

MODÉLISATION ET INVERSION EN GÉOPHYSIQUE

5 - Inversion linéaire

Bernard Giroux
(bernard.giroux@ete.inrs.ca)

Institut national de la recherche scientifique
Centre Eau Terre Environnement

Version 1.0.2
Hiver 2018

Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Régression linéaire

- La façon la plus courante de résoudre un problème d'inversion linéaire est basée sur la mesure de la distance entre les données observées \mathbf{d}^{obs} et les données prédites \mathbf{d}^{pre} ;
- Cette distance est fonction de l'erreur de prédiction, définie pour une i^{e} observation par

$$e_i = d_i^{\text{obs}} - d_i^{\text{pre}}. \quad (1)$$

- La méthode des moindres-carrés est l'approche la plus fréquente pour estimer les paramètres du modèle \mathbf{m}^{est} ;
 - On cherche dans ce cas les paramètres qui donneront l'erreur E la plus faible, où

$$E = \sum_{i=0}^{N-1} e_i^2 = \mathbf{e}^T \mathbf{e}. \quad (2)$$

- L'erreur E est la *distance euclidienne* au carré du vecteur \mathbf{e} .

Régression linéaire

Aperçu

Distance

Moindres-carrés

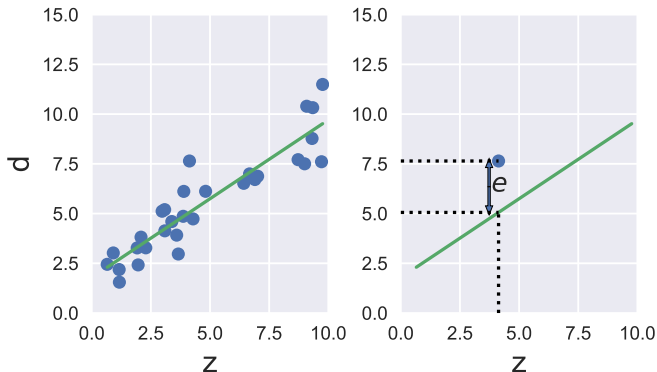
Existence de la solution

Hypothèses *a priori*

Varianse des
paramètres

Inverse généralisée

- L'exemple suivant montre l'ajustement de points par une droite, obtenu par moindres-carrés.



- La distance euclidienne est une mesure parmi d'autres;
 - On peut par exemple considérer la somme des valeurs absolues.
- On utilise *norme* pour désigner une mesure de distance;
- La norme d'un vecteur est notée $\|\mathbf{e}\|$
- On dénombre :

$$\text{norme } L_1 : \quad \|\mathbf{e}\|_1 = \left[\sum_i |e_i| \right] \quad (3)$$

$$\text{norme } L_2 : \quad \|\mathbf{e}\|_2 = \left[\sum_i |e_i|^2 \right]^{1/2} \quad (4)$$

$$\text{norme } L_n : \quad \|\mathbf{e}\|_n = \left[\sum_i |e_i|^n \right]^{1/n} \quad (5)$$

- Lorsque $n \rightarrow \infty$, seule la valeur la plus élevée a un poids non nul, i.e.

$$\text{norme } L_\infty : \quad \|\mathbf{e}\|_\infty = \max_i |e_i| \quad (6)$$

Régression linéaire

Aperçu

Distance

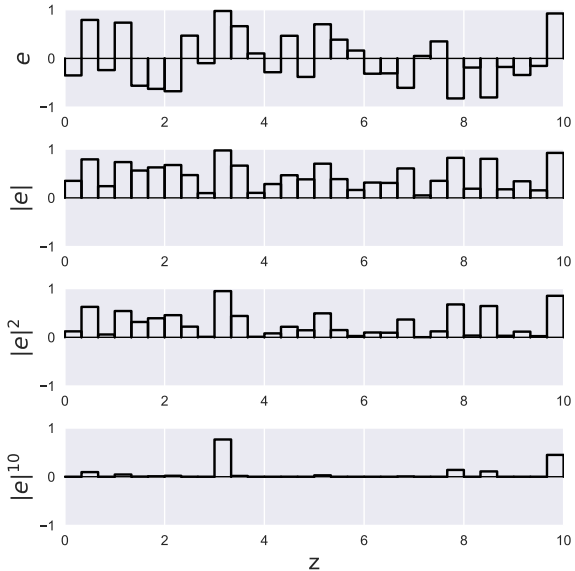
Moindres-carrés

Existence de la solution

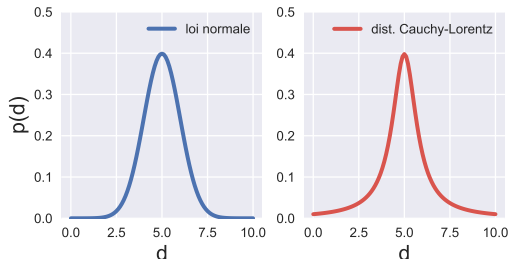
Hypothèses *a priori*

Variance des
paramètres

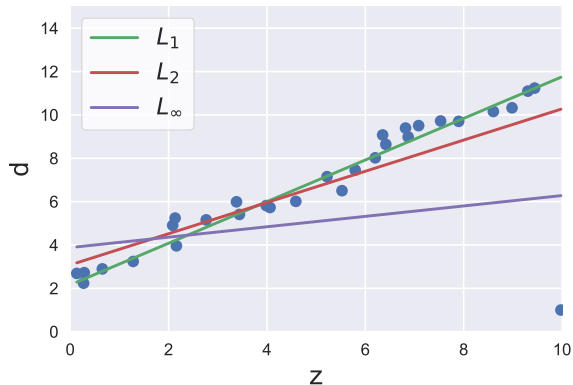
Inverse généralisée



- Le choix d'une norme dépend principalement de l'importance donnée aux données aberrantes ;
- Une norme plus élevée donne un poids plus élevé aux erreur de prédiction e_i plus élevées.
- La norme L_2 implique que les données sont distribuées selon une loi normale ;
 - Une distribution normale est assez peu étalée.



- Si les données contiennent quelques points aberrants, la distribution sera plus étalée et les résultats peuvent être complètement erronés.



Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

- Une droite est définie par une ordonnée à l'origine (m_0) et par une pente (m_1), i.e.

$$d_i = m_0 + m_1 z_i. \quad (7)$$

- Il y a donc deux paramètres du modèle, $M=2$.
- Typiquement, on dispose de beaucoup plus que deux points, i.e. $N > M$.
- À moins que les points ne s'alignent parfaitement, on ne peut trouver une droite qui passe par tout les points ;
- On a affaire à un problème *surdéterminé*, il n'y a pas de solution pour laquelle $\mathbf{e} = 0$.

Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

 Hypothèses *a priori*

 Variance des
paramètres

Inverse généralisée

- On cherche alors une solution approximative, où le niveau d'approximation est défini par

$$E = \mathbf{e}^T \mathbf{e} = \sum_{i=0}^{N-1} (d_i - m_0 - m_1 z_i)^2. \quad (8)$$

- On cherche donc le minimum de $E(m_0, m_1)$, qui est obtenu en égalant les dérivées de E à zéro et en solutionnant :

$$\frac{\partial E}{\partial m_0} = \frac{\partial}{\partial m_0} \sum_{i=0}^{N-1} (d_i - m_0 - m_1 z_i)^2 \quad (9)$$

$$= 2Nm_0 + 2m_1 \sum_{i=0}^{N-1} z_i - 2 \sum_{i=0}^{N-1} d_i = 0 \quad (10)$$

$$\frac{\partial E}{\partial m_1} = \frac{\partial}{\partial m_1} \sum_{i=0}^{N-1} (d_i - m_0 - m_1 z_i)^2 \quad (11)$$

$$= 2m_0 \sum_{i=0}^{N-1} z_i + 2m_1 \sum_{i=0}^{N-1} z_i^2 - 2 \sum_{i=0}^{N-1} z_i d_i = 0. \quad (12)$$

- On peut généraliser les moindres-carrés à n'importe quel système linéaire;
- L'erreur vaut alors

$$E = \mathbf{e}^T \mathbf{e} = (\mathbf{d} - \mathbf{Gm})^T (\mathbf{d} - \mathbf{Gm}) = \sum_{i=0}^{N-1} \left[d_i - \sum_{j=0}^{M-1} G_{ij} m_j \right] \left[d_i - \sum_{k=0}^{M-1} G_{ik} m_k \right] \quad (13)$$

- En multipliant les termes et changeant l'ordre des sommations, on trouve

$$E = \underbrace{\sum_{j=0}^{M-1} \sum_{k=0}^{M-1} m_j m_k \sum_{i=0}^{N-1} G_{ij} G_{ik}}_{T_1} - 2 \underbrace{\sum_{j=0}^{M-1} m_j \sum_{i=0}^{N-1} G_{ij} d_i}_{T_2} + \underbrace{\sum_{i=0}^{N-1} d_i d_i}_{T_3} \quad (14)$$

Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

 Hypothèses *a priori*

 Variance des
paramètres

Inverse généralisée

- Les dérivées sont maintenant calculées
- Pour le 1^e terme, on a

$$\frac{\partial T_1}{\partial m_q} = \sum_{j=0}^{M-1} \sum_{k=0}^{M-1} [\delta_{jq} m_k + m_j \delta_{kq}] \sum_{i=0}^{N-1} G_{ij} G_{ik} \quad (15)$$

$$= 2 \sum_{k=0}^{M-1} m_k \sum_{i=0}^{N-1} G_{iq} G_{ik} \quad (16)$$

où

$$\delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad (17)$$

provient du fait que $\partial m_i / \partial m_j$ vaut 1 si $i = j$ et 0 si $i \neq j$.

Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

- Pour le 2^e terme, on a

$$-2 \frac{\partial T_2}{\partial m_q} = -2 \sum_{j=0}^{M-1} \delta_{jq} \sum_{i=0}^{N-1} G_{ij} d_i = -2 \sum_{i=0}^{N-1} G_{iq} d_i \quad (18)$$

- Le 3^e terme ne contient pas de m , alors $\frac{\partial T_3}{\partial m_q} = 0$.
- En combinant les 3 termes, on trouve

$$\frac{\partial E}{\partial m_q} = 0 = 2 \sum_{k=0}^{M-1} m_k \sum_{i=0}^{N-1} G_{iq} G_{ik} - 2 \sum_{i=0}^{N-1} G_{iq} d_i \quad (19)$$

- Sous forme matricielle, cela donne

$$\mathbf{G}^T \mathbf{G} \mathbf{m} - \mathbf{G}^T \mathbf{d} = 0. \quad (20)$$

Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

- Dans l'équation (20), $\mathbf{G}^T \mathbf{G}$ est une matrice carrée de taille $M \times M$ qui multiplie un vecteur \mathbf{m} de M éléments;
- $\mathbf{G}^T \mathbf{d}$ est aussi un vecteur de M éléments;
- En supposant que $[\mathbf{G}^T \mathbf{G}]^{-1}$ existe, l'estimateur des paramètres du modèle est

$$\mathbf{m}^{\text{est}} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d} \quad (21)$$

Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

Hypothèses *a priori*Variance des
paramètres

Inverse généralisée

- Les commandes suivantes permettent de générer un ensemble de points plus ou moins alignés le long d'une droite :

```
N = 30
zmin = 0
zmax = 10
z = np.sort(zmin + zmax*np.random.rand(N, 1), axis=0)

a = 2.0
b = 1.0
m = np.asarray([a, b])
sd = 0.5
dobs = m[0] + m[1] * z + sd*np.random.randn(N, 1)

plt.plot(z, dobs, 'o')
plt.xlabel('z', fontsize=16)
plt.ylabel('d', fontsize=16)
plt.show()
```

Exercice – ajuster une droite

Régression linéaire

Aperçu

Distance

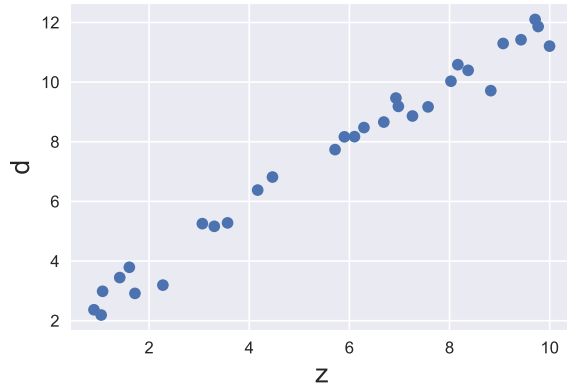
Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée



Régression linéaire

Aperçu

Distance

Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

- Étapes à suivre :
 - Construire la matrice \mathbf{G} ;
 - Calculer $\mathbf{A} = \mathbf{G}^T \mathbf{G}$;
 - Calculer $\mathbf{b} = \mathbf{G}^T \mathbf{d}_{\text{obs}}$;
 - Calculer l'inverse de \mathbf{A} ;
 - Calculer $\mathbf{m}_{\text{est}} = \mathbf{A}^{-1} \mathbf{b}$.
- Visualisez le résultat avec

```
dpre = G.dot(mest)
```

```
plt.plot(z, dobs, 'o')
plt.plot(z, dpre, '-', linewidth=4)
plt.xlabel('z', fontsize=16)
plt.ylabel('d', fontsize=16)
plt.show()
```

Exercice – ajuster une droite

Régression linéaire

Aperçu

Distance

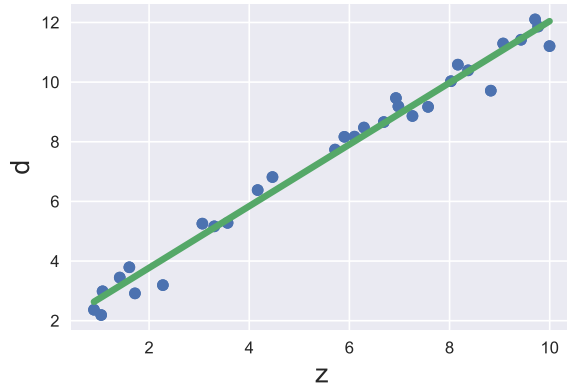
Moindres-carrés

Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée



Existence de la solution moindres-carrés

Régression linéaire

Aperçu

Distance

Moindres-carrés

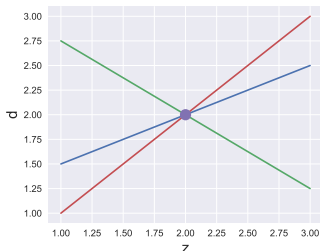
Existence de la solution

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

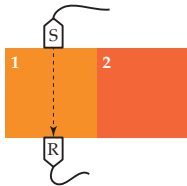
- La solution des moindres-carrés a été retenue parce qu'il n'y a pas de solution exacte à notre problème;
- C'est la méthode qui nous donne la "meilleure" solution, au sens où la norme L_2 est minimisée;
- En utilisant $\mathbf{m}^{\text{est}} = [\mathbf{G}^T \mathbf{G}]^{-1} \mathbf{G}^T \mathbf{d}$, on assume qu'il n'y a qu'une seule "meilleure" solution;
- La méthode échoue s'il existe plusieurs solutions qui donne la même erreur E .



Ajustement d'une droite avec un seul point :

- Une infinité de droites passe par le point;
- Pour chaque droite, $E = 0$.

- On peut classer les problèmes inverses en fonction de l'information contenue dans le système $\mathbf{Gm} = \mathbf{d}$
- Le problème est **indéterminé** (*underdetermined*) lorsque le nombre de paramètres M est supérieur au nombre de données indépendantes $N, M > N$;
 - La matrice $[\mathbf{G}^T \mathbf{G}]^{-1}$ est singulière (non inversible).



- Lorsque $M < N$, le problème est **surdéterminé** (*overdetermined*);
 - Les moindres-carrés sont appropriés.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

Hypothèses *a priori*

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération
Égalité

Variance des
paramètres

Inverse généralisée

- Lorsqu'un problème est indéterminé, il existe une infinité de solutions et **il faut ajouter une information au système** pour arriver à une solution satisfaisante ;
- Cette information est nommée **information *a priori*** ;
 - Par exemple, pour ajuster une droite avec un seul point, on peut assumer que la droite doit passer à l'origine.
 - Un autre exemple est de supposer que les paramètres doivent être à l'intérieur d'une plage de valeurs donnée, e.g. des densités entre 1000 et 3500 kg/m³.
- Le choix d'une hypothèse *a priori* n'est pas toujours évident et dépend clairement de l'application.

Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- Une hypothèse *a priori* fréquente est que le modèle \mathbf{m} doit être “simple” ;
 - se justifie si on considère que les données seules sont insuffisantes.
- Une mesure de simplicité est la longueur euclidienne de \mathbf{m} :

$$L = \mathbf{m}^T \mathbf{m} = \sum m_i^2. \quad (22)$$

- Le problème devient celui de minimiser L sous la contrainte que $\mathbf{e} = \mathbf{d} - \mathbf{Gm} = 0$.
- La méthode des multiplicateurs de Lagrange permet de trouver la solution.
- La fonction à minimiser est

$$\Phi(\mathbf{m}) = L + \sum_{i=0}^{N-1} \lambda_i e_i = \sum_{i=0}^{M-1} m_i^2 + \sum_{i=0}^{N-1} \lambda_i \left[d_i - \sum_{j=0}^{M-1} G_{ij} m_j \right] \quad (23)$$

où λ_i sont les multiplicateurs de Lagrange.

- Le minimum est obtenu en dérivant par rapport à m

$$\frac{\partial \Phi}{\partial m_q} = \sum_{i=0}^{M-1} 2 \frac{\partial m_i}{\partial m_q} m_i - \sum_{i=0}^{N-1} \lambda_i \sum_{j=0}^{M-1} G_{ij} \frac{\partial m_j}{\partial m_q} = 2m_q - \sum_{i=0}^{N-1} \lambda_i G_{iq} \quad (24)$$

- En égalant (24) à zéro, on obtient, sous forme matricielle

$$2\mathbf{m} = \mathbf{G}^T \boldsymbol{\lambda} \quad (25)$$

- En insérant dans $\mathbf{d} = \mathbf{Gm}$, on trouve

$$\boldsymbol{\lambda} = 2 \left[\mathbf{G} \mathbf{G}^T \right]^{-1} \mathbf{d} \quad (26)$$

qui nous permet de finalement trouver, *l'estimateur de longueur minimum*

$$\mathbf{m}^{\text{est}} = \mathbf{G}^T \left[\mathbf{G} \mathbf{G}^T \right]^{-1} \mathbf{d} \quad (27)$$

Régression linéaire

Hypothèses *a priori*

Problème purement indéterminé

Problème partiellement indéterminé

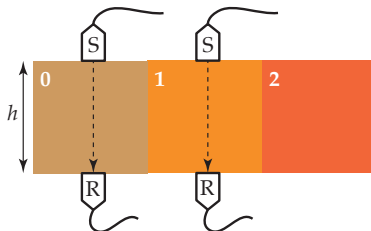
Pondération

Égalité

Variance des paramètres

Inverse généralisée

- Trouvez les paramètres du modèle de la figure suivante, pour $h = 2$ et $\mathbf{d}^{\text{obs}} = [0.5, 0.46]$.



Régression linéaire

Hypothèses *a priori*

Problème purement indéterminé

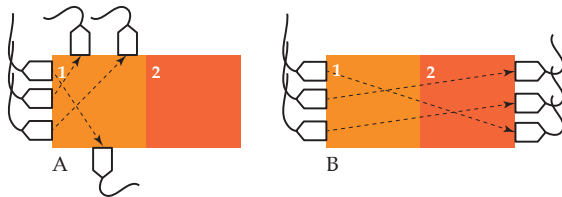
Problème partiellement indéterminé

Pondération

Égalité

Variance des paramètres

Inverse généralisée



- En pratique, les problèmes inverses ne sont jamais complètement surdéterminés ou purement indéterminés.
 - Une cellule du modèle peut être traversée par plusieurs rais alors qu'une autre n'est traversée par aucun rai (A);
 - Si tout les segments de rais sont de la même longueur (B), seulement la lenteur moyenne peut être déterminée.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Si le problème n'est pas trop indéterminé, on peut minimiser une combinaison de l'erreur de prédiction et de la longueur du modèle (indépendamment des paramètres individuels) :

$$\Phi(\mathbf{m}) = E + \varepsilon^2 L = \mathbf{e}^T \mathbf{e} + \varepsilon^2 \mathbf{m}^T \mathbf{m}, \quad (28)$$

où le poids ε^2 détermine l'importance relative de L par rapport à E .

- Si ε est très élevé, l'emphasis est mise sur la partie indéterminée
 - se fait au détriment de $E \rightarrow$ le modèle estimé sera loin du modèle vrai.
- Si ε est très faible, l'information *a priori* n'est pas propagée et la partie indéterminée le reste.
- En général, on cherche ε par essai-erreur.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- En minimisant $\Phi(\mathbf{m})$ par rapport aux paramètres du modèle, on trouve

$$[\mathbf{G}^T \mathbf{G} + \varepsilon^2 \mathbf{I}] \mathbf{m}^{\text{est}} = \mathbf{G}^T \mathbf{d} \quad (29)$$

que l'on récrit

$$\mathbf{m}^{\text{est}} = [\mathbf{G}^T \mathbf{G} + \varepsilon^2 \mathbf{I}]^{-1} \mathbf{G}^T \mathbf{d} \quad (30)$$

- \mathbf{m}^{est} est nommé solution des moindres-carrés *amortis* (*damped least squares*).
- La solution est stabilisée par l'amortissement, et on dit que le problème est *régularisé*.
 - On retrouve le terme *régularisation de Tikhonov* pour décrire ce type d'utilisation d'information *a priori*.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

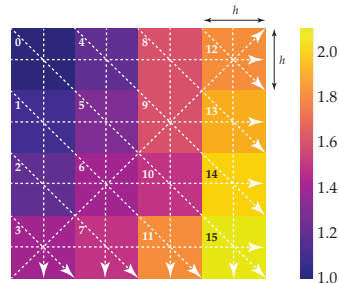
Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Examinons un exemple de problème partiellement indéterminé.
- Le modèle comporte 16 paramètres ;
- La taille h vaut 2 ;
- 16 mesures ont été effectuées.



Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Définition du modèle et des points de mesure.

```
mtrue = np.array([1.0, 1.1, 1.2, 1.4,
                  1.2, 1.3, 1.4, 1.5,
                  1.6, 1.6, 1.5, 1.8,
                  1.8, 1.9, 2.0, 2.1])
```

```
Tx = h*np.array([[0.0, 0.5], [0.0, 1.5],
                 [0.0, 2.5], [0.0, 3.5],
                 [0.5, 0.0], [1.5, 0.0],
                 [2.5, 0.0], [3.5, 0.0],
                 [0.0, 3.0], [0.0, 2.0],
                 [0.0, 1.0], [0.0, 0.0],
                 [1.0, 0.0], [2.0, 0.0],
                 [3.0, 0.0], [0.0, 4.0]])
```

```
Rx = h*np.array([[4.0, 0.5], [4.0, 1.5],
                 [4.0, 2.5], [4.0, 3.5],
                 [0.5, 4.0], [1.5, 4.0],
                 [2.5, 4.0], [3.5, 4.0],
                 [1.0, 4.0], [2.0, 4.0],
                 [3.0, 4.0], [4.0, 4.0],
                 [4.0, 3.0], [4.0, 2.0],
                 [4.0, 1.0], [4.0, 0.0]])
```

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Construction de la matrice G .

```
G = np.zeros((16, nx*nz))

G[0, :4] = h
G[1, 1::4] = h
G[2, 2::4] = h
G[3, 3::4] = h
G[4, :4] = h
G[5, 4:8] = h
G[6, 8:12] = h
G[7, 12:16] = h
G[8, 3] = np.sqrt(2*h*h)
G[9, 2:8:5] = np.sqrt(2*h*h)
G[10, 1:12:5] = np.sqrt(2*h*h)
G[11, :5] = np.sqrt(2*h*h)
G[12, 4::5] = np.sqrt(2*h*h)
G[13, 8::5] = np.sqrt(2*h*h)
G[14, 12] = np.sqrt(2*h*h)
G[15, 3:13:3] = np.sqrt(2*h*h)
```

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Générez les données et ajoutez un bruit gaussien avec $\sigma^2 = 0.05$
- Comparez la solution des moindres-carrés ordinaires avec les moindres-carrés amortis pour
 - $\varepsilon = 10$
 - $\varepsilon = 1$
 - $\varepsilon = 0.1$
 - $\varepsilon = 0.001$
 - $\varepsilon = 10^{-15}$

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Dans plusieurs cas, la longueur $L = \mathbf{m}^T \mathbf{m}$ n'est pas une mesure appropriée de la simplicité du modèle;
- Par exemple, si on cherche à évaluer les fluctuations par rapport à une moyenne connue
 - il est préférable de minimiser la distance par rapport à cette moyenne $\langle \mathbf{m} \rangle$, i.e.

$$L = (\mathbf{m} - \langle \mathbf{m} \rangle)^T (\mathbf{m} - \langle \mathbf{m} \rangle) \quad (31)$$

- Dans d'autres cas, on sait que le modèle est continu et varie lentement spatialement
 - on peut alors minimiser
 - l'inclinaison (*steepness*) : dérivée première de \mathbf{m}
 - la rugosité (*roughness*) : dérivée seconde de \mathbf{m}

Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- L'inclinaison ou la rugosité peuvent être calculées à partir d'une matrice \mathbf{D} telle que (pour l'inclinaison)

$$\mathbf{D}\mathbf{m} = \frac{1}{\Delta x} \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_{M-1} \end{bmatrix} \quad (32)$$

- Pour la rugosité, les lignes contiennent $(\Delta x)^{-2} [\dots \quad 1 \quad -2 \quad 1 \quad \dots]$
- Le terme à minimiser est alors

$$L = (\mathbf{D}\mathbf{m})^T (\mathbf{D}\mathbf{m}) = \mathbf{m}^T \mathbf{D}^T \mathbf{D} \mathbf{m} = \mathbf{m}^T \mathbf{W}_m \mathbf{m} \quad (33)$$

- La matrice \mathbf{W}_m donne un poids différent aux paramètres du modèle.

Régression linéaire

Hypothèses *a priori*

Problème purement indéterminé

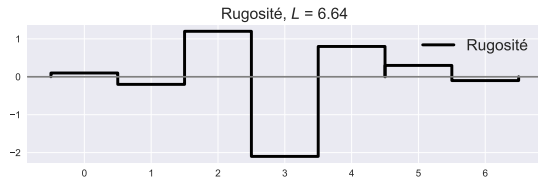
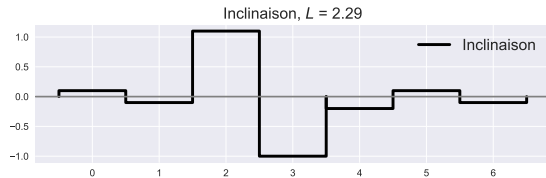
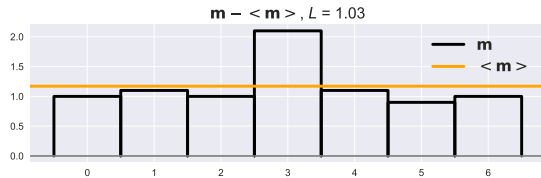
Problème partiellement indéterminé

Pondération

Égalité

Variance des paramètres

Inverse généralisée



Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- La mesure de la simplicité du modèle peut être généralisée à

$$L = (\mathbf{m} - \langle \mathbf{m} \rangle)^T \mathbf{W}_m (\mathbf{m} - \langle \mathbf{m} \rangle) \quad (34)$$

- D'une façon similaire, il est possible de pondérer certains terme de l'erreur de prédiction;
 - utile lorsque certaines mesures sont plus précises que d'autres.
- L'erreur de prédiction généralisée s'écrit alors

$$E = \mathbf{e}^T \mathbf{W}_e \mathbf{e}. \quad (35)$$

- \mathbf{W}_e est généralement une matrice diagonale;
 - Par exemple, pour 5 mesures où on sait que la 3^e est deux fois plus précise, on aura

$$\mathbf{W}_e = \begin{bmatrix} 1 & & & & \\ & 1 & & & \\ & & 2 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \quad (36)$$

Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- La solution des moindres-carrés pondérés, i.e. lorsque $E = \mathbf{e}^T \mathbf{W}_e \mathbf{e}$, vaut

$$\mathbf{m}^{\text{est}} = \left(\mathbf{G}^T \mathbf{W}_e \mathbf{G} \right)^{-1} \mathbf{G}^T \mathbf{W}_e \mathbf{d}. \quad (37)$$

- Lorsque le système est partiellement indéterminé, l'amortissement est inclus et la solution est

$$\mathbf{m}^{\text{est}} = \left(\mathbf{G}^T \mathbf{W}_e \mathbf{G} + \varepsilon^2 \mathbf{W}_m \right)^{-1} \left(\mathbf{G}^T \mathbf{W}_e \mathbf{d} + \varepsilon^2 \mathbf{W}_m \langle \mathbf{m} \rangle \right) \quad (38)$$

- Pour résoudre ce système, on peut le simplifier en posant

$$\mathbf{F} = \begin{bmatrix} \mathbf{W}_e^{1/2} \mathbf{G} \\ \varepsilon \mathbf{D} \end{bmatrix} \quad \text{et} \quad \mathbf{f} = \begin{bmatrix} \mathbf{W}_e^{1/2} \mathbf{d} \\ \varepsilon \mathbf{D} \langle \mathbf{m} \rangle \end{bmatrix} \quad (39)$$

- Il suffit alors de résoudre $\mathbf{F} \mathbf{m}^{\text{est}} = \mathbf{f}$ par la méthode des moindres-carrés ordinaire : $\mathbf{m}^{\text{est}} = (\mathbf{F}^T \mathbf{F})^{-1} \mathbf{F}^T \mathbf{f}$.

Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- Le problème est de retrouver une fonction sinus à partir de points aléatoirement distribués.

```
M = 101
Dz = 1.0
z = Dz*np.arange(M)
zmax = z.max()
mtrue = np.sin(3*np.pi*z/zmax)
```

- Les observations sont :

```
ind = np.array([0, 8, 14, 16, 36, 48, 60, 72, 84, 90, M-1])
N = ind.size

zobs = z[ind]
sigmad = 0.0 # pas de bruit dans les données
dobs = np.sin(3*np.pi*zobs/zmax) + \
    sigmad*np.random.randn(N)
```

- Pour simplifier, on attribue un poids égal à chaque observation, i.e. W_e est une matrice identité.

Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- Nous avons $M=101$ et $N=11$, le système est indéterminé ;
- On sait qu'une fonction sinus est lisse, on peut minimiser la rugosité.
- **G** contient simplement des 1 aux indices des points de mesure.

```
i = np.arange(N)
j = ind
s = np.ones(i.shape)
G = sp.coo_matrix((s, (i, j)), shape=(N, M))
```

- La matrice de rugosité **D** (de taille $M \times M$) contient les termes $(\Delta x)^{-2}[\dots 1 \ -2 \ 1 \ \dots]$ centrés sur le paramètre où la dérivée est évaluée.
 - Aux extrémités, on utilise une dérivée première.
- Construisez **D** et résolvez pour trois valeurs de ε , soit 1.0, 0.01, 100.0.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Il arrive parfois qu'on
 - connaisse la valeur du modèle en un point donné ;
 - sache qu'une certaine fonction des paramètres est égale à une constante.
- On peut exprimer ces contraintes sous la forme $\mathbf{Hm} = \mathbf{h}$, par exemple :
 - la moyenne des paramètres est égale à h_0 :

$$\mathbf{Hm} = \frac{1}{M} [1 \ 1 \ \dots \ 1] \begin{bmatrix} m_0 \\ m_1 \\ \vdots \\ m_{M-1} \end{bmatrix} = [h_0] = \mathbf{h} \quad (40)$$

- Une valeur donnée m_k est connue :

$$\mathbf{Hm} = [0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0] \begin{bmatrix} m_0 \\ \vdots \\ m_k \\ \vdots \\ m_{M-1} \end{bmatrix} = [h_k] = \mathbf{h} \quad (41)$$

Régression linéaire

 Hypothèses *a priori*

 Problème purement
indéterminé

 Problème partiellement
indéterminé

Pondération

Égalité

 Variance des
paramètres

Inverse généralisée

- La méthode des multiplicateurs de Lagrange permet de trouver la solution.
- On minimise E avec la contrainte que $\mathbf{Hm} - \mathbf{h} = 0$ en formant la fonction suivante :

$$\Phi(m) = \sum_{i=0}^{N-1} \left[\sum_{j=0}^{M-1} G_{ij}m_j - d_i \right]^2 + 2 \sum_{i=0}^{p-1} \lambda_i \left[\sum_{j=0}^{M-1} H_{ij}m_j - h_i \right] \quad (42)$$

où p est le nombre de contraintes.

- Les dérivées par rapport aux paramètres,

$$\frac{\partial \Phi(m)}{\partial m_q} = 2 \sum_{i=0}^{M-1} m_i \sum_{j=0}^{N-1} G_{jq}G_{ji} - 2 \sum_{i=0}^{N-1} G_{iq}d_i + 2 \sum_{i=0}^{p-1} \lambda_i H_{iq}, \quad (43)$$

sont égalées à zéro pour trouver le minimum.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Sous forme matricielle, le système d'équation est

$$\underbrace{\begin{bmatrix} \mathbf{G}^T \mathbf{G} & \mathbf{H}^T \\ \mathbf{H} & 0 \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} \mathbf{m} \\ \lambda \end{bmatrix}}_{\mathbf{x}} = \underbrace{\begin{bmatrix} \mathbf{G}^T \mathbf{d} \\ \mathbf{h} \end{bmatrix}}_{\mathbf{b}} \quad (44)$$

- Ce système est habituellement résolu avec un solveur itératif.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- La résolution avec les multiplicateurs de Lagrange se prête mal à la situation où on souhaite appliquer une pondération au modèle ;
 - on pourrait par exemple vouloir lisser le modèle en plus d'imposer une contrainte d'égalité.
- Une approche par moindres-carrés amortis est possible, il suffit d'ajouter les termes appropriés :

$$\mathbf{F} = \begin{bmatrix} \mathbf{W}_e^{1/2} \mathbf{G} \\ \varepsilon \mathbf{D} \\ \gamma \mathbf{H} \end{bmatrix} \quad \text{et} \quad \mathbf{f} = \begin{bmatrix} \mathbf{W}_e^{1/2} \mathbf{d} \\ \varepsilon \mathbf{D} \langle \mathbf{m} \rangle \\ \gamma \mathbf{h} \end{bmatrix} \quad (45)$$

où γ permet d'ajuster la pondération de la contrainte d'égalité.

Régression linéaire

Hypothèses *a priori*

Problème purement
indéterminé

Problème partiellement
indéterminé

Pondération

Égalité

Variance des
paramètres

Inverse généralisée

- Problème : ajuster une droite devant passer par un point connu (z' , d').
- Les paramètres du modèle sont l'ordonnée à l'origine m_0 et la pente m_1 ;
 - et la contrainte est que $d' = m_0 + m_1 z'$.
- Les données sont :

```
N = 30
zmin = 0
zmax = 10
z = np.sort(zmin + zmax*np.random.rand(N, 1), axis=0)

# d = a + b*z + bruit
a = 2.0
b = 1.0
sd = 0.5
dobs = a + b * z + sd*np.random.randn(N, 1)

# contraintes, z' & d'
zp = 8
dp = 6
```

Régression linéaire

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Variance des paramètres

- Les données contiennent invariablement un bruit qui va entraîner une erreur dans l'estimation des paramètres du modèle
- Comment le bruit dans les données se propage-t-il dans les paramètres ?
- On peut
 - généraliser les estimateurs linéaires vus précédemment à une forme $\mathbf{m}^{\text{est}} = \mathbf{M}\mathbf{d} + \mathbf{v}$
 - quantifier le bruit dans les données par la matrice de covariance $[\text{cov } \mathbf{d}]$
- On peut alors montrer que

$$[\text{cov } \mathbf{m}] = \mathbf{M}[\text{cov } \mathbf{d}]\mathbf{M}^T \quad (46)$$

- On assume souvent que les données sont non corrélées et qu'elles ont une la même variance σ_d^2 ;
- Les covariance de paramètres pour les moindres-carrés vaut alors

$$[\text{cov } \mathbf{m}] = \left[\left(\mathbf{G}^T \mathbf{G} \right)^{-1} \right] \sigma_d^2 \mathbf{I} \left[\left(\mathbf{G}^T \mathbf{G} \right)^{-1} \right]^T = \sigma_d^2 \left(\mathbf{G}^T \mathbf{G} \right)^{-1} \quad (47)$$

- Pour l'estimateur de longueur minimum nous avons

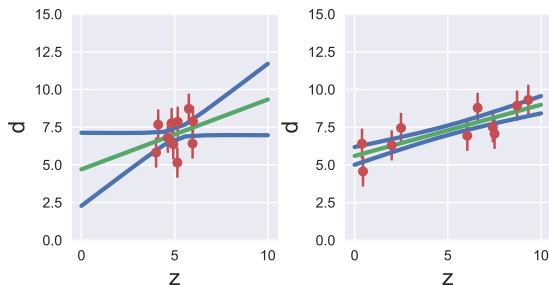
$$\begin{aligned} [\text{cov } \mathbf{m}] &= \left[\mathbf{G}^T \left(\mathbf{G} \mathbf{G}^T \right)^{-1} \right] \sigma_d^2 \mathbf{I} \left[\mathbf{G}^T \left(\mathbf{G} \mathbf{G}^T \right)^{-1} \right]^T \\ &= \sigma_d^2 \mathbf{G}^T \left(\mathbf{G} \mathbf{G}^T \right)^{-2} \mathbf{G} \end{aligned} \quad (48)$$

- Un problème se pose pour estimer σ_d^2 ;
 - On peut se baser sur la résolution des appareils de mesures, e.g. un gravimètre précis à $\pm 5 \mu\text{Gal}$, on parle alors de *variance a priori* ;
 - On peut aussi se baser sur la distribution des erreurs de prédiction \mathbf{e} obtenues après inversion (*a posteriori*), avec

$$\sigma_d^2 \approx \frac{1}{N - M} \sum_{i=0}^{N-1} e_i^2. \quad (49)$$

- La variance *a posteriori* tend cependant à être surestimée en raison des imprécisions du modèle.
- Le constat final demeure néanmoins : *les paramètres du modèles sont corrélés et de variance inégale.*
- L'opérateur \mathbf{G} joue un rôle central dans la propagation des erreurs.

- Exemple de l'influence de G sur la variance des paramètres :
- la variance des données est la même pour tout les points ;
- l'étalement des coordonnées en z dicte la variance des paramètres (courbes bleues = 1σ).



Régression linéaire

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Résolution

Inverse généralisée

- La **décomposition en valeurs singulières** (SVD) permet d'étudier et de résoudre les problèmes indéterminés et mal conditionnés ;
- Pour une matrice \mathbf{G} de taille $N \times M$, la SVD est

$$\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T \quad (50)$$

où

- \mathbf{U} est une matrice $N \times N$ orthogonale où les colonnes forment les vecteurs de base de l'espace des données ;
- \mathbf{V} est une matrice $M \times M$ orthogonale où les colonnes forment les vecteurs de base de l'espace des paramètres ;
- \mathbf{S} est une matrice $N \times M$ diagonale contenant les valeurs singulières de \mathbf{G} .

- Les valeurs singulières sont habituellement classées en ordre décroissant sur la diagonale de \mathbf{S} ;
- Certaines valeurs singulières peuvent être égales à zéro, ce qui fait qu'on peut partitionner \mathbf{S} selon

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_p & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad (51)$$

où p est le nombre de valeurs non nulles.

- Similairement, \mathbf{U} et \mathbf{V} peuvent être partitionnées par colonnes, selon $[\mathbf{U}_p \ \mathbf{U}_0]$ et $[\mathbf{V}_p \ \mathbf{V}_0]$, pour ne garder que les colonnes non multipliées par la partie nulle de \mathbf{S} ;
- On a alors la forme compacte

$$\mathbf{G} = \mathbf{U}_p \mathbf{S}_p \mathbf{V}_p^T \quad (52)$$

Régression linéaire

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Résolution

- Les colonnes de \mathbf{U}_p sont dans l'espace colonne de $R(\mathbf{G})$ et sont linéairement indépendantes.
- Comme il y a p vecteurs dans la base, le rang de \mathbf{G} est p .
- On peut montrer que $N(\mathbf{G}^T) + R(\mathbf{G}) = R^n$, et que les $N - p$ colonnes de \mathbf{U}_0 forment la base du noyau de \mathbf{G}^T .
- On nomme $N(\mathbf{G}^T)$ le noyau des données.
- Similairement, on nomme $N(\mathbf{G})$ le noyau du modèle.

- La SVD peut être utilisée pour calculer l'inverse généralisée de \mathbf{G} , aussi appelée pseudo-inverse de Moore-Penrose :

$$\mathbf{G}^{\dagger} = \mathbf{V}_p \mathbf{S}_p^{-1} \mathbf{U}_p^T \quad (53)$$

- La solution est alors

$$\mathbf{m}_{+} = \mathbf{G}^{\dagger} \mathbf{d} \quad (54)$$

- Une propriété intéressante de (54) est que \mathbf{G}^{\dagger} existe toujours, et donc qu'une solution existe toujours.
 - Les valeurs singulières nulles "correspondent" aux colonnes de \mathbf{G} linéairement dépendantes, la SVD "filtre" pour ne garder que les colonnes indépendantes.

Régression linéaire

Hypothèses *a priori*Variance des
paramètres

Inverse généralisée

Résolution

On peut montrer que

- Lorsque $N = M = p$, $\mathbf{G}^+ = \mathbf{G}^{-1}$ et la solution est unique et les paramètres s'ajustent parfaitement aux données.
- Lorsque $N = p$ et $p < M$, \mathbf{G}^+ est équivalent à la solution de longueur minimum. Pour des raisons de précision numérique, on favorise en pratique l'utilisation de la SVD pour solutionner le système.
- Lorsque $M = p$ et $p < N$, \mathbf{G}^+ est équivalent à la solution des moindres-carrés.
- Lorsque $p < N$ et $p < M$, \mathbf{G}^+ est équivalent à la solution de longueur minimum.

- On a vu que la covariance des paramètres est

$$[\text{cov } \mathbf{m}] = \sigma_d^2 (\mathbf{G}^T \mathbf{G})^{-1}$$
- Pour l'inverse généralisée on a

$$[\text{cov } \mathbf{m}_+] = \mathbf{G}^+ [\text{cov } \mathbf{d}] (\mathbf{G}^+)^T \quad (55)$$

$$= \sigma_d^2 \mathbf{G}^+ (\mathbf{G}^+)^T \quad (56)$$

$$= \sigma_d^2 \mathbf{V}_p \mathbf{S}_p^{-2} \mathbf{V}_p^T \quad (57)$$

$$= \sigma_d^2 \sum_{i=0}^{p-1} \frac{V_{:,i} V_{:,i}^T}{s_i^2} \quad (58)$$

- Malheureusement, \mathbf{m}_+ *n'est pas* un estimateur non biaisé de \mathbf{m}_{vrai} (sauf si $p = M$)
 - Cela est dû au fait que \mathbf{m}_{vrai} peut contenir des projections non nulles dans des vecteurs de base de \mathbf{V} qui ne sont pas utilisés par l'inverse généralisée (portion tronquée).
- On peut quantifier ce biais avec la **matrice de résolution du modèle**;
 - permet de déterminer à quel point \mathbf{m}_+ s'approche de \mathbf{m}_{vrai} , en *assumant qu'il n'y a pas d'erreur dans les données*.
- Partant de \mathbf{m}_{vrai} , on a que $\mathbf{d}_{\text{vrai}} = \mathbf{G}\mathbf{m}_{\text{vrai}}$ et donc que

$$\mathbf{m}_+ = \mathbf{G}^\dagger \mathbf{d}_{\text{vrai}} \quad (59)$$

$$= \mathbf{G}^\dagger \mathbf{G} \mathbf{m}_{\text{vrai}} \quad (60)$$

$$= \mathbf{R}_m \mathbf{m}_{\text{vrai}} \quad (61)$$

- \mathbf{R}_m permet donc de quantifier à quel point \mathbf{m}_+ s'approche de \mathbf{m}_{vrai} ;
 - si \mathbf{R}_m est une matrice identité, le modèle vrai peut être retrouvé parfaitement et la résolution est "parfaite".
- En pratique, on examine la diagonale de \mathbf{R}_m pour voir si les éléments sont proches de 1 ;
 - si c'est le cas, les paramètres correspondants sont bien résolus ;
 - dans le cas inverse, les paramètres sont une moyenne pondérée des paramètres vrais.
- On peut aussi mener un test de résolution avec un modèle impulsional \mathbf{m}_i (vecteur de 0 avec un seul élément i égal à 1) ;
 - Le produit de \mathbf{R}_m avec \mathbf{m}_i fait ressortir la contribution des colonnes de \mathbf{R}_m sur le i^{e} paramètre.

Régression linéaire

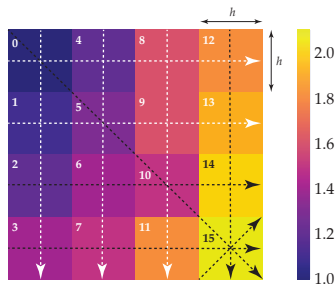
Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Résolution

- Examinons la signification de la matrice de résolution avec un exemple en tomographie.
- Le modèle comporte 16 paramètres ;
- La taille h vaut 2 ;
- 10 mesures ont été effectuées.



Régression linéaire

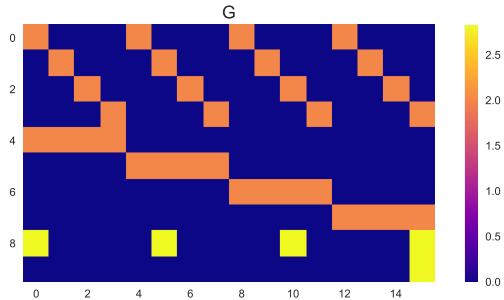
Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Résolution

- La matrice G a la forme suivante :



- Le rang de la matrice est 9.

Régression linéaire

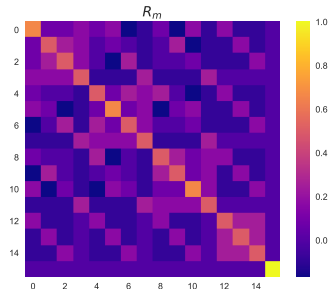
Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Résolution

- La matrice de résolution contient les éléments les plus élevés sur sa diagonale.
- La résolution est 1 seulement pour le 16^e paramètre.
- Les autres paramètres contiennent des contributions des cellules voisines.



Régression linéaire

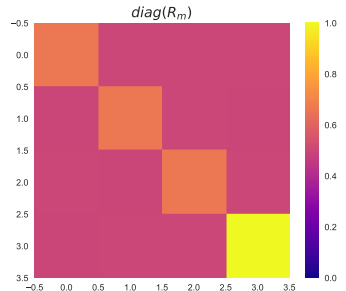
Hypothèses *a priori*

Variance des
paramètres

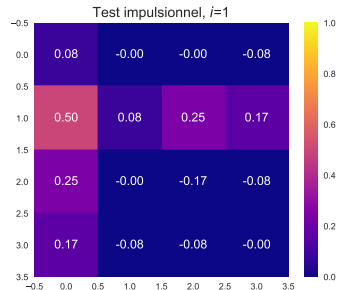
Inverse généralisée

Résolution

- La résolution est plus élevée aux cellules traversés par le long rai oblique.



- Test impulsionnel pour $\mathbf{m}_i = [0 \ 1 \ 0 \ 0 \ \dots \ 0]$
- La 2^e cellule ne peut être complètement distinguée de ses voisines ;
- Les cellules traversés par le long rai oblique contribuent moins.



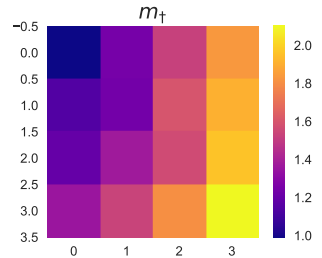
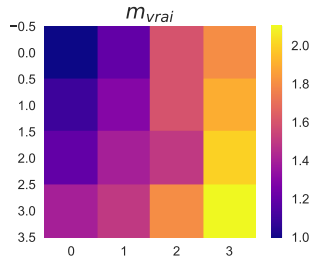
Régression linéaire

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée
Résolution

- Malgré la résolution imparfaite, le modèle estimé est proche du modèle vrai.



- Idéalement, on voudrait que \mathbf{m}_+ nous permette de retrouver exactement les données observées.
- D'une façon similaire à la résolution du modèle, on peut évaluer individuellement le poids des données observées dans les données prédites par \mathbf{m}_+ .
- Soit \mathbf{d}_+ le vecteur des données produit par \mathbf{m}_+ , i.e.

$$\mathbf{d}_+ = \mathbf{G}\mathbf{m}_+ \quad (62)$$

- Puisque $\mathbf{m}_+ = \mathbf{G}^+\mathbf{d}$, on a que

$$\mathbf{d}_+ = \mathbf{G}\mathbf{G}^+\mathbf{d} \quad (63)$$

$$= \mathbf{R}_d\mathbf{d} \quad (64)$$

- Si $\mathbf{R}_d = \mathbf{I}$, l'erreur de prédiction est nulle.
- À l'inverse, \mathbf{R}_d donne une mesure de la capacité de l'estimateur à reproduire les données ;
- Si par exemple \mathbf{R}_d contient une ligne égale à

$$[\dots 0000.10.80.1000 \dots]$$

où 0.8 apparait sur le i^e colonne, alors

$$d_i^{\text{pre}} = \sum_j R_d(i, j) d_j^{\text{obs}} = 0.1 d_{i-1}^{\text{obs}} + 0.8 d_i^{\text{obs}} + 0.1 d_{i+1}^{\text{obs}} \quad (65)$$

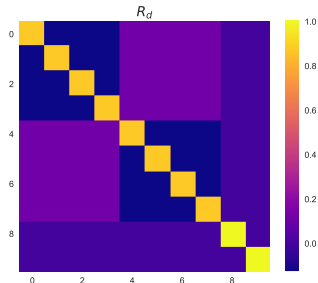
Régression linéaire

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée
Résolution

- Examinons R_d pour l'exemple précédent
- Les valeurs sur la diagonale sont assez proches de 1, sauf pour les 8^e et 9^e données où $R_d \approx 1$
- Pour les 7 autres données, il y a une composante non nulle des autres termes ;
 - les données prédites sont une moyenne pondérée des données observées



Régression linéaire

Hypothèses *a priori*

Variance des
paramètres

Inverse généralisée

Résolution

- Il est important de rappeler que \mathbf{R}_m et \mathbf{R}_d *ne dépendent pas* des données et des modèles, mais qu'elles sont dues *exclusivement* à \mathbf{G} ;
- Ces matrices sont donc le *reflet de la physique du problème et de la géométrie d'acquisition des données*;
- En pratique, la capacité à retrouver \mathbf{m}_{vrai} dépend autant de la résolution que de la propagation du bruit dans les paramètres du modèle.
- \mathbf{R}_m et \mathbf{R}_d sont des outils très pratiques pour la conception des géométries d'acquisition.