Formulação matemática de um problema inverso não-linear

Prof. André L. A. dos Reis

Rio de Janeiro 2022

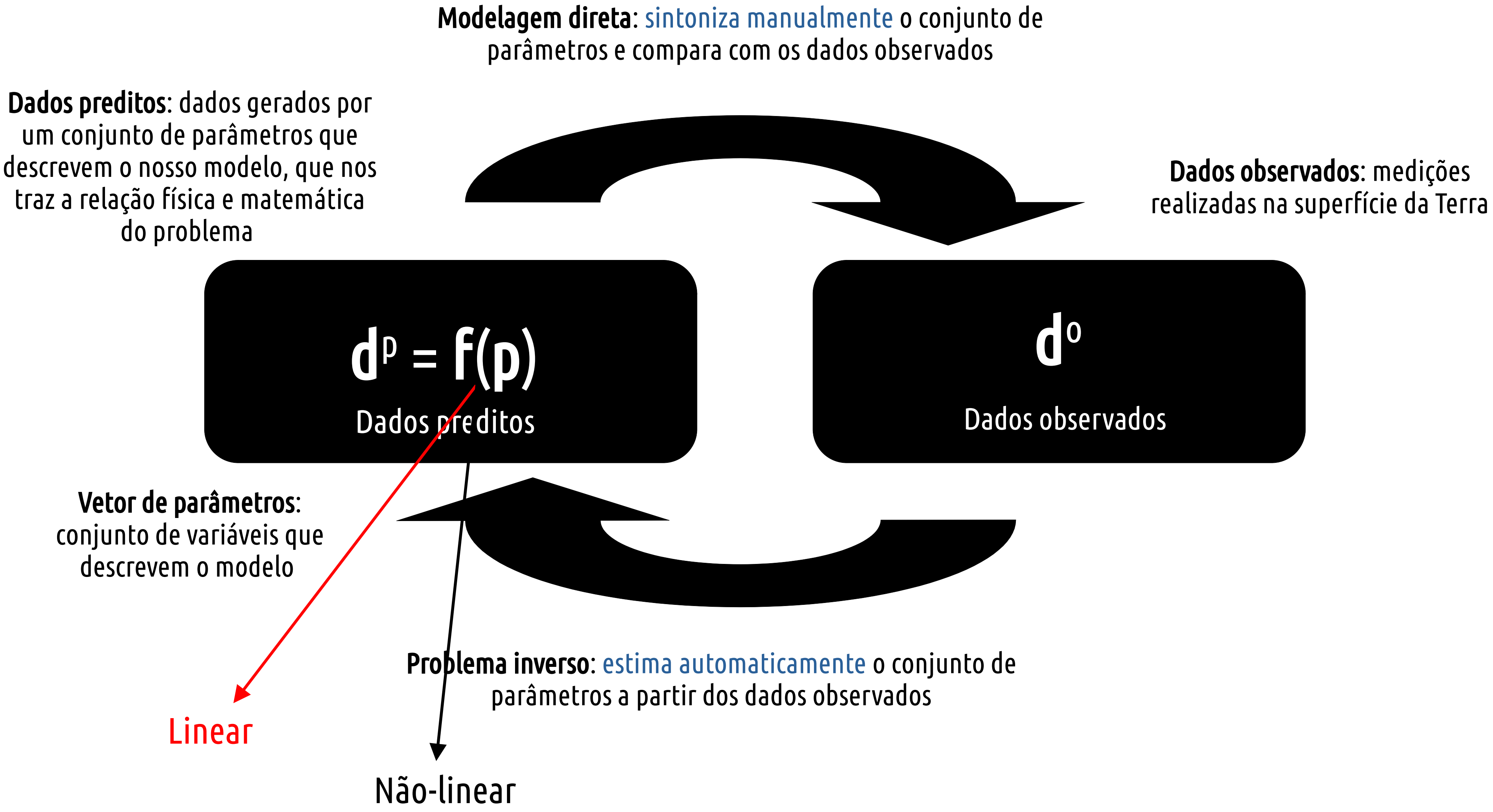
Objetivos da aula

* A formulação de um problema inverso não linear

* Métodos por gradiente:
    - Steepest descent
    - Newton
    - Gauss-Newton
    - Levenberg-Marquardt

* Diferença entre os métodos

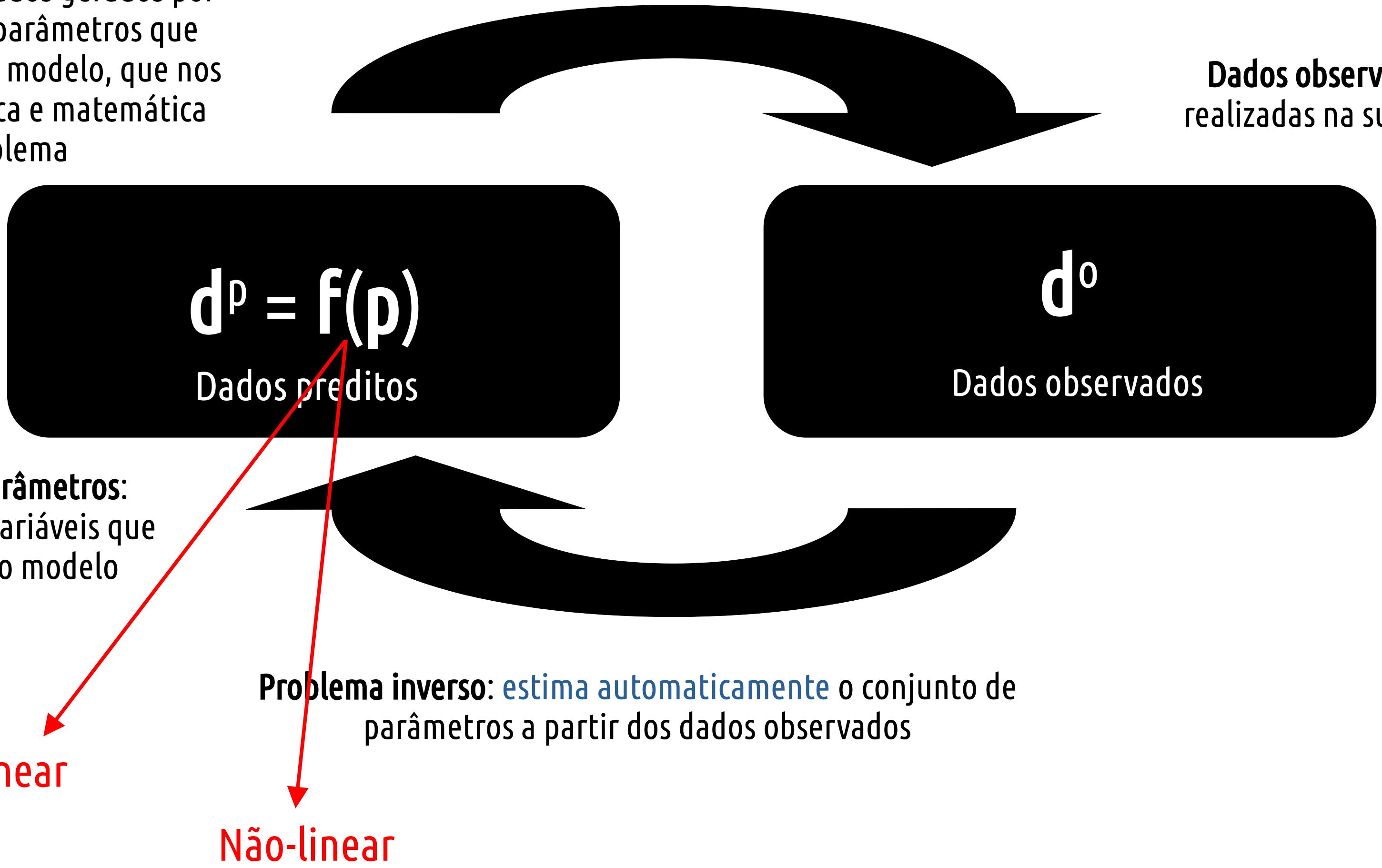
Formulação de um problema inverso não-linear


Modelagem direta: sintoniza manualmente o conjunto de parâmetros e compara com os dados observados
Dados preditos: dados gerados por um conjunto de parâmetros que descrevem o nosso modelo, que nos traz a relação física e matemática do problema
Dados observados: medições realizadas na superfície da Terra
$d^p = f(p)$
Dados preditos
d^o
Dados observados
Vetor de parâmetros: conjunto de variáveis que descrevem o modelo
Problema inverso: estima automaticamente o conjunto de parâmetros a partir dos dados observados
Linear
Não-linear

Modelagem direta: sintoniza manualmente o conjunto de parâmetros e compara com os dados observados

Dados preditos: dados gerados por um conjunto de parâmetros que descrevem o nosso modelo, que nos traz a relação física e matemática do problema

Dados observados: medições realizadas na superfície da Terra



Vetor de parâmetros: conjunto de variáveis que descrevem o modelo

Problema inverso: estima automaticamente o conjunto de parâmetros a partir dos dados observados

Introdução ao problema não-linear

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

Vetor de dados observados

$$\nabla\psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função
de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste

Problema linear

$$\mathbf{d}^p = \mathbf{G}\mathbf{p}$$

$$(\mathbf{G}^T\mathbf{G})\mathbf{p}^\dagger = \mathbf{G}^T\mathbf{d}^o$$

Estimador de mínimos quadrados

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

??????

Introdução ao problema não-linear

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

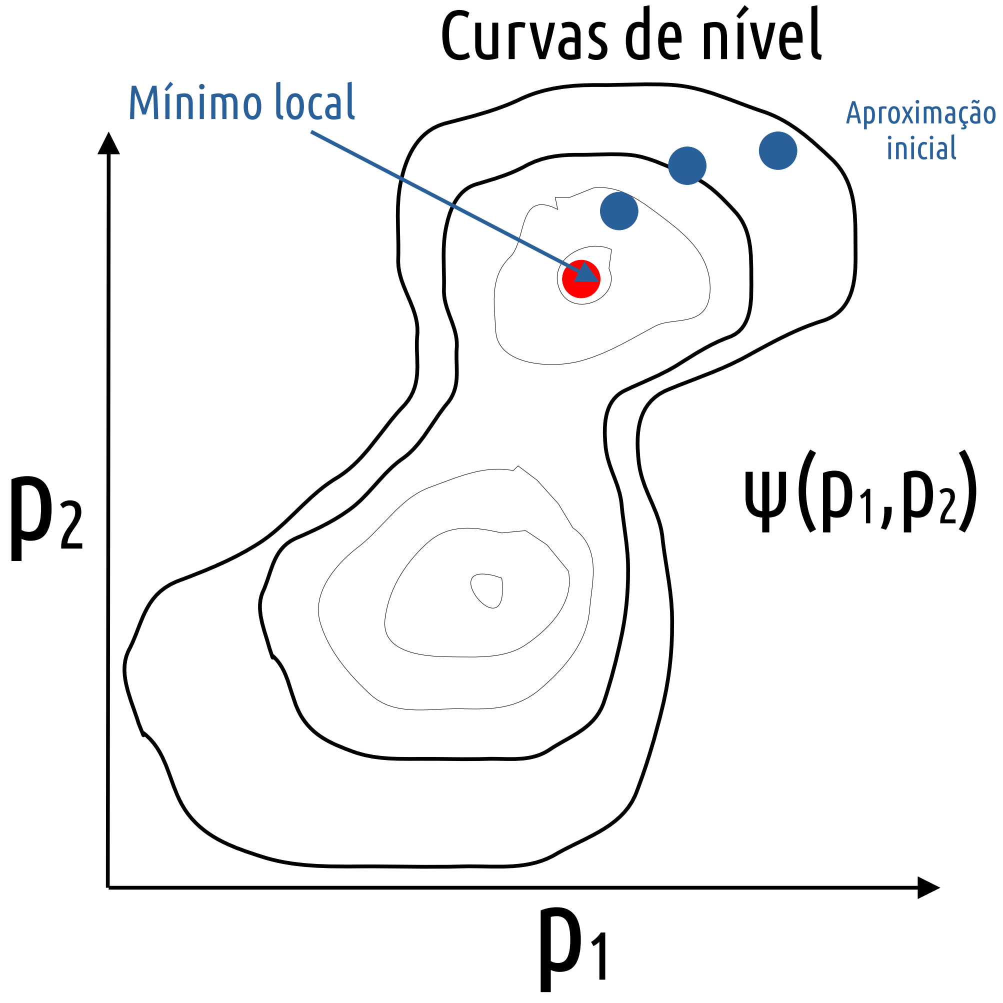
Vetor de dados observados

$$\nabla \psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste



Introdução ao problema não-linear

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

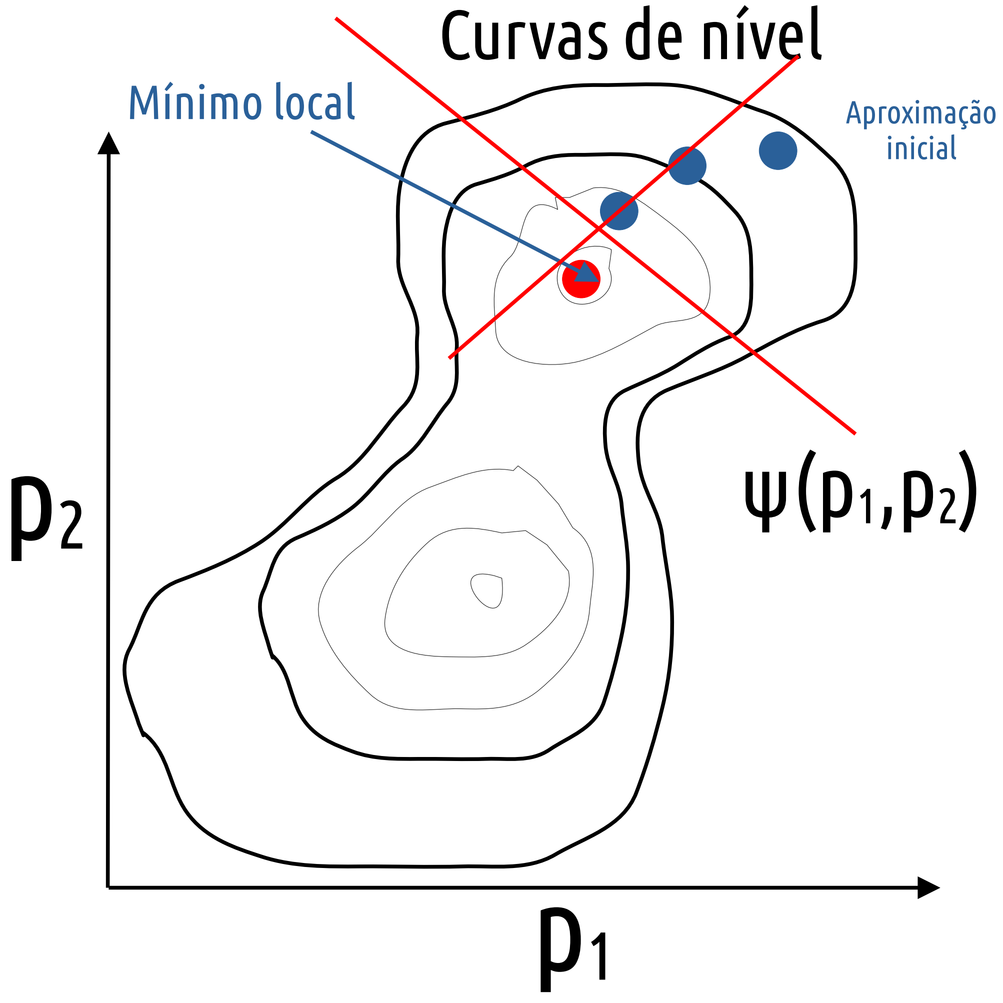
Vetor de dados observados

$$\nabla\psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste



Introdução ao problema não-linear

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

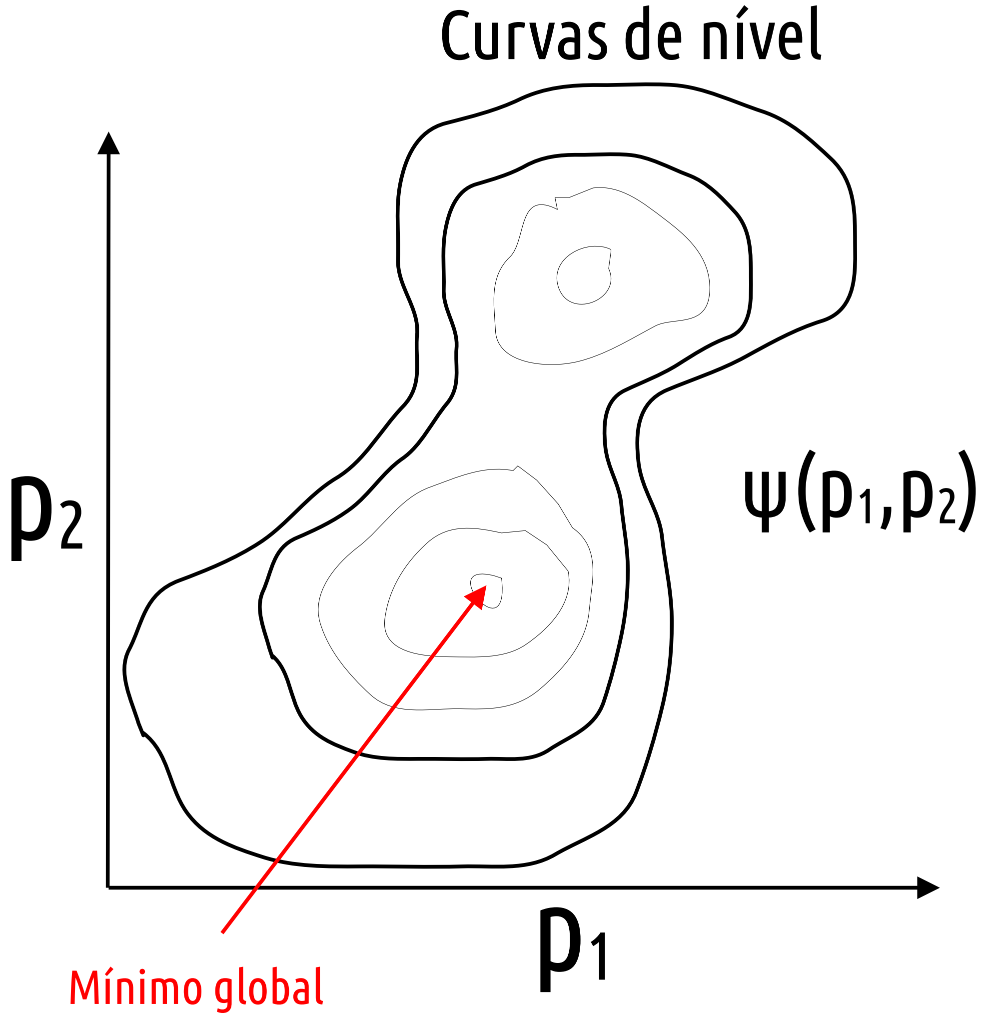
Vetor de dados observados

$$\nabla\psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste



Introdução ao problema não-linear

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

Vetor de dados observados

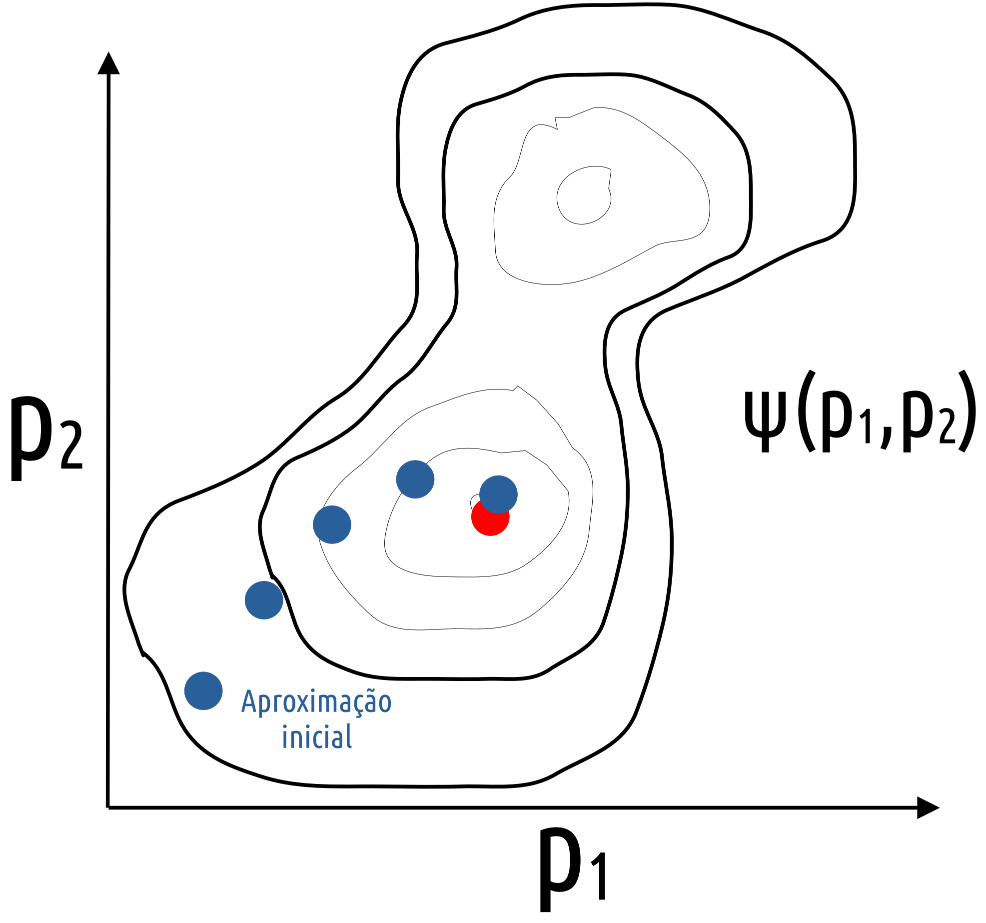
$$\nabla \psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste

Curvas de nível



Introdução ao problema não-linear

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

Vetor de dados observados

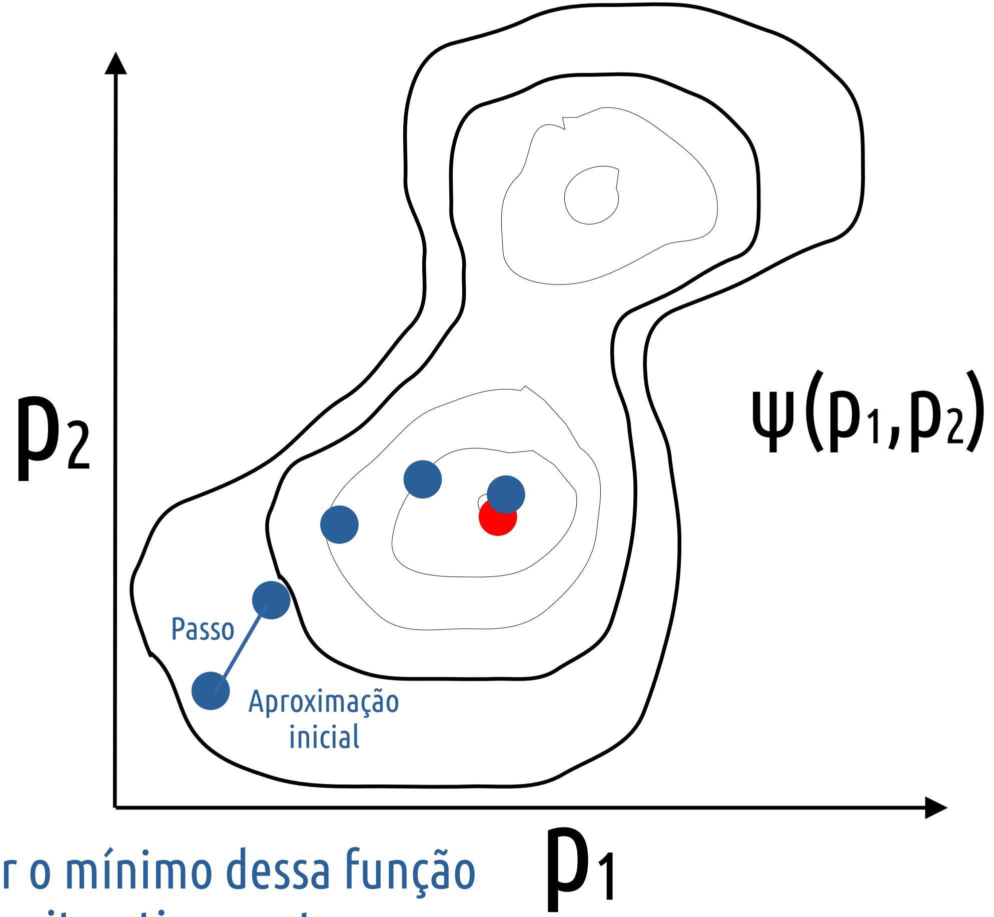
$$\nabla\psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função
de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste

Curvas de nível



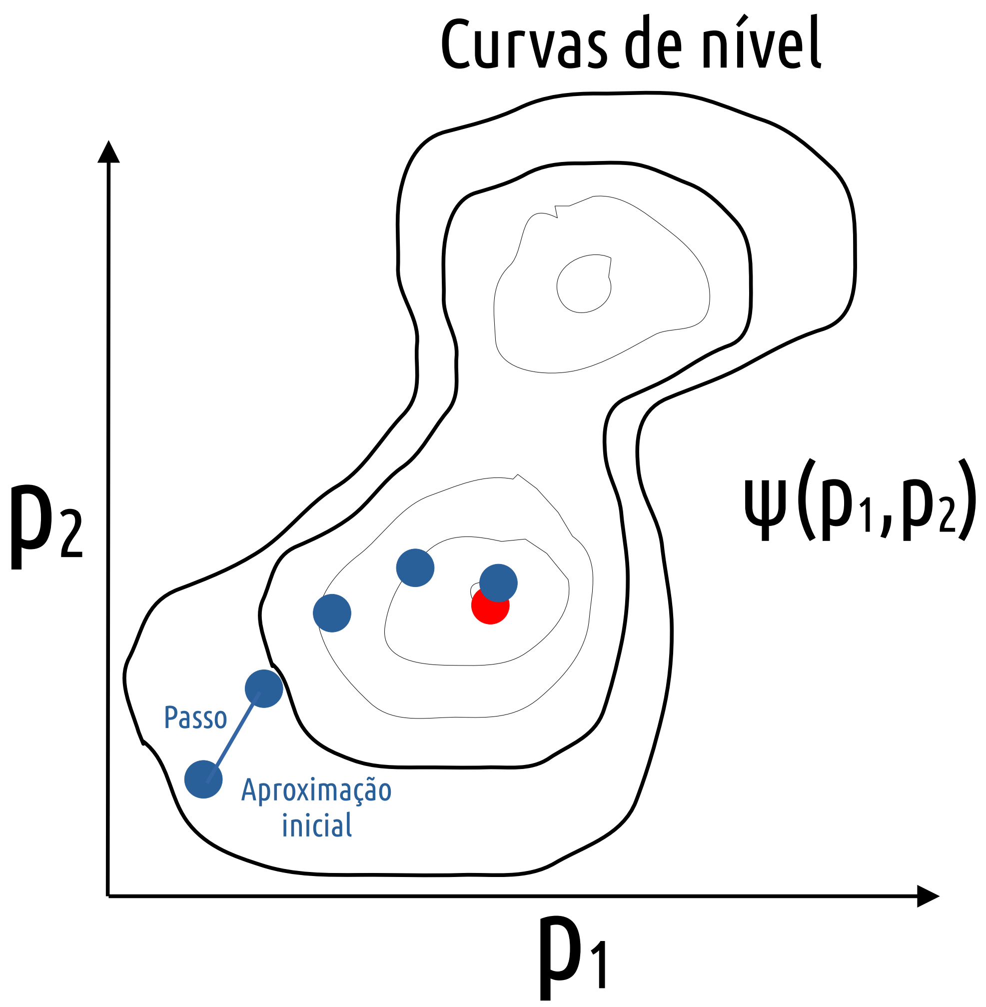
Achar o mínimo dessa função
iterativamente

Introdução ao problema não-linear

Existem dois tipos de métodos que minimizam estas funções:

Métodos por gradiente
(Determinísticos)

Métodos Heurísticos



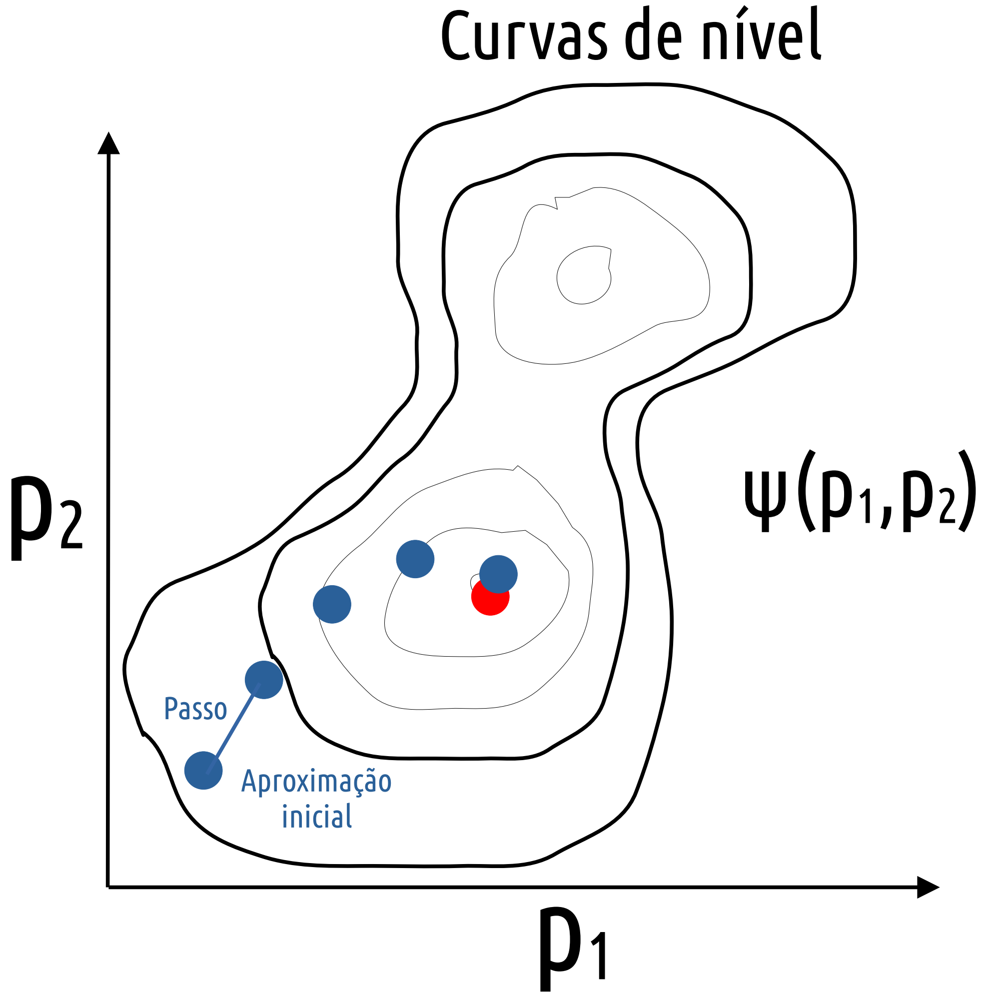
Introdução ao problema não-linear

Métodos por gradiente :

- Steepest descent
- Newton
- Gauss-Newton
- Levenberg-Marquardt

Métodos Heurísticos :

- Simulated Annealing
- Ant Colony
- Algoritmo Genético



Métodos por gradiente

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{f}(\mathbf{p})\|_2^2$$

Expandindo a função de ajuste até segunda ordem, teremos:

$$\psi(\mathbf{p_0} + \mathbf{\Delta p}) \approx \psi(\mathbf{p_0}) + \mathbf{J}(\mathbf{p_0})\mathbf{\Delta p} + \frac{1}{2}\mathbf{\Delta p}^T\mathbf{H}(\mathbf{p_0})\mathbf{\Delta p}$$

$$\mathbf{H}(\mathbf{p_0})\mathbf{\Delta p} = -\mathbf{J}(\mathbf{p_0})$$

Resolveremos este sistema de equações

Métodos por gradiente

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

Vetor de dados observados

$$\nabla \psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função
de ajuste

$$\psi(\mathbf{p}) = ||\mathbf{d}^o - \mathbf{d}^p||_2^2$$

Função de ajuste

Problema linear

$$\mathbf{d}^p = \mathbf{G}\mathbf{p}$$

$$(\mathbf{G}^T\mathbf{G})\mathbf{p}^\dagger = \mathbf{G}^T\mathbf{d}^o$$

Estimador de mínimos quadrados

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

Métodos por gradiente

$$\mathbf{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_M \end{bmatrix}$$

Vetor de parâmetros

$$\mathbf{d}^p = \begin{bmatrix} d_1^p \\ \vdots \\ d_N^p \end{bmatrix}$$

Vetor de dados preditos

$$\mathbf{d}^o = \begin{bmatrix} d_1^o \\ \vdots \\ d_N^o \end{bmatrix}$$

Vetor de dados observados

$$\nabla \psi(\mathbf{p}) = \mathbf{0}$$

O gradiente da função
de ajuste

$$\psi(\mathbf{p}) = \|\mathbf{d}^o - \mathbf{d}^p\|_2^2$$

Função de ajuste

Problema linear

$$\mathbf{d}^p = \mathbf{G}\mathbf{p}$$

$$(\mathbf{G}^T\mathbf{G})\mathbf{p}^\dagger = \mathbf{G}^T\mathbf{d}^o$$

Estimador de mínimos quadrados

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\boldsymbol{\Delta}\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Steepest Descent : $1/\eta$

Newton : $\mathbf{H}(\mathbf{p}_0)$

Gauss-Newton : $\mathbf{J}^T(\mathbf{p}_0)\mathbf{J}(\mathbf{p}_0)$

Levenberg-Marquardt : $\mathbf{J}^T(\mathbf{p}_0)\mathbf{J}(\mathbf{p}_0) + \lambda\mathbf{I}$

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\boldsymbol{\Delta}\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Vamos às contas!

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Método	Convergência
Steepest Descent	0
Levenberg-Marquardt	1
Gauss-Newton	2
Newton	3

Lenta = 0

Rápida = 3

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\Delta\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Método	Chute inicial
Steepest Descent	Pode ser longe
Levenberg-Marquardt	Pode ser longe
Gauss-Newton	Deve ser próxima
Newton	Deve ser próxima

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^{p} \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\boldsymbol{\Delta}\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Método	Passo
Steepest Descent	Depende de um parâmetro e do gradiente
Levenberg-Marquardt	Depende do gradiente, da Hessiana e de um parâmetro
Gauss-Newton	Depende do gradiente e da Hessiana
Newton	Depende do gradiente e da Hessiana

Diferença entre os métodos

Problema não-linear

$$\mathbf{d}^p \neq \mathbf{G}\mathbf{p}$$

$$\mathbf{H}(\mathbf{p}_0)\boldsymbol{\Delta}\mathbf{p} = -\mathbf{J}(\mathbf{p}_0)$$

Sistema de equações

Método	Custo computacional
Steepest Descent	0
Levenberg-Marquardt	2
Gauss-Newton	1
Newton	3

Baixo = 0

Alto = 3

Como regularizar um problema inverso não-linear? E os exemplos?

Até breve!