ESE415 OPTIMIZATION

Tutorial 2

Xiaojian Xu

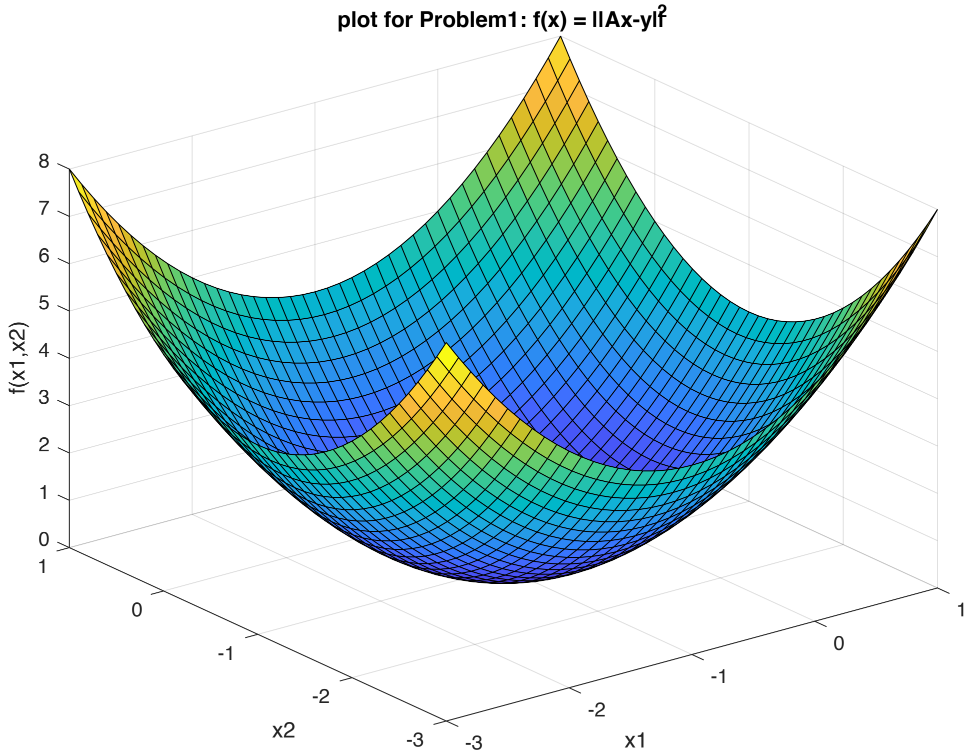
02/08/2019

Guideline

HW1 RECAP

HW2 RECITATION

HW1 Review (Problem 5)



Problem 5

Consider the following objective function

$$f(\boldsymbol{x}) = \frac{1}{2}\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|^2 \quad \text{where} \quad \boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \boldsymbol{y} = \begin{bmatrix} -2 \\ 0 \end{bmatrix} \quad \text{with} \quad \boldsymbol{x} \in \mathbb{R}^2.$$

(a) Find all the minimizers of f analytically. Are they local or global minimizers?

(b) Download and modify the Matlab script `GradientMethod.m` to minimize f with the gradient method. Set the parameters of the algorithm to $\boldsymbol{x}^0 = \texttt{xitInit} = (1, 5)$, $\gamma = \texttt{stepSize} = 0.1$, $\texttt{maxIter} = 100$, and $\texttt{tol} = 10^{-6}$. Submit the printout of your code with you responses.

(c) For each of the following values of the step-size parameter $\gamma > 0$ comment on the convergence of the algorithm: (*i*) $\gamma = 0.1$, (*ii*) $\gamma = 0.4$, (*iii*) $\gamma = 0.99$, and (*iv*) $\gamma = 1.0001$. Submit the generated plots with your responses.

(d) **(Bonus)** By using the update equation of the gradient method for this problem, justify analytically the behavior of the method for the step-sizes in (c). Is it possible to make the gradient method converge in a single iteration for this problem?

HW1. P5 Solutions & Explanations

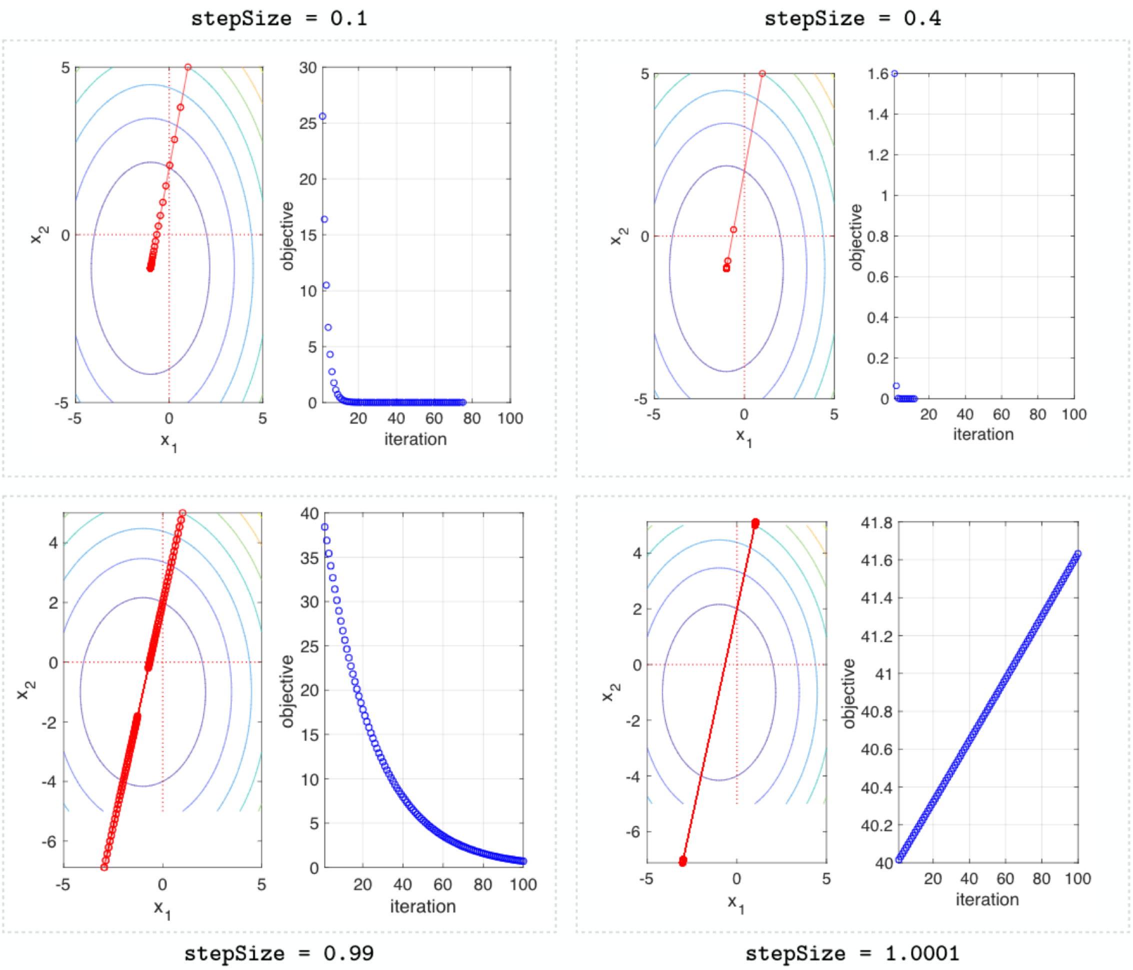
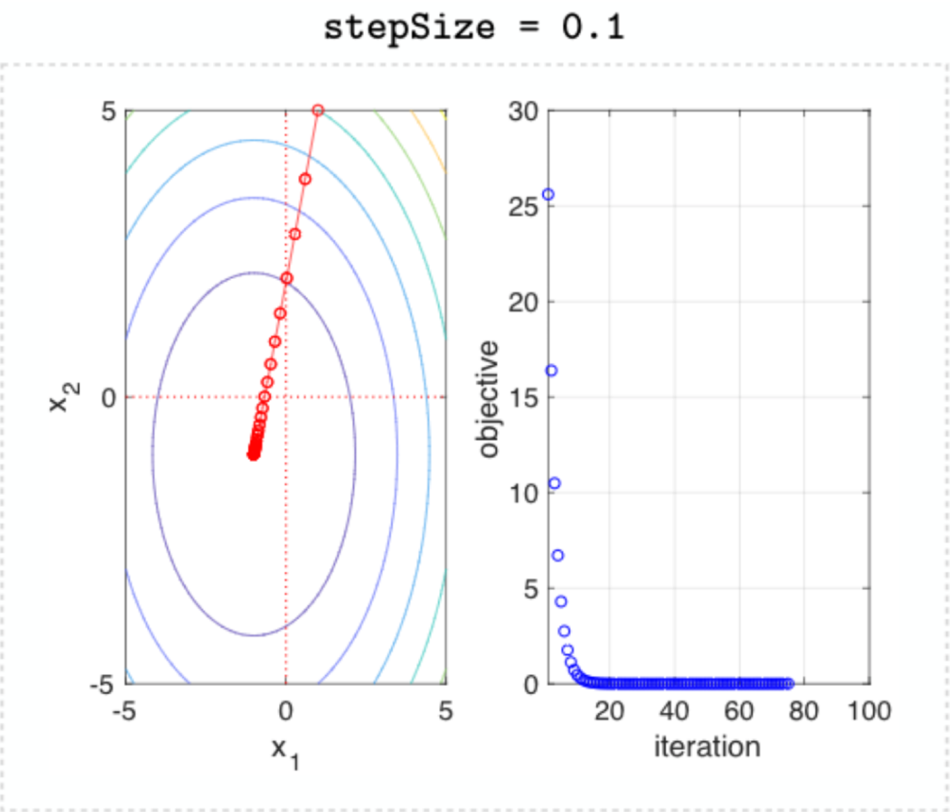
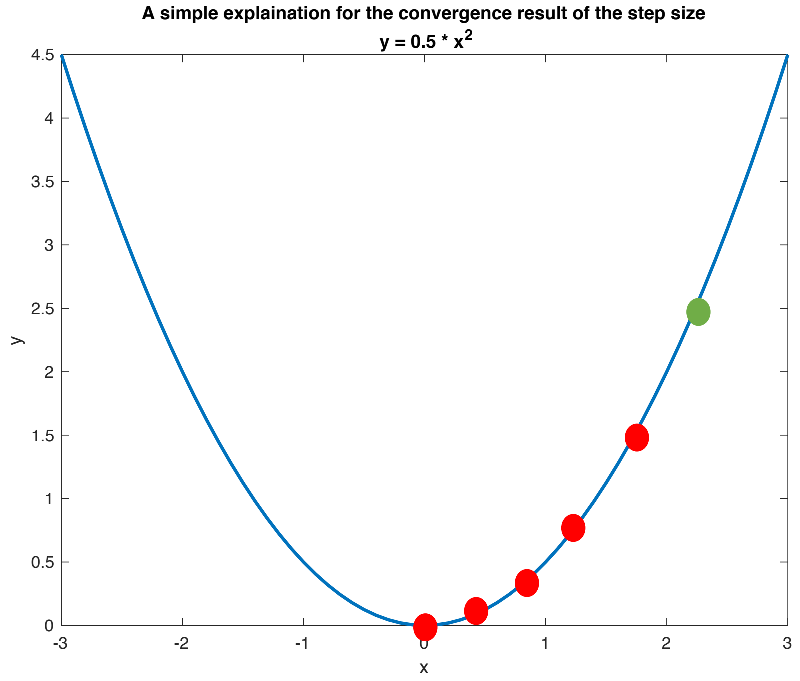


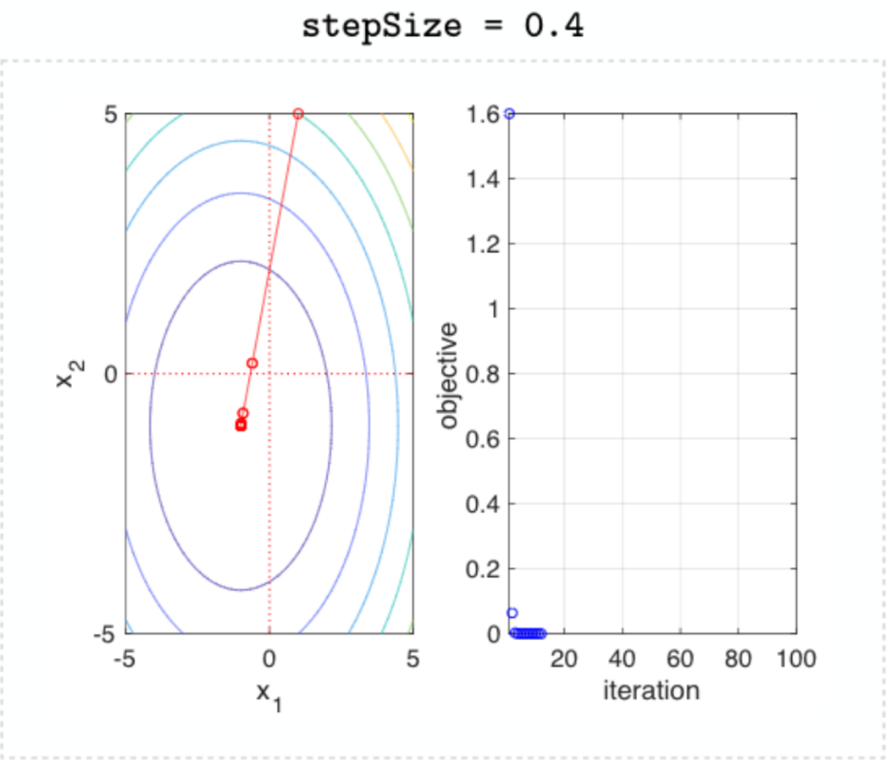
Figure 2: Illustration of the convergence the gradient method for different values of the parameter γ.

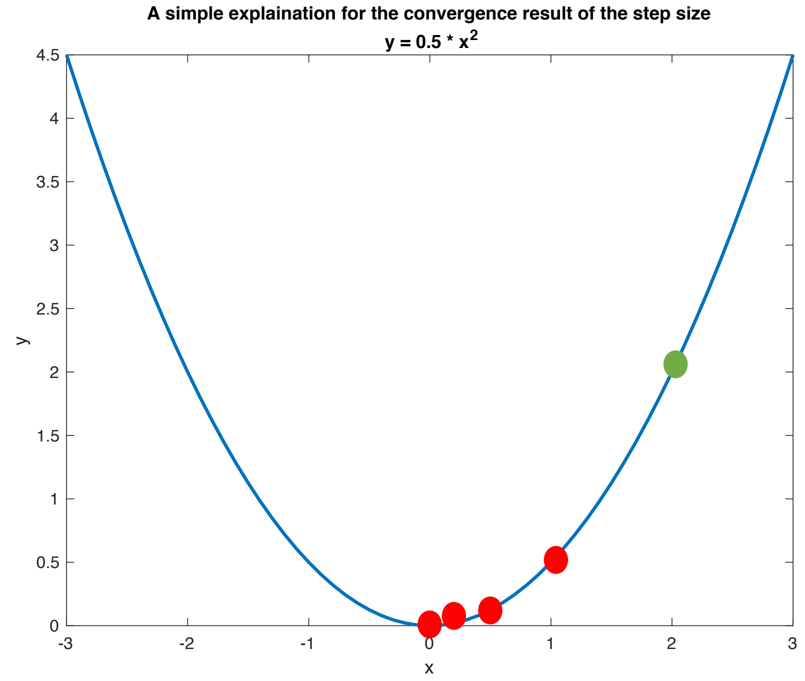




HW1. P5
Solutions & Explanations

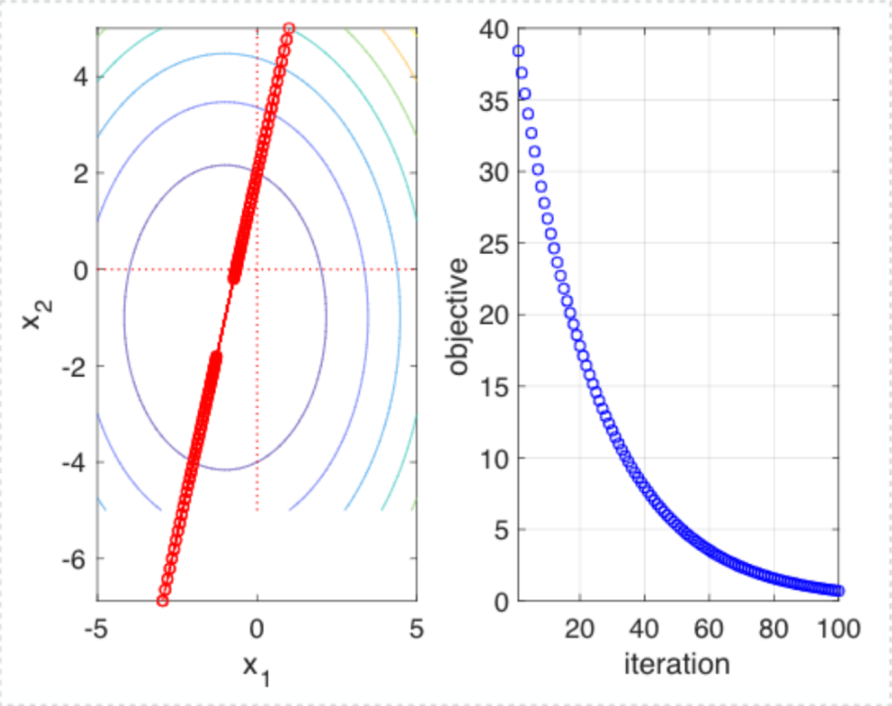
https://people.maths.ox.ac.uk/hauser/hauser_lecture2.pdf



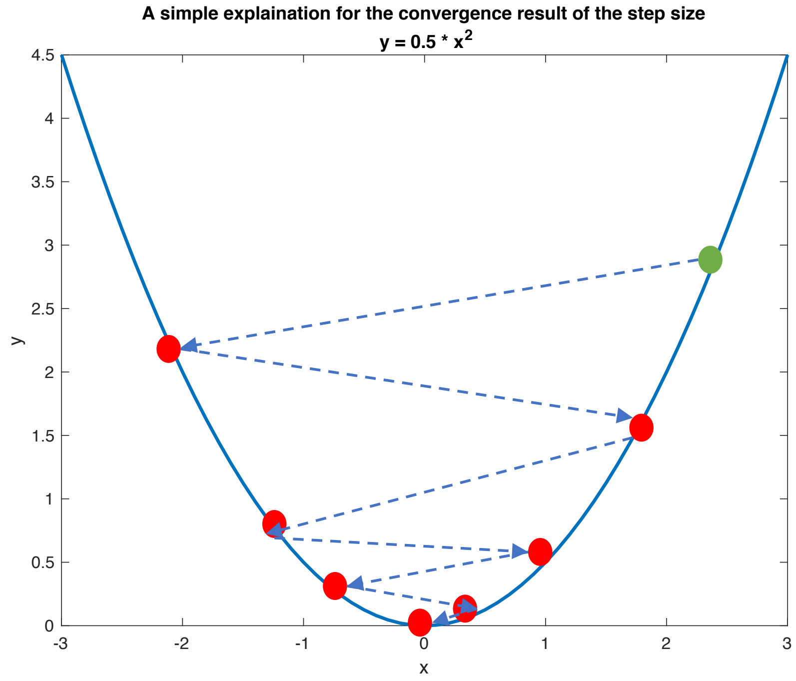


HW1. P5
Solutions & Explanations

https://people.maths.ox.ac.uk/hauser/hauser_lecture2.pdf

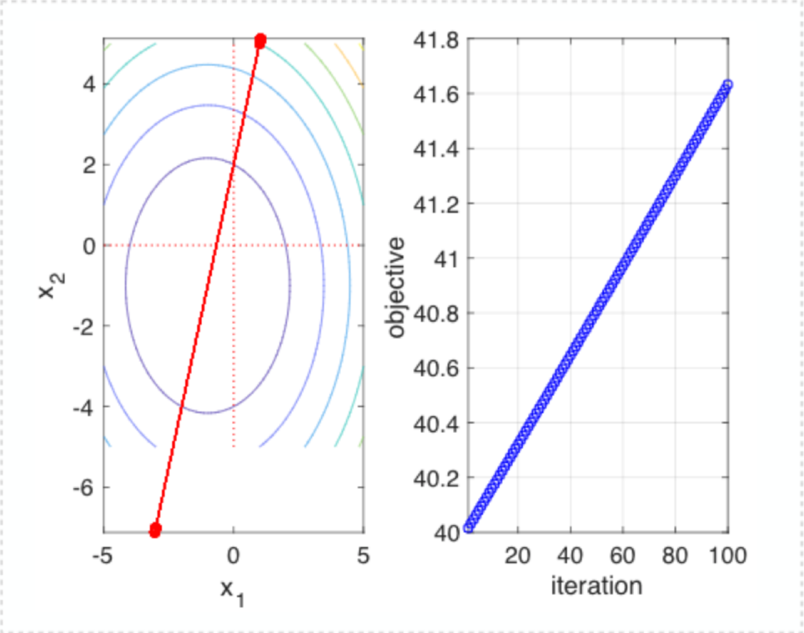


stepSize = 0.99

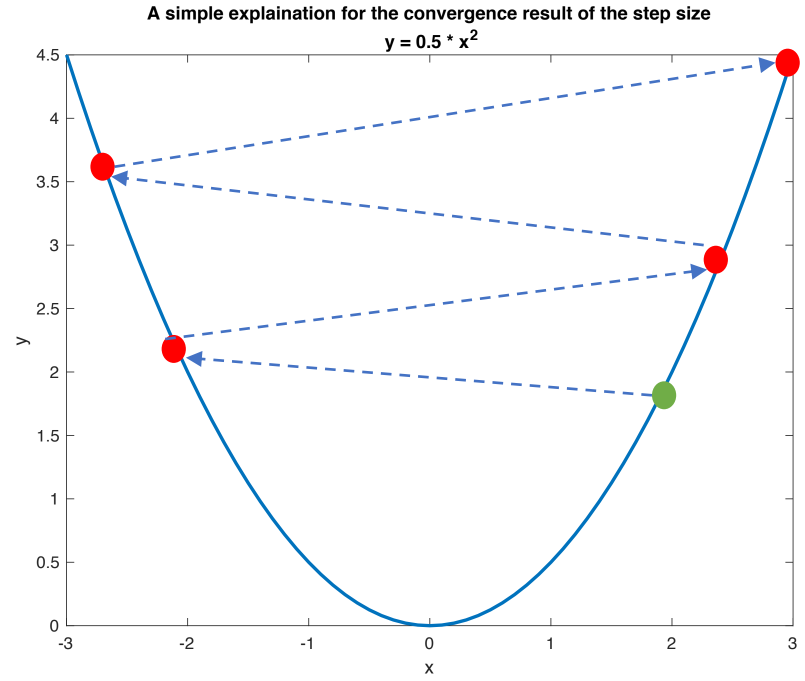


HW1. P5
Solutions & Explanations

https://people.maths.ox.ac.uk/hauser/hauser_lecture2.pdf



stepSize = 1.0001



HW1. P5
Solutions & Explanations

https://people.maths.ox.ac.uk/hauser/hauser_lecture2.pdf

HW1. P5 Solutions & Explanations

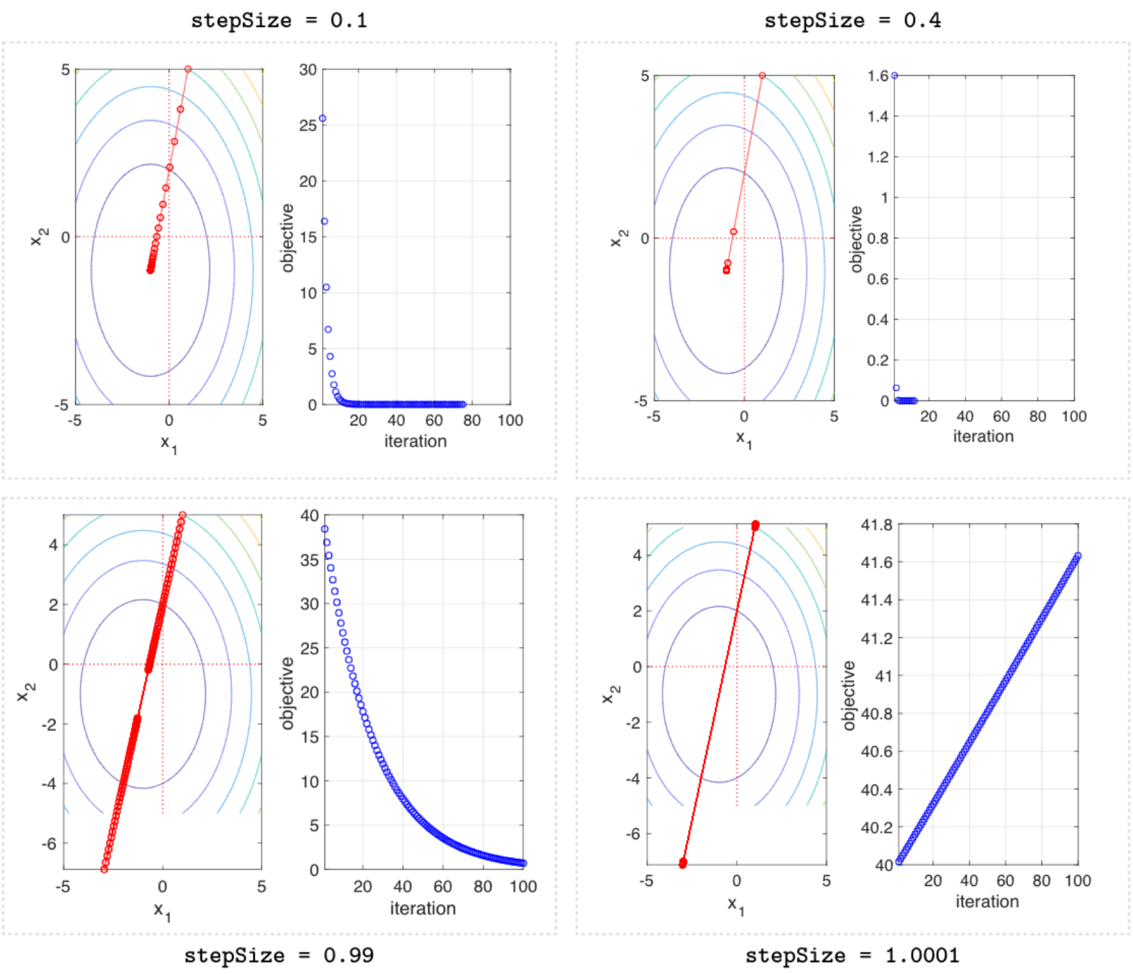


Figure 2: Illustration of the convergence the gradient method for different values of the parameter γ.

HW2 Recitation (P1)

Consider the function $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_{\ell_1} = \sum_{i=1}^{n} |x_i|$ where $\boldsymbol{x} \in \mathbb{R}^n$. For all the answers below explain your reasoning.

a) Is f coercive?

b) Is f a norm on $\mathbb{R}^n$?

c) Is f convex on $\mathbb{R}^n$?

d) Consider another function $g(\boldsymbol{x}) = \|\boldsymbol{x}\|_{\mathsf{n}}$, where $\|\cdot\|_{\mathsf{n}}$ is an arbitrary norm. Is g convex?

Hints

Definition [edit]

Given a vector space V over a subfield F of the complex numbers, a **norm** on V is a nonnegative-valued scalar function p: $V \to [0,+\infty)$ with the following properties:[1]

For all $a \in F$ and all $\mathbf{u}, \mathbf{v} \in V$,

1. $p(\mathbf{u} + \mathbf{v}) \leq p(\mathbf{u}) + p(\mathbf{v})$ (being *subadditive* or satisfying the *triangle inequality*).
2. $p(a\mathbf{v}) = |a|\, p(\mathbf{v})$ (being *absolutely homogeneous* or *absolutely scalable*).
3. If $p(\mathbf{v}) = 0$ then $\mathbf{v}=\mathbf{0}$ is the zero vector (being *positive definite* or being *point-separating*).

Solution (P1)

a) Yes. One way to see this is to note that $\|\boldsymbol{x}\|_{\ell_1} \geq \|\boldsymbol{x}\|_{\ell_2}$. As a side note (not necessarily for grading), this can be shown as

$$\|\boldsymbol{x}\|_{\ell_2}^2 = \sum_{i=1}^{n} |x_i|^2 \leq \left(\sum_{i=1}^{n} |x_i|^2 + 2 \sum_{i,j,i \neq j} |x_i||x_j| \right) = \|\boldsymbol{x}\|_{\ell_1}^2.$$

b) We check the properties of the norm:

- *absolute homogeneity*:

$$f(\alpha \boldsymbol{x}) = \sum_{i=1}^{n} |\alpha x_i| = |\alpha| \sum_{i=1}^{n} |x_i| = |\alpha| f(\boldsymbol{x}).$$

- *nonnegativity*: It is clear that $f(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x}$, since it is the sum of absolute values. Additionally, $f(\boldsymbol{x}) = 0$, means that each component of the sum must be zero, which implies that $\boldsymbol{x} = \mathbf{0}$.
- *triangular inequality*: Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$f(\boldsymbol{x} + \boldsymbol{y}) = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} (|x_i| + |y_i|) = f(\boldsymbol{x}) + f(\boldsymbol{y}).$$

c) Consider $\theta \in [0, 1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, then

$$f(\theta \boldsymbol{x} + (1 - \theta \boldsymbol{y})) \leq f(\theta \boldsymbol{x}) + f((1 - \theta)\boldsymbol{y}) = |\theta| f(\boldsymbol{x}) + |(1 - \theta)| f(\boldsymbol{y}) = \theta f(\boldsymbol{x}) + (1 - \theta) f(\boldsymbol{y}),$$

where in the first inequality we used the triangular inequality, in the first equality we used the absolute homogeneity, and in the second equality we used $\theta \in [0, 1]$.

d) Since any norm will satisfy the triangular inequality and absolute homogeneity, g is convex.

HW2 Recitation (P3)

Prove the following results.

a) Let $f \in \mathcal{F}^1(\mathbb{R}^n)$. Show that if $\nabla f(\boldsymbol{x}^*) = 0$, then $\boldsymbol{x}^*$ is a global minimizer of f.

b) Let $f \in \mathcal{F}^1(\mathbb{R}^m)$, and $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$. Show that $\varphi(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}) \in \mathcal{F}^1(\mathbb{R}^n)$.

c) Let $f_1, \ldots, f_K \in \mathcal{F}^1(\mathbb{R}^n)$. Show that $f(\boldsymbol{x}) = \max_{k \in [1,\ldots,K]}\{f_k(\boldsymbol{x})\}$ is convex.

Hint

Remark. We introduce a set of shorthand notations. First, we will denote with $\mathcal{F}^k(C)$ the class of functions that are both convex and k times continuously differentiable on a convex set $C \subseteq \mathbb{R}^n$. We will denote with $\mathcal{F}_L^k(C)$ the class of functions that additionally have Lipschitz continuous derivatives.

Solution (P2)

a) Let $\boldsymbol{x}^*$ be a critical point. Since $f \in \mathcal{F}^1(\mathbb{R}^n)$, we have for all $\boldsymbol{x} \in \mathbb{R}^n$

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}^*) + \nabla f(\boldsymbol{x}^*)^\mathsf{T}(\boldsymbol{x} - \boldsymbol{x}^*) = f(\boldsymbol{x}^*),$$

where in the first inequality we used the convexity and the second the fact that $\nabla f(\boldsymbol{x}^*) = 0$.

b) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Denote $\bar{\boldsymbol{x}} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ and $\bar{\boldsymbol{y}} = \boldsymbol{A}\boldsymbol{y}+\boldsymbol{b}$. Since $\nabla \varphi(\boldsymbol{x}) = \boldsymbol{A}^\mathsf{T}\nabla f(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}) = \boldsymbol{A}^\mathsf{T}\nabla f(\bar{\boldsymbol{x}})$, we have

$$\begin{aligned}
\varphi(\boldsymbol{y}) = f(\bar{\boldsymbol{y}}) &\geq f(\bar{\boldsymbol{x}}) + \nabla f(\bar{\boldsymbol{x}})^\mathsf{T}(\bar{\boldsymbol{y}} - \bar{\boldsymbol{x}}) \\
&= \varphi(\boldsymbol{x}) + \nabla f(\bar{\boldsymbol{x}})^\mathsf{T}\boldsymbol{A}(\boldsymbol{y} - \boldsymbol{x}) \\
&= \varphi(\boldsymbol{x}) + [\boldsymbol{A}^\mathsf{T}\nabla f(\bar{\boldsymbol{x}})]^\mathsf{T}(\boldsymbol{y} - \boldsymbol{x}) \\
&= \varphi(\boldsymbol{x}) + \nabla \varphi(\boldsymbol{x})^\mathsf{T}(\boldsymbol{y} - \boldsymbol{x}).
\end{aligned}$$

c) For any $\theta \in [0, 1]$ and $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$

$$\begin{aligned}
f(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y}) &= \max_{k\in[1,\ldots,K]} \{f_k(\theta\boldsymbol{x} + (1-\theta)\boldsymbol{y})\} \\
&\leq \max_{k\in[1,\ldots,K]} \{\theta f_k(\boldsymbol{x}) + (1-\theta) f_k(\boldsymbol{y})\} \\
&\leq \theta \max_{k\in[1,\ldots,K]} \{f_k(\boldsymbol{x})\} + (1-\theta) \max_{k\in[1,\ldots,K]} \{f_k(\boldsymbol{y})\} \\
&= \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}).
\end{aligned}$$

Thanks