

# Apresentação do Projeto

## Optimização e Algoritmos

Instituto Superior Técnico

Grupo: 28

Francisco Melo, 84053

Rodrigo Rego, 89213

18 de dezembro de 2018

## 1 Parte 1

- Formulação do Problema
- Regularizador  $\ell_2^2$
- Regularizador  $\ell_2$
- Regularizador  $\ell_1$
- Formulação com discos -  $\ell_2$
- Passagem exata pelos pontos
- Mudança de Pesos Iterativa

## 2 Parte 2

- Regressão Logística - Método do Gradiente
- Regressão Logística - Método de Newton
- Localização em rede - Levenberg-Marquardt

## 3 Parte 3

- KKT - Tarefa 1
- KKT - Tarefa 2

# Formulação do Problema

## Principais objetivos

- 1 **Navegação:** de uma posição inicial para uma final tendo em conta a dinâmica do robot;
- 2 **Controlo limitado:** forças aplicadas ao motor são menores que  $U_{máx.}$ ;
- 3 **Waypoints:** posição do robot em instantes  $\tau_k$  deve estar próxima dos *waypoints*;
- 4 **Controlo simples:** sinal de controlo com poucas mudanças ao longo do tempo.

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K \|Ex(\tau_k) - w_k\|_2^2 + \lambda \sum_{t=1}^{T-1} \|u(t) - u(t-1)\|_2^2 && (1) \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\max}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned}$$

# Regularizador $\ell_2^2$ - Resultados (1)

$\lambda$	Mudanças do Sinal	Desvio Médio
$10^{-3}$	79	0.1257
$10^{-2}$	79	0.8242
$10^{-1}$	79	2.1958
1	79	3.6826
10	79	5.6317
100	79	10.9042
1000	79	15.3304

**Tabela:** Mudanças do sinal de controlo e desvio médio com  $\ell_2^2$ .

# Regularizador $\ell_2^2$ - Resultados (2)

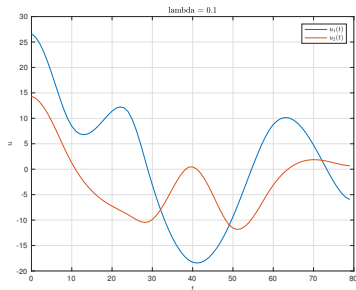
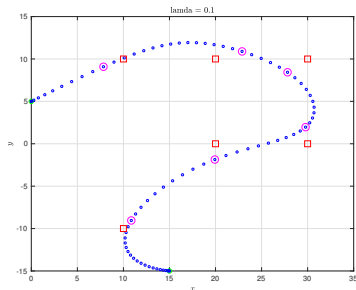


Figura: Caso  $\lambda = 10^{-1}$ .

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K \|Ex(\tau_k) - w_k\|_2^2 + \lambda \sum_{t=1}^{T-1} \|u(t) - u(t-1)\|_2 && (2) \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\text{max}}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned}$$

# Regularizador $\ell_2$ - Resultados (1)

$\lambda$	Mudanças do Sinal	Desvio Médio
$10^{-3}$	10	0.0075
$10^{-2}$	8	0.0747
$10^{-1}$	11	0.7021
1	4	2.8876
10	4	5.3689
100	4	12.5914
1000	1	16.2266

**Tabela:** Mudanças do sinal de controlo e desvio médio com  $\ell_2$ .



# Regularizador $\ell_2$ - Resultados (2)

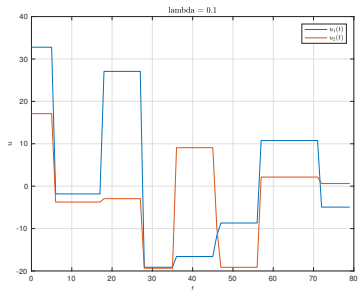
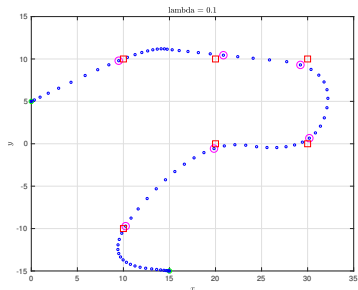


Figura: Caso  $\lambda = 10^{-1}$ .

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K \|Ex(\tau_k) - w_k\|_2^2 + \lambda \sum_{t=1}^{T-1} \|u(t) - u(t-1)\|_1 && (3) \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\max}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned}$$

# Regularizador $\ell_1$ - Resultados (1)

$\lambda$	Mudanças do Sinal	Desvio Médio
$10^{-3}$	13	0.0107
$10^{-2}$	12	0.1055
$10^{-1}$	14	0.8863
1	11	2.8732
10	5	5.4362
100	3	13.0273
1000	2	16.0463

**Tabela:** Mudanças do sinal de controlo e desvio médio com  $\ell_1$ .

# Regularizador $\ell_1$ - Resultados (2)

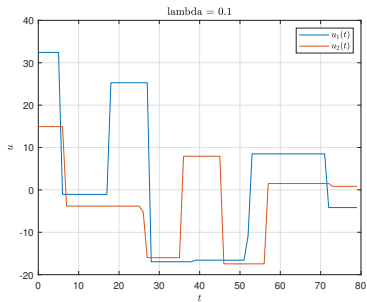
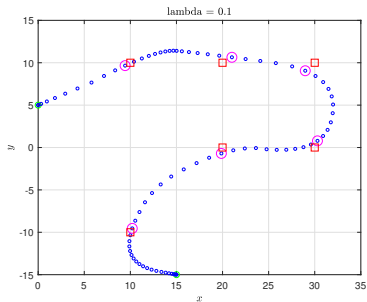


Figura: Caso  $\lambda = 10^{-1}$ .

# Conclusões - Impacto do parâmetro $\lambda$ (1)

## Trajétória

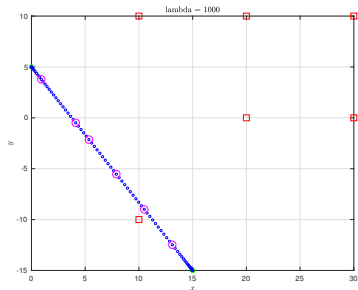
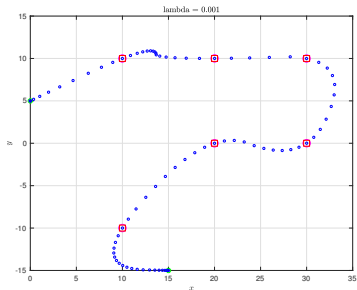


Figura: Evolução da trajetória para valores extremos de  $\lambda$  ( $\ell_2$ ).

# Conclusões - Impacto do parâmetro $\lambda$ (2)

## Sinal de Controle

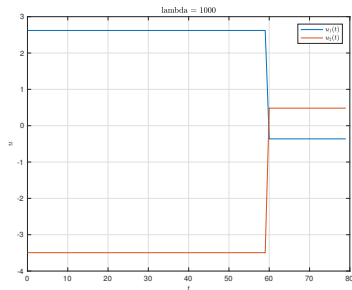
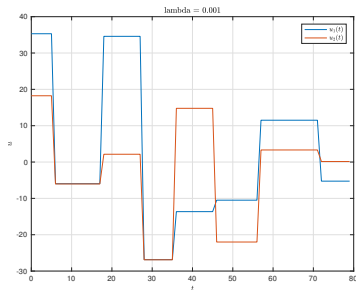


Figura: Evolução do sinal de controlo para valores extremos de  $\lambda$  ( $\ell_2$ ).

## Conclusões - Regularizador (3)

$$\|d\|_2^2 = \sum_{k=1}^K d_k^2 \quad \|d\|_2 = \sqrt{\sum_{k=1}^K d_k^2} \quad \|d\|_1 = \sum_{k=1}^K |d_k| \quad (4)$$

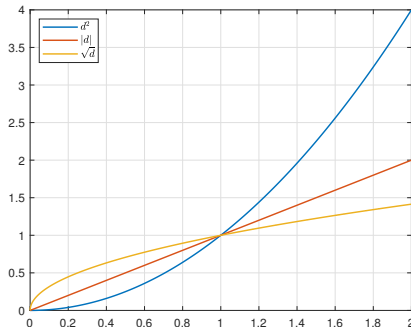


Figura: Operações dos diferentes regularizadores.

# Conclusões - Regularizador (4)

## Conclusões chave

- $\ell_2^2$ : penaliza sobretudo grandes desvios e diminui o peso de pequenos desvios;
- $\ell_2$ : normaliza o peso de grandes e pequenos desvios (penalização mais uniformemente distribuída);
- $\ell_1$ : considera de uma forma linear o peso dos desvios na penalização

∴ Com  $\ell_2$  obteve-se melhores resultados (sinal de controlo mais simples).



# Forma fechada de $d(p, D(c, r))$

$$d(p, D(c, r)) = \|p - c\|_2 - r \quad (5)$$

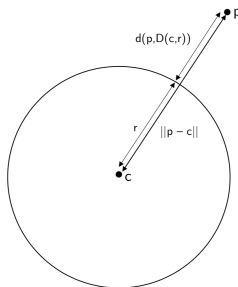


Figura: Figura auxiliar da distância mínima ao ponto  $p$ .

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K d(Ex(\tau_k), D(c_k, r_k))^2 + \lambda \sum_{t=1}^{T-1} \|u(t) - u(t-1)\|_2 \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\max}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned} \tag{6}$$

## $\ell_2$ Discos - Resultados (1)

$\lambda$	Mudanças de Sinal	Desvio Médio
$10^{-3}$	13	2.0055
$10^{-2}$	11	2.05414
$10^{-1}$	7	2.44833
1	5	3.51969
10	5	5.7842
100	2	13.2397
1000	2	16.227

**Tabela:** Mudanças do sinal de controlo e desvio médio com  $\ell_2$  e discos.

# $\ell_2$ Discos - Resultados (2)

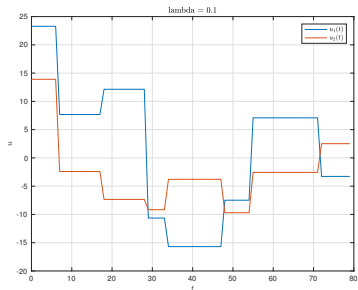
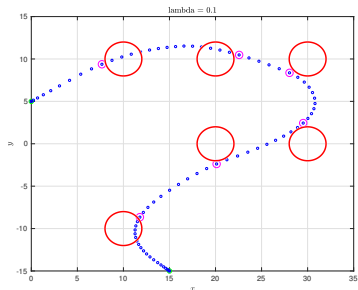


Figura: Caso  $\lambda = 10^{-1}$ .

## $\ell_2$ Discos

- Para todos os casos de  $\lambda$  a trajetória é mais suave;
- O número de mudanças do sinal de controlo é semelhante;
- O desvio médio aos *waypoints* é maior.

## $\ell_2$ Pontos

- Para valores de  $\lambda$  pequenos a trajetória é menos suave;
- O número de mudanças do sinal de controlo é semelhante;
- O desvio médio aos *waypoints* é inferior.

# Trajectoria - $\ell_2$ Discos vs. $\ell_2$ Pontos ( $\lambda = 10^{-3}$ )

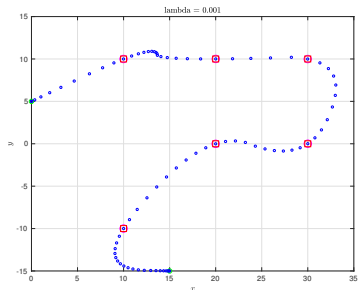
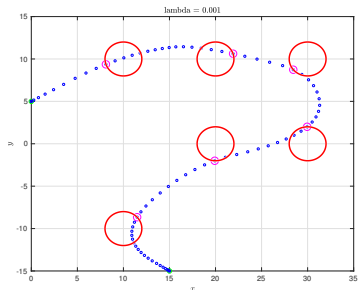


Figura: Caso  $\lambda = 10^{-3}$ .

Efeito da redução de  $U_{max} = 15$

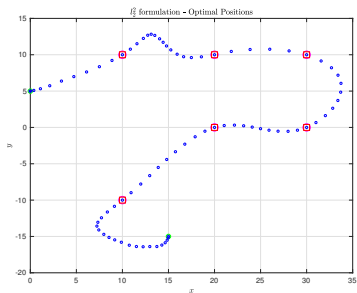
- A redução do valor desta restrição reduz o espaço de soluções;
- Para  $U_{max} = 15$  concluiu-se que não há solução que cumpra com todas as restrições do problema.

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K \|Ex(\tau_k) - w_k\|_2^2 \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\max}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned} \tag{7}$$

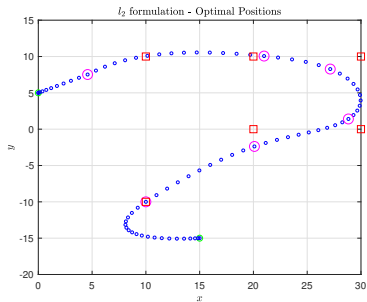


$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K \|Ex(\tau_k) - w_k\|_2 \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\max}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned} \tag{8}$$

# Passagem exata pelos pontos - $\ell_2^2$ vs. $\ell_2$ (1)



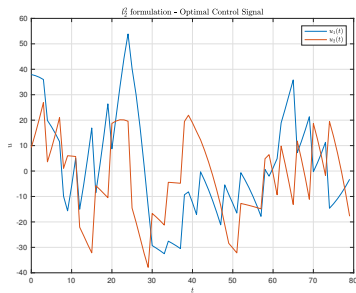
(a)



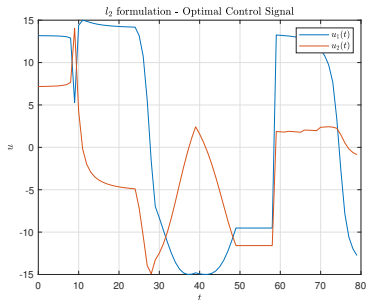
(b)

Figura: Comparação da trajetória: (a)  $\ell_2^2$  e (b)  $\ell_2$ .

# Passagem exata pelos pontos - $\ell_2^2$ vs. $\ell_2$ (2)



(a)



(b)

Figura: Comparação do sinal de controlo: (a)  $\ell_2^2$  e (b)  $\ell_2$ .

# Passagem exata pelos pontos - $\ell_2^2$ vs. $\ell_2$ (3)

Regularizador	Waypoints Capturados
$\ell_2^2$	6
$\ell_2$	1

**Tabela:** Número de pontos capturados.

$$\begin{aligned} & \underset{x,u}{\text{minimize}} && \sum_{k=1}^K \frac{1}{\|E_{X^{(m)}}(\tau_k) - w_k\|_2 + \epsilon} \|E_{X^{(m)}}(\tau_k) - w_k\|_2 \\ & \text{subject to} && x(0) = x_{\text{initial}} \\ & && x(T) = x_{\text{final}} \\ & && \|u(t)\|_2 \leq U_{\max}, \quad \text{for } 0 \leq t \leq T-1 \\ & && x(t+1) = Ax(t) + Bu(t), \quad \text{for } 0 \leq t \leq T-1. \end{aligned} \tag{9}$$

# Mudança de Pesos Iterativa - Resultados (1)

$m$	N.º de <i>waypoints</i> capturados
1	1
2	2
3	2
4	3
5	3
6	3
7	3
8	3
8	3
10	3

**Tabela:** N.º de pontos capturados com a mudança de pesos iterativa.

# Mudança de Pesos Iterativa - Resultados (2)

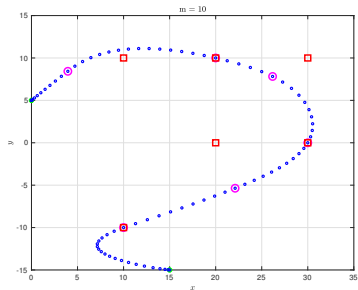
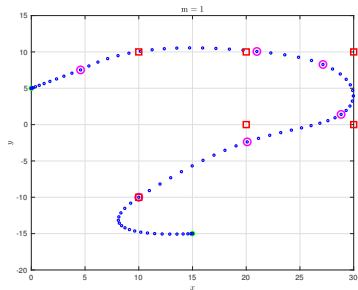


Figura: Comparação da trajetória para  $m = 1$  vs.  $m = 10$ .

# Mudança de Pesos Iterativa - Resultados (3)

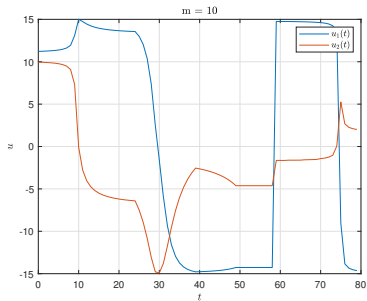
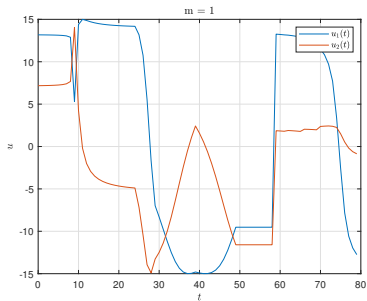


Figura: Comparação do sinal de controlo para  $m = 1$  vs.  $m = 10$ .



# Mudança de Pesos Iterativa - Comentários (3)

Peso:

$$\frac{1}{\|Ex^{(m)}(\tau_k) - w_k\|_2 + \epsilon} \quad (10)$$

- Quando a distância entre o robot (nos instantes  $\tau_k$ ) e os *waypoints* (mais  $\epsilon$ ) é menor que 1  $\implies$  pesos maiores.

Isto implica que sempre o que robot estiver perto de um *waypoint* a sua posição vai ser capturada pelo mesmo numa região de raio 1.

Quando o robot se encontra distante, os pesos são pequenos e atribui-se menos penalização à passagem por *waypoints*, considerando-se mais a chegada à posição final.

# Método do Gradiente

- Algoritmo *line search*, iterações evoluem segundo:

$$x_{k+1} = x_k + \alpha_k d_k \quad (11)$$

- Onde  $d_k = -\nabla f(x_k)$  :  $\nabla f(x_k)^T d_k = -\|\nabla f(x_k)\|^2 < 0$
- $\alpha_k > 0$ , calculado pela subrotina *backtracking*

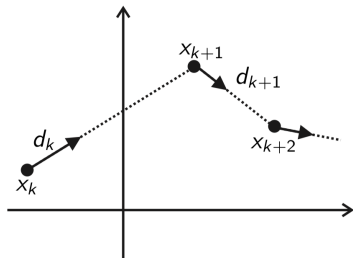


Figura: Método do gradiente - *line search*.

# Método do Gradiente - data1.mat

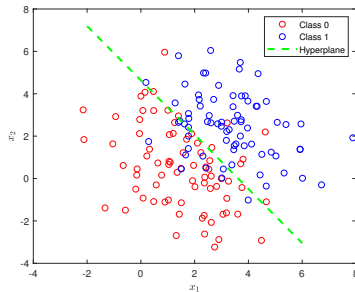
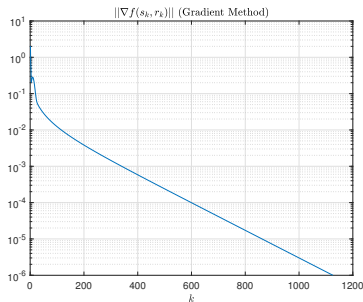


Figura: Resultados do método do gradiente para data1.mat.

$$s = (1.3495, 1.0540) \quad r = 4.8815$$

# Método do Gradiente - data2.mat

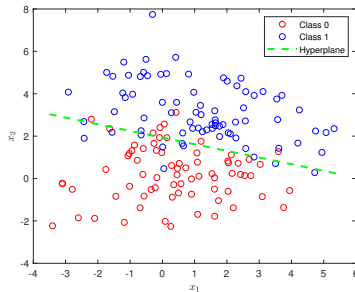
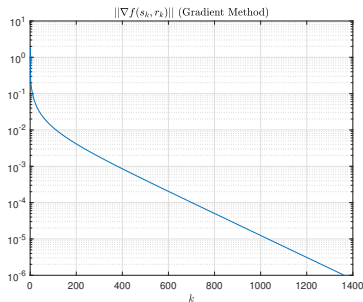


Figura: Resultados do método do gradiente para data2.mat.

$$s = (0.7402, 2.3577) \quad r = 4.5553$$

# Método do Gradiente - data3.mat

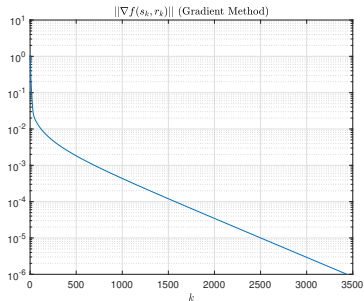


Figura: Resultados do método do gradiente para data3.mat.

$$s = (-1.3082, 1.4078) \quad r = 0.8049$$

# Método do Gradiente - data4.mat

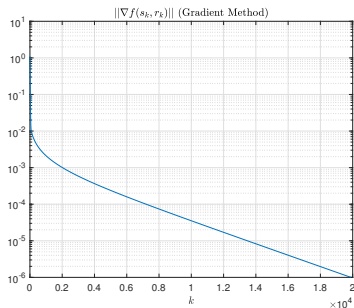


Figura: Resultados do método do gradiente para data4.mat.

$$s = (0.1098, -0.6423) \quad r = 0.1019$$

# Método de Newton

- Iterações evoluem segundo:

$$x_{k+1} = x_k - (\nabla^2 f(x_k))^{-1} \nabla f(x_k) \quad (12)$$

- Onde  $d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$
- Se  $\nabla^2 f(x_k) \succ 0$  (valores próprios positivos), então

$$\nabla f(x_k)^T d_k < 0$$

significa que  $d_k$  tem direção descendente

Parece-se com uma iteração *line search*...

$$x_{k+1} = x_k + \alpha_k d_k$$

$$\alpha_k = 1 \text{ e } d_k = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$$

# Método de Newton - data1.mat

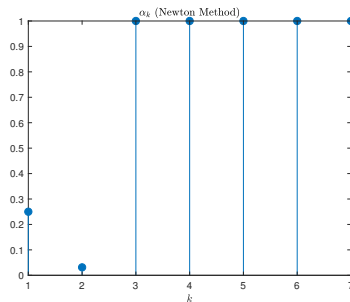
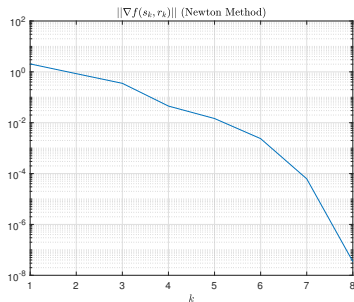


Figura: Resultados do método de Newton para data1.mat.

$$s = (1.3496, 1.0540) \quad r = 4.8817$$



# Método de Newton - data2.mat

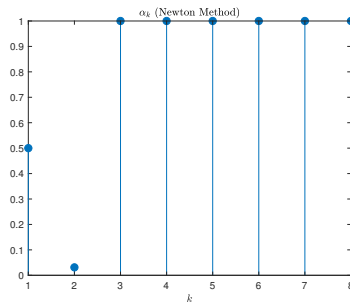
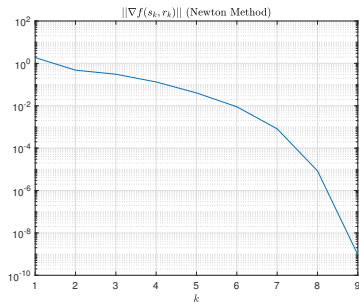


Figura: Resultados do método de Newton para data2.mat.

$$s = (0.7402, 2.3577) \quad r = 4.5554$$

# Método de Newton - data3.mat

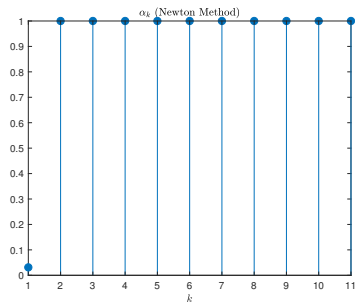
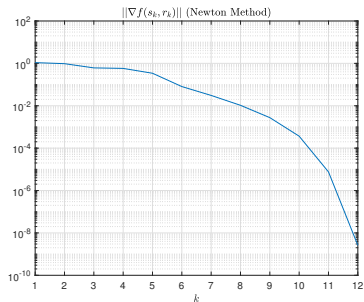


Figura: Resultados do método de Newton para data3.mat.

$$s = (-1.3083, 1.4079) \quad r = 0.8049$$

# Método de Newton - data4.mat

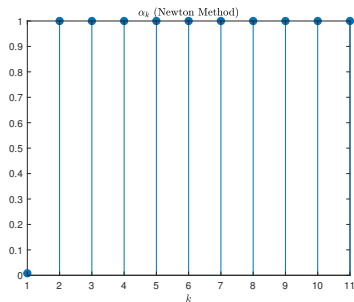
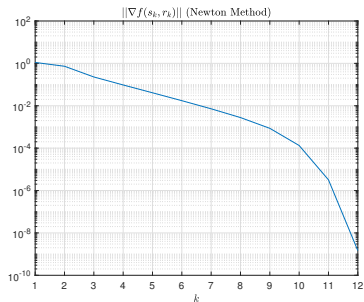


Figura: Resultados do método de Newton para data4.mat.

$$s = (0.1099, -0.6424) \quad r = 0.1019$$

# Método do Gradiente vs. Newton

	Método do Gradiente			Método de Newton		
	Tempo (ms)	Iterações	Tempo/Iter. (ms)	Tempo (ms)	Iterações	Tempo/Iter. (ms)
data1.mat	76,1	1 126	0,0676	16,9	8	2,113
data2.mat	87,7	1 363	0,0643	16,4	9	1,822
data3.mat	776,2	3 437	0,2258	39,9	12	7,492
data4.mat	656 847,4	19 893	33,0190	3 673,1	12	306,092
Custo por Iteração	$\mathcal{O}(n)$			$\mathcal{O}(n^3)$		

**Tabela:** Avaliação do desempenho.

- O método de Newton converge em menos iterações (geralmente próximo de 10 iterações), mas com mais custo por iteração;
- O método do Gradiente converge em muitas mais iterações, mas com um custo muito menor por iteração.

# Método de Levenberg-Marquardt

- Considera problemas de mínimos quadrados não lineares:

$$\underset{x \in \mathbf{R}^n}{\text{minimize}} \quad \underbrace{f_1(x)^2 + f_2(x)^2 + \cdots + f_p(x)^2}_{f(x)}. \quad (13)$$

$f_p : \mathbf{R}^n \rightarrow \mathbf{R}$  são diferenciáveis

- Se todas  $f_p$  forem funções afim, o problema traduz-se num problema de mínimos quadrados, cuja solução advém de  $A \setminus b$

$$\hat{x}_{k+1} = \underset{x \in \mathbf{R}^n}{\operatorname{argmin}} \quad \sum_{p=1}^P (f_p(x_k) + \nabla f_p(x_k)^T (x - x_k))^2 + \lambda_k \|x - x_k\|^2$$

## Pontos chave

- cada  $f_p$  é substituída pela sua linearização em  $x_k$
- $\lambda_k$  penaliza desvios em relação a  $x_k$  ( $\lambda_k > 0$ )

# Método de LM - lndata1.mat

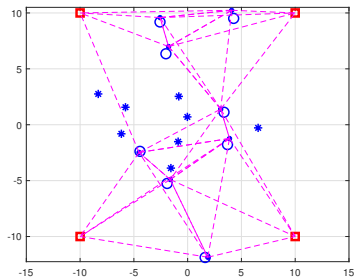
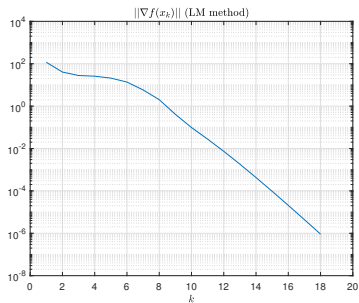


Figura: Resultados do método de LM para lndata1.mat.

# Método de LM - lndata2.mat

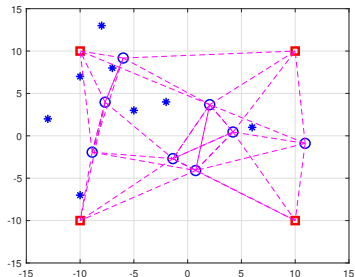
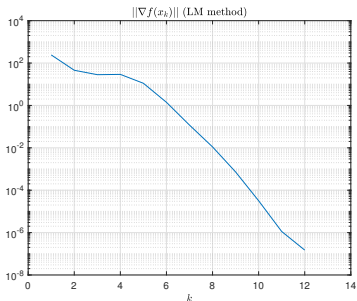


Figura: Resultados do método de LM (inicialização aleatória) para lndata2.mat.

- Escolheu-se a inicialização que resultou no mínimo custo = 4.4945;
- Diferentes inicializações podem convergir para diferentes mínimos locais.

# Formulação do Problema - Tarefa 1

$$\begin{aligned} & \underset{y \in \mathbb{R}^n}{\text{minimize}} && \|p - y\|_2 \\ & \text{subject to} && y \in D(c, r). \end{aligned} \quad (14)$$

$$\begin{aligned} & \underset{y \in \mathbb{R}^n}{\text{minimize}} && \frac{1}{2} \|p - y^*\|_2^2 \\ & \text{subject to} && \|y^* - c\|_2^2 - r^2 \leq 0 \end{aligned} \quad (15)$$

## Condições KKT

$$\begin{cases} \nabla F(y^*) + \mu \nabla g(y^*) = 0 \\ g(y^*) \leq 0 \\ \mu \geq 0 \\ \mu g(y^*) = 0 \end{cases}$$



Diferentes soluções

$$y^* = c \pm \frac{(p - c)}{\|p - c\|_2} r \quad \vee \quad y^* = p \quad (16)$$

Solução ( $p$  no exterior do disco)

$$y^* = c + \frac{(p - c)}{\|p - c\|_2} r \quad (17)$$

# Tentativa de Resolução do Problema - Tarefa 2

$$\begin{aligned} & \underset{x,u}{\text{minimize}} \sum_{k=1}^K \|Ex(\tau_k) - w_k\|_2^2 + \lambda \sum_{t=1}^{T-1} \|u(t) - u(t-1)\|_2^2 = \\ & = \underset{x,u}{\text{minimize}} \sum_{k=1}^K (Ex(\tau_k) - w_k)^T (Ex(\tau_k) - w_k) + \lambda \sum_{t=1}^{T-1} (u(t) - u(t-1))^T (u(t) - u(t-1)) \\ & = \underset{X}{\text{minimize}} \quad \frac{1}{2} (X - P)^T A_m (X - P) \end{aligned}$$
$$A_m = \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & & \lambda & \\ & & & & \ddots \\ & & & & & \lambda \end{bmatrix} \quad X = \begin{bmatrix} Ex(\tau_1) \\ \vdots \\ Ex(\tau_K) \\ u(1) \\ \vdots \\ u(T-1) \end{bmatrix} \quad P = \begin{bmatrix} w_1 \\ \vdots \\ w_K \\ u(0) \\ \vdots \\ u(T-2) \end{bmatrix}$$

# Forma Fechada - Tarefa 2

Reformulação das restrições

$$\underbrace{\begin{bmatrix} 0_1 & \dots & 0_K & A^{T-1}B & \dots & B \end{bmatrix}}_C \underbrace{\begin{bmatrix} Ex(\tau_1) \\ \vdots \\ Ex(\tau_K) \\ u(1) \\ \vdots \\ u(T-1) \end{bmatrix}}_X = \underbrace{\begin{bmatrix} P_2 \\ V_2 \end{bmatrix}}_d - A^T \underbrace{\begin{bmatrix} P_1 \\ V_1 \end{bmatrix}}_d$$

Forma fechada

$$X = P - A_m^{-1} C^T (C A_m^{-1} C^T)^{-1} (C P - d)$$