = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 4 & 4 \end{bmatrix}, p = 1$ f) $A' = \begin{bmatrix} 1 & -2 \\ 0 & 1 \\ 2 & 0 \\ 0 & 2 \end{bmatrix}, p = 2$

Note: Because $\|u\|_1 = \sum_i |u_i|$, we readily have $\|u\|_1 + \|v\|_1 = \| \begin{bmatrix} u \\ v \end{bmatrix} \|_1$.

Q 4. We observe a set of T consecutive time-domain samples of a signal stored in vector $y \in \mathbb{R}^T$. These observations are modeled as $y = x + n$, where $x \in \mathbb{R}^T$ is the signal of interest and $n \in \mathbb{R}^T$ is noise. We know that x is a piecewise linear signal. Also, the magnitude of its increments from one time instant to the next one usually does not exceed 1. Finally, we know that the noise vector n is near zero. Choose a suitable formulation to estimate x from y.

a)

$$\underset{x}{\text{minimize}} \quad \|y - x\|_2^2 + \lambda \sum_{t=3}^{T} |x_t - 2x_{t-1} + x_{t-2}| + \mu \sum_{t=2}^{T} (-(x_t - x_{t-1}) + 1)_+ \tag{5}$$

b)

$$\underset{x}{\text{minimize}} \quad \|x\|_2^2 + \lambda \sum_{t=2}^{T} |x_t - x_{t-1}| + \mu \sum_{t=2}^{T} (|x_t - x_{t-1}| - 1)^2 \tag{6}$$

c) ✓

$$\underset{x}{\text{minimize}} \quad \|y - x\|_2^2 + \lambda \sum_{t=3}^{T} |x_t - 2x_{t-1} + x_{t-2}| + \mu \sum_{t=2}^{T} (|x_t - x_{t-1}| - 1)_+ \tag{7}$$

d)

$$\underset{x}{\text{minimize}} \quad \|y - x\|_2^2 + \lambda \sum_{t=3}^{T} |x_t - 2x_{t-1} + x_{t-2}| + \mu \sum_{t=2}^{T} (-|x_t - x_{t-1}| + 1)_+ \tag{8}$$

Q 5. At the k-th iteration of the Levenberg-Marquardt algorithm the residual $r(x_k)$ and its Jacobian $J_r(x_k)$ are given by

$$r(x_k) = \begin{bmatrix} -2 \\ -1 \end{bmatrix}, \qquad J_r(x_k) = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}$$

Determine the line search direction d_k to find x_{k+1} for regularization factor $\lambda_k = 0$.

a) $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ b) $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ c) $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ d) $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ e) $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ f) $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$

Note: For $\lambda_k = 0$ (this would be a limit case when $\lambda > 0$ becomes very small near the optimal point) the iteration of Levenberg-Marquardt solves $\min_x \|J_r(x_k)(x - x_k) + r(x_k)\|^2$. The line search direction d_k is the least-squares solution of the linear system $J_r(x_k)d_k = -r(x_k)$, which, in this case, is a plain linear system because the Jacobian $J_r(x_k)$ is square and invertible.

Q 6. We wish to select between the gradient method or the Newton method for unconstrained minimization of a cost function. The initial point x_0 and the corresponding function value, gradient, and Hessian are given below

$$x_0 = \begin{bmatrix} 3 \\ \frac{1}{2} \end{bmatrix}, \quad f(x_0) = \sqrt{2}, \quad \nabla f(x_0) = \begin{bmatrix} 5 \\ 4 \end{bmatrix}, \quad \nabla^2 f(x_0) = Q \Lambda Q^T, \quad Q = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

Assuming that the outlook at x_0 is representative of how the minimization will evolve in subsequent iterations, choose among the different values for diagonal matrix Λ given below the one for which you expect the Newton method to provide the greatest advantage in convergence speed over the gradient method.

a) $\begin{bmatrix} 0.1 & 0 \\ 0 & 1 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 0 \\ 0 & 10^3 \end{bmatrix}$ **c)** $\begin{bmatrix} 0.1 & 0 \\ 0 & 10 \end{bmatrix}$

d) $\begin{bmatrix} \sqrt{10} & 0 \\ 0 & \frac{1}{\sqrt{10}} \end{bmatrix}$ **e)** $\begin{bmatrix} 10^3 & 0 \\ 0 & 10^3 \end{bmatrix}$ **f)** $\begin{bmatrix} 10^{-4} & 0 \\ 0 & 1 \end{bmatrix}$

Note: The performance gap between the Newton method and the gradient method widens as the condition number of the Hessian increases. The condition number is the ratio between the largest and smallest values in the diagonal of Λ.