# Optimisation non linéaire sans contraintes

Recherche opérationnelle

GC-SIE

# Moindres carrés

# Moindres carrés

$$\min f(x) = \frac{1}{2}\|g(x)\|^2 = \frac{1}{2}\sum_{i=1}^{m}\|g_i(x)\|^2$$

- $g_i$: $IR^m \to IR^{r(i)}$ est continûment différentiable, i=1,…,m
- On a souvent r(i) = 1.



# Moindres carrés

Exemple :

Estimation des paramètres d'un modèle

- Soit un modèle mathématique

  z=h(x,y)
- x est le vecteur des paramètres inconnus.
- y est le vecteur d'entrée du modèle.
- z est le vecteur de sortie du modèle.
- On dispose de m observations $(y_i,z_i)$

# Moindres carrés

- Question : quelles sont les valeurs des paramètres telles que le modèle reproduise le mieux les
  observations ?

$$\min \sum_{i=1}^{m} (z_i - h(x, y_i))^2$$



# Moindres carrés

Exemple :

- On veut mesurer la résistivité du cuivre.
- On dispose d'une barre de 1m de cuivre, de section $1cm^2$.
- L'expérience consiste à envoyer des courants de diverses intensités et de mesurer la différence de potentiel.
- Le modèle mathématique est donné par la loi d'Ohm.

# Moindres carrés

- Paramètre inconnu : résistance R
- Entrée du modèle : intensité I
- Sortie du modèle : diff. potentiel V
- Modèle mathématique :

$$V = R\, I = (\lambda/S)\, \rho\, I$$

où $\lambda$ est la longueur, S la section et $\rho$ la résistivité du cuivre.

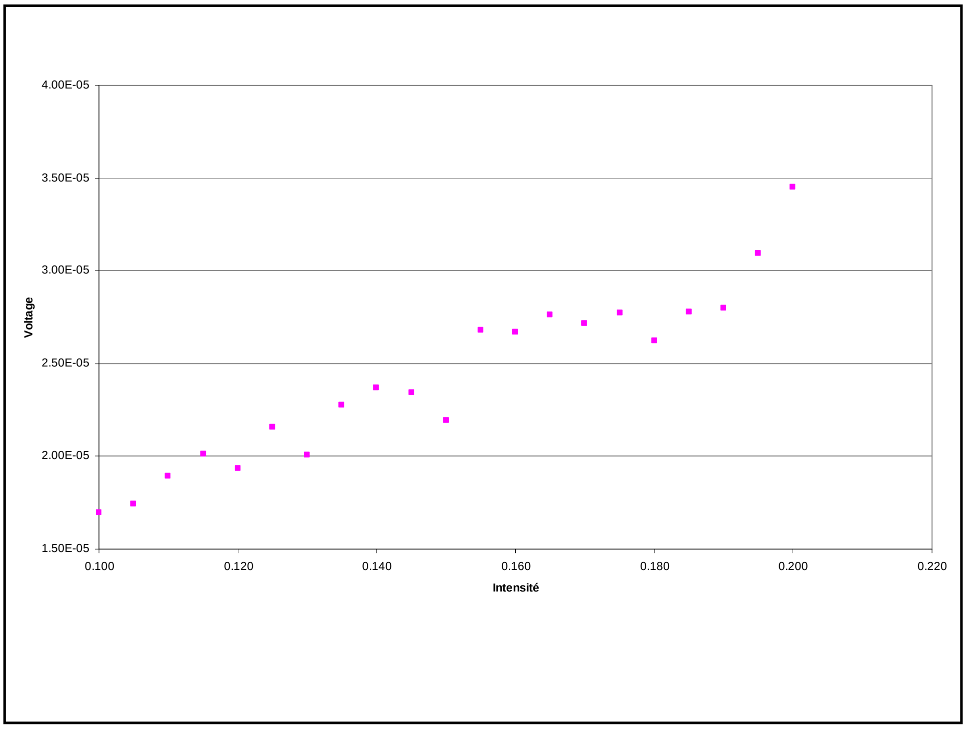


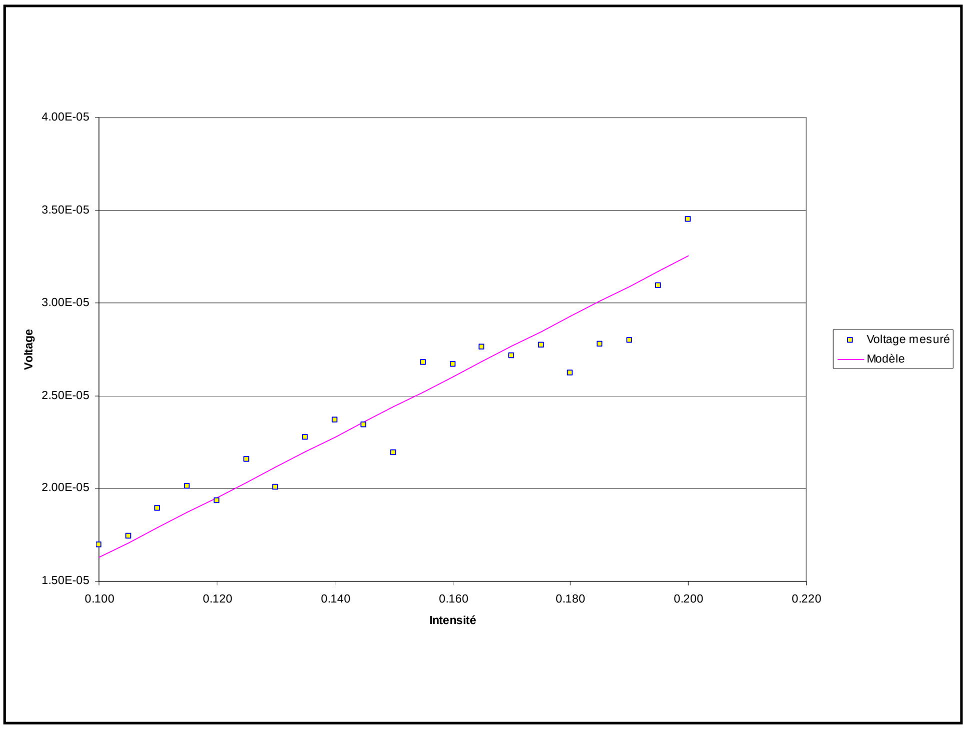
# Moindres carrés

- Données récoltées :

| Intensité | Voltage |
|---|---|
| 0.100 | 1.70E-05 |
| 0.105 | 1.74E-05 |
| 0.110 | 1.89E-05 |
| 0.115 | 2.01E-05 |
| 0.120 | 1.94E-05 |
| 0.125 | 2.16E-05 |
| 0.130 | 2.01E-05 |
| 0.135 | 2.28E-05 |
| 0.140 | 2.37E-05 |
| 0.145 | 2.34E-05 |

| Intensité | Voltage |
|---|---|
| 0.150 | 2.19E-05 |
| 0.155 | 2.68E-05 |
| 0.160 | 2.67E-05 |
| 0.165 | 2.76E-05 |
| 0.170 | 2.72E-05 |
| 0.175 | 2.77E-05 |
| 0.180 | 2.62E-05 |
| 0.185 | 2.78E-05 |
| 0.190 | 2.80E-05 |
| 0.195 | 3.09E-05 |
| 0.200 | 3.45E-05 |


4.00E-05
3.50E-05
3.00E-05
2.50E-05
2.00E-05
1.50E-05
Voltage
0.100
0.120
0.140
0.160
0.180
0.200
0.220
Intensité


4.00E-05
3.50E-05
3.00E-05
2.50E-05
2.00E-05
1.50E-05
Voltage
0.100
0.120
0.140
0.160
0.180
0.200
0.220
Intensité
Voltage mesuré
Modèle

# Moindres carrés

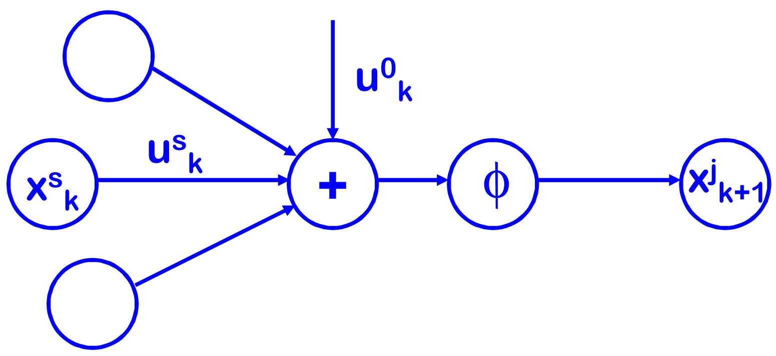
Réseaux de neurones.

- Modèle spécifié par un système multi-niveaux.
- Le niveau consiste en $n_k$ unités d'activitation ou neurone.
- Chaque unité d'activation est une relation entrée-sortie

$$\phi : \mathrm{IR} \rightarrow \mathrm{IR}$$



# Moindres carrés



# Moindres carrés

- La sortie de la $j^{\text{ième}}$ unité d'activation du niveau k+1 est notée $x^j_{k+1}$.
- L'entrée est une fonction linéaire des sorties du niveau k.
- Donc

$$x^j_{k+1} = \phi\left(u_k^{0j} + \sum_{s=1}^{n_k} x_k^s u_k^{sj}\right) \quad j = 1, \ldots, n_{k+1}.$$



# Moindres carrés

- Les $u_k^s$ sont appelés « poids »
- Ce sont les paramètres à déterminer.
- Pour un ensemble de paramètres donnés, et si N est le nombre de niveaux, à chaque vecteur d'entrée $x_0$ du niveau 0 correspond un vecteur de sortie $x_N$ du niveau N.

# Moindres carrés

- Le réseau de neurones peut donc être considéré comme un modèle mathématique

$$z=h(x,y)$$

- où
  - x est le vecteur de poids
  - y est le vecteur d'entrées au niveau 0
  - z est le vecteur de sorties au niveau N



# Moindres carrés

- La phase d'entrainement du réseau, ou phase d'apprentissage peut donc être considérée comme la résolution d'un problème de moindres carrés.
- Exemples typiques de fonctions d'activation :

  Fonction sigmoidale :

$$\phi(\xi) = \frac{1}{1+e^{-\xi}}$$

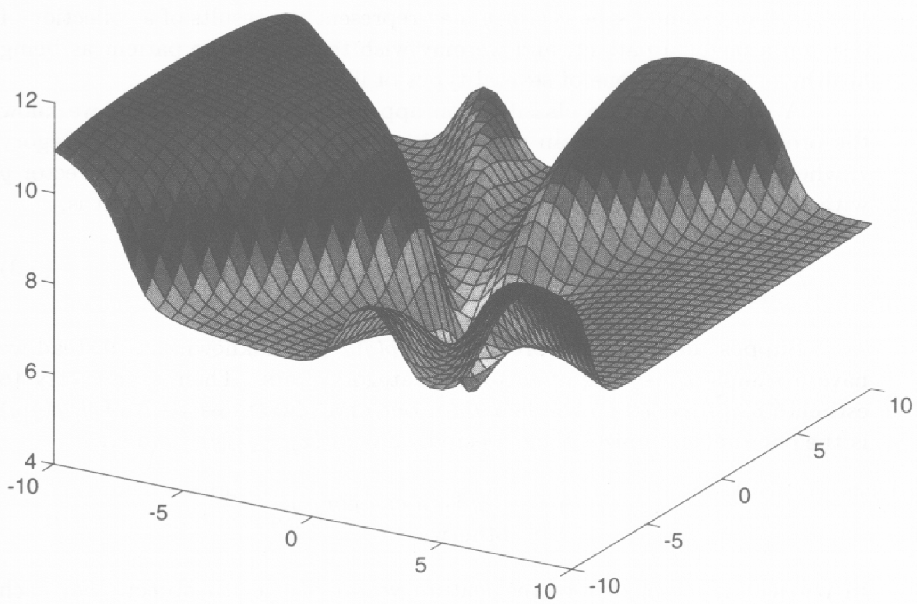
  Fonction hyperbolique tangente :

$$\phi(\xi) = \frac{e^{\xi}-e^{-\xi}}{e^{\xi}+e^{-\xi}}$$

# Moindres carrés

- Le problème d'entrainement de réseaux neuronaux est souvent très compliqué. Les fonctions de coûts associées sont non-convexes et possèdent souvent des minima locaux multiples.
- Exemple à deux paramètres.






Source: Bertsekas (1995) Nonlinear programming, Athena Scientific

# Gauss-Newton

$$\min f(x) = \frac{1}{2}\|g(x)\|^2 = \frac{1}{2}\sum_{i=1}^{m}\|g_i(x)\|^2$$

Idée

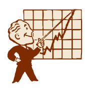
- Travailler sur g et non sur f.
- Linéarisation de g :

$$m(x) = g(x_k) + \nabla g(x_k)^T(x - x_k)$$



# Gauss-Newton

- Minimiser la norme de m(x)



$$\begin{aligned}\tfrac{1}{2}\|m(x)\|^2 = & \\ & \tfrac{1}{2}\|g(x_k)\|^2 \\ + & \ (x - x_k)^T\nabla g(x_k)g(x_k) \\ + & \ \tfrac{1}{2}(x - x_k)^T\nabla g(x_k)\nabla g(x_k)^T(x - x_k)\end{aligned}$$

## Gauss-Newton

- Si f(x) = ½ ¦¦m(x)¦¦², alors

$$\nabla f(x) = \nabla g(x_k)\nabla g(x_k)^T(x - x_k) + \nabla g(x_k)g(x_k)$$

- **Le minimum est atteint en**

$$x = x_k - \left(\nabla g(x_k)\nabla g(x_k)^T\right)^{-1}\nabla g(x_k)g(x_k)$$

**si la matrice est inversible.**



## Gauss-Newton

- Une itération Gauss-Newton pure est

$$x_{k+1} = x_k - (\nabla g(x_k)\nabla g(x_k)^T)^{-1}\nabla g(x_k)g(x_k)$$

- **$\nabla g(x_k)g(x_k)$ est le gradient de ½¦¦g(x)¦¦² en $x_k$**
- **Si $\nabla g(x_k)\nabla g(x_k)^T$ est définie positive, nous avons donc une direction de descente.**

# Gauss-Newton

- Tout comme la méthode de Newton pure pour le cas général, la méthode de Gauss-Newton pure pour les moindres carrés peut ne pas converger.
- Solution :

$$x_{k+1} = x_k - \alpha_k(\nabla g(x_k)\nabla g(x_k)^T + \Delta_k)^{-1}\nabla g(x_k)g(x_k)$$



# Gauss-Newton

$$x_{k+1} = x_k - \alpha_k(\nabla g(x_k)\nabla g(x_k)^T + \Delta_k)^{-1}\nabla g(x_k)g(x_k)$$

- $\alpha_k$ est choisi par la règle d'Armijo.
- $\Delta_k$ est une matrice diagonale telle que $\nabla g(x_k)\nabla g(x_k)^T+\Delta_k$ soit défini positif.
- Méthode de Levenberg-Marquardt :

  $\Delta_k$ = multiple de l'identité

# Gauss-Newton

Cas linéaire

- $g(x)=Cx-z$
- $\nabla g(x) = C^T$
- $x_{k+1} = x_k-(C^TC)^{-1}C^T(Cx_k-z)$

  $= (C^TC)^{-1}C^Tz \quad \forall k$
- La solution est obtenue en une itération
- Note: le système d'équations

  $$C^TCx=C^Tz$$

  est appelé équations normales.



# Gauss-Newton

Relation avec la méthode de Newton

- Soit $g:IR^n \rightarrow IR^m$
- $f(x)= ½ \lVert g(x) \rVert^2$

$$\begin{aligned} \nabla f(x) &= \nabla g(x)^T g(x) \\ &= \sum_{i=1}^{m} \nabla g_i(x) g_i(x) \end{aligned}$$

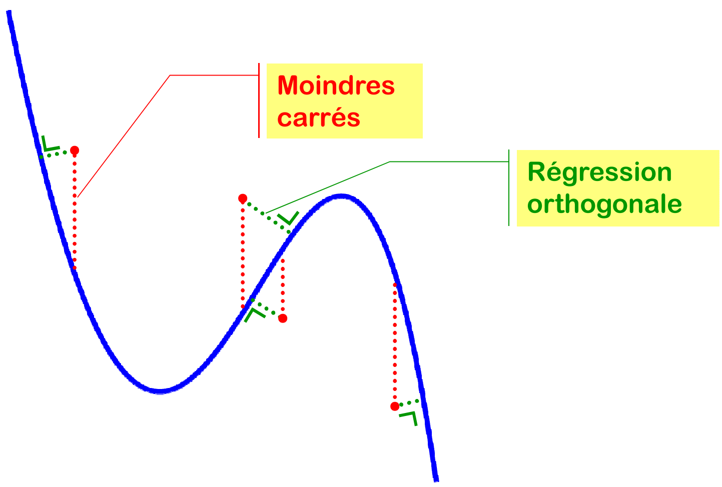
## Gauss-Newton

$$\begin{aligned}\nabla^2 f(x) &= \Sigma_{i=1}^m [\nabla^2 g_i(x) g_i(x) \\ &+ \nabla g_i(x) \nabla g_i(x)^T] \\ \\ &= \nabla g(x) \nabla g(x)^T \\ &+ \Sigma_{i=1}^m \nabla^2 g_i(x) g_i(x)\end{aligned}$$

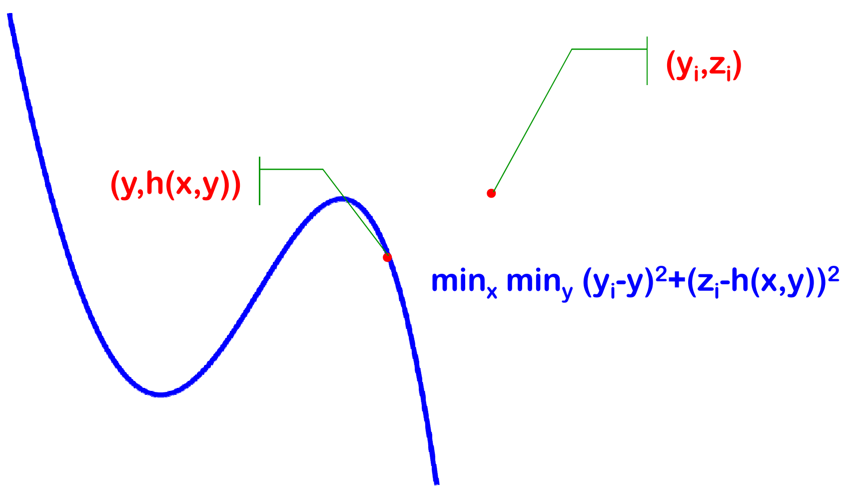
Gauss-Newton = Newton
en négligeant le second terme



## Régression orthogonale



# Régression orthogonale






# Régression orthogonale

Notes :

- Cette régression est utilisée lorsque des erreurs sont présentes aussi dans les entrées du modèle.
- On suppose que les erreurs dans les entrées et les erreurs dans les sorties sont indépendantes, de moyenne nulle.
- Même si le modèle est linéaire, le problème de moindres carrés n'est pas linéaire.

# Régression orthogonale

- Exemples :
  - Résistivité du cuivre
    - Moindres carrés : r = 1.61046 $10^{-8}$
    - Régression orthogonale : 1.60797 $10^{-8}$
  - Modèle sphérique :
    - $z = x_1 + [x_2^2-(y-x_3)^2]^{½}$



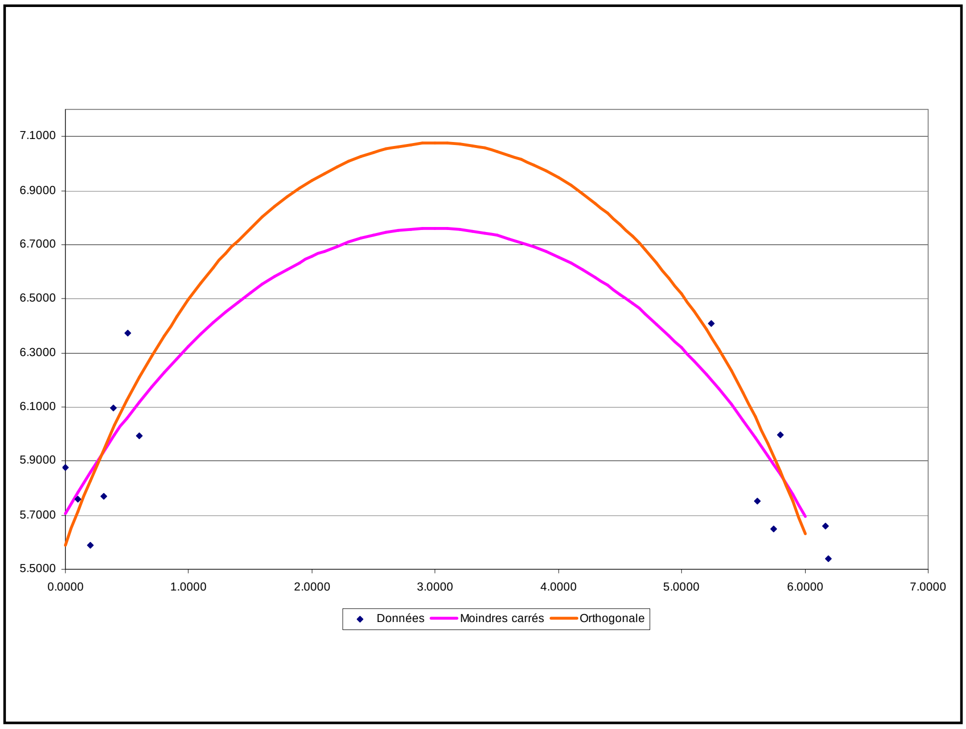
# Régression orthogonale

| y | z |
|---|---|
| 0.0000 | 5.8776 |
| 0.0974 | 5.7608 |
| 0.2042 | 5.5878 |
| 0.3090 | 5.7680 |
| 0.3911 | 6.0966 |
| 0.5092 | 6.3744 |
| 0.5972 | 5.9919 |
| 5.2431 | 6.4085 |
| 5.8022 | 5.9953 |
| 5.6136 | 5.7521 |
| 5.7435 | 5.6481 |
| 6.1645 | 5.6582 |
| 6.1940 | 5.5383 |

**Modèle réel :**
- $\mathbf{x_1 = 3}$
- $\mathbf{x_2 = 4}$
- $\mathbf{x_3 = 3}$

Moindres carrés :
- $x_1 = 1.9809$
- $x_2 = 4.7794$
- $x_3 = 2.9938$

Orthogonale :
- $x_1 = 3.2759$
- $x_2 = 3.8001$
- $x_3 = 3.0165$


7.1000
6.9000
6.7000
6.5000
6.3000
6.1000
5.9000
5.7000
5.5000
0.0000
1.0000
2.0000
3.0000
4.0000
5.0000
6.0000
7.0000
Données
Moindres carrés
Orthogonale