

Preuves

Paramètres effectifs et modèles non-linéaires

Alexandre CONSTANTIN¹ Rodrigo CABRAL FARIAS² Jean-Marc BROSSIER¹ Olivier MICHEL¹

¹Univ. Grenoble-Alpes, CNRS, Grenoble-INP, GIPSA Lab, 38000 Grenoble

²Univ. Côte d'Azur, CNRS, Laboratoire I3S, Sophia-Antipolis 06900

Résumé – Nous présentons une synthèse de la littérature autour des questions de degrés de liberté et en proposons un estimateur applicable à tout modèle paramétrique. Des résultats sur un perceptron avec une couche cachée montrent des performances équivalentes à l'algorithme de Ye pour une complexité numérique réduite.

Abstract – A summary of the litterature on degrees of freedom is presented and an estimator for non-linear models is proposed. Results on a one hidden layer perceptron show equivalent behavior as Ye's algorithm, with lower numerical complexity.

A Degrés de liberté en linéaire

Afin de montrer l'égalité (3) :

$$dP = \sum_{i=1}^P \frac{\lambda_i}{\lambda_i + \eta},$$

dans le cas linéaire, où $\{\lambda_i\}_{i=1,\dots,P}$ sont les valeurs propres de la Hessienne de la fonction de coût sans régularisation et η la force de la régularisation 'ridge', nous rappelons que :

$$\mathbf{S}_\eta = \mathbf{X} (\mathbf{X}^\top \mathbf{X} + \eta \mathbf{I})^{-1} \mathbf{X}^\top,$$

avec $\mathbf{X} \in \mathbb{R}^{N \times P}$ et \mathbf{I} la matrice identité, ici de taille P .

La matrice $\mathbf{X}^\top \mathbf{X}$ étant symétrique définie positive, elle admet une décomposition en valeur propre $\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1}$ telle que $\mathbf{\Lambda}$ est la matrice diagonale contenant les P valeurs propres $\{\lambda_i\}_{i=1}^P$ de $\mathbf{X}^\top \mathbf{X}$. Il vient alors, en utilisant la linéarité cyclique et la décomposition en valeur propre :

$$\begin{aligned} \text{Tr}(\mathbf{S}_\eta) &= \text{Tr} \left(\mathbf{X} (\mathbf{X}^\top \mathbf{X} + \eta \mathbf{I})^{-1} \mathbf{X}^\top \right) \\ &= \text{Tr} \left(\mathbf{X}^\top \mathbf{X} (\mathbf{X}^\top \mathbf{X} + \eta \mathbf{I})^{-1} \right) \\ &= \text{Tr} \left(\mathbf{P}\mathbf{\Lambda}\mathbf{P}^{-1} (\mathbf{P}(\mathbf{\Lambda} + \eta \mathbf{I})\mathbf{P}^{-1})^{-1} \right) \\ &= \text{Tr} \left(\mathbf{\Lambda}\mathbf{P}^{-1}\mathbf{P} (\mathbf{\Lambda} + \eta \mathbf{I})^{-1} \mathbf{P}^{-1}\mathbf{P} \right) \\ &= \text{Tr} \left(\mathbf{\Lambda} (\mathbf{\Lambda} + \eta \mathbf{I})^{-1} \right) \\ &= \sum_{i=1}^P \frac{\lambda_i}{\lambda_i + \eta}. \end{aligned}$$

Pour démontrer l'égalité entre la définition (4) des degrés de liberté et dP , nous insistons sur le fait que les échantillons doivent être indépendants, en effet :

$$\begin{aligned} df &= \frac{1}{\sigma_y^2} \sum_{i=1}^N \text{cov}(y_i, \sum_{k=1}^N (\mathbf{S}_\eta)_{i,k} y_k) \\ &= \frac{1}{\sigma_y^2} \sum_{i=1}^N \sum_{k=1}^N (\mathbf{S}_\eta)_{i,k} \text{cov}(y_i, y_k) \end{aligned}$$

en utilisant $\hat{y} = \mathbf{S}_\eta \mathbf{y}$, la linéarité à droite de la covariance (linéarité d'une espérance) et $\text{cov}(y_i, y_k) = 0$ pour $i \neq k$ par indépendance. Il en découle naturellement :

$$\begin{aligned} df &= \frac{1}{\sigma_y^2} \sum_{i=1}^N (\mathbf{S}_\eta)_{i,i} \text{cov}(y_i, y_i) \\ &= \frac{1}{\sigma_y^2} \sum_{i=1}^N (\mathbf{S}_\eta)_{i,i} \sigma_y^2 \\ &= \text{Tr}(\mathbf{S}_\eta) \\ &= dP, \end{aligned}$$

B Preuve de la proposition 1

L'estimateur \hat{df} est un estimateur de :

$$df' = \text{Tr} \left[\mathbb{E} \left(-\mathbf{J}_\theta [m_\theta(\mathbf{y})] [\mathbf{H}_\theta \mathcal{L}]^{-1} \mathbf{J}_y [\nabla_\theta \mathcal{L}] \right) \right],$$

où $\mathbf{J}_x[\mathbf{y}] \in \mathbb{R}^{m \times n}$ est la matrice jacobienne telle que, pour $\mathbf{y} \in \mathbb{R}^m$ et $\mathbf{x} \in \mathbb{R}^n$, $(\mathbf{J}_x[\mathbf{y}])_{i,j} = \partial y_i / \partial x_j$.

Nous voulons donc montrer que :

$$df = \sum_{i=1}^N \mathbb{E} \left(\frac{\partial \hat{y}_i}{\partial y_i} \right) = df'.$$

Or $\sum_{i=1}^N \mathbb{E} \left(\frac{\partial \hat{y}_i}{\partial y_i} \right) = \text{Tr} [\mathbb{E} (\mathbf{J}_y \hat{\mathbf{y}})]$, c'est-à-dire, par égalité des dimensions des matrices, montrer que :

$$\mathbf{J}_y[\hat{\mathbf{y}}] = -\mathbf{J}_\theta [m_\theta(\mathbf{y})] [\mathbf{H}_\theta \mathcal{L}]^{-1} \mathbf{J}_y [\nabla_\theta \mathcal{L}].$$

Pour cela nous rappelons que les sorties prédites sont fonction du modèle, à \mathbf{X} fixé, tel que :

$$\hat{\mathbf{y}} = m_{\hat{\theta}}(\mathbf{X}),$$

où $\hat{\theta}$ sont les paramètres, au point de convergence, tel que $\nabla_\theta \mathcal{L}(\hat{\theta}) = \mathbf{0}$. Et $\hat{\theta}$ est une fonction implicite, notée f , de \mathbf{y} (à \mathbf{X} fixé car seules les perturbations en \mathbf{y} nous intéressent).

Or $\mathbf{J}_y[\hat{\mathbf{y}}] = \mathbf{J}_{f(\mathbf{y})}[\hat{\mathbf{y}}] \mathbf{J}_y[f(\mathbf{y})]$ par la composition des fonctions m et f de la matrice Jacobienne. D'une part, $\mathbf{J}_{f(\mathbf{y})}[\hat{\mathbf{y}}] =$

$\mathbf{J}_\theta[m_{\hat{\theta}}(\mathbf{y})]$ est évaluable en $\theta = \hat{\theta}$. D'autre part, par le théorème des fonctions implicites avec $\nabla_\theta \mathcal{L}(\hat{\theta}) = \mathbf{0}$, il vient :

$$\mathbf{J}_y[f(\mathbf{y})] = - \left[J_{f(\mathbf{y})}[\nabla_\theta \mathcal{L}]|_{(\mathbf{y}, \hat{\theta})} \right]^{-1} J_y[\nabla_\theta \mathcal{L}]|_{(\mathbf{y}, \hat{\theta})}.$$

Le terme $J_y[\nabla_\theta \mathcal{L}]$ est évaluable de manière directe car le gradient est connu. Par définition $J_{f(\mathbf{y})}[\nabla_\theta \mathcal{L}]$ est la matrice Hessienne de la fonction de coût évaluée à $\theta = \hat{\theta}$. D'où l'égalité $df' = df$.

Notre estimateur est égal au degrés de liberté dans le cas linéaire (noté dP), car :

$$\begin{cases} \mathcal{L} = \|\mathbf{y} - \mathbf{X}\theta\|_2^2 + \eta \|\theta\|_2^2 \\ \nabla_\theta \mathcal{L} = -2\mathbf{X}^\top(\mathbf{y} - \mathbf{X}\theta) + 2\eta\theta \\ m_\theta = \mathbf{X}\theta, \end{cases}$$

ce qui permet d'obtenir :

$$df = \text{Tr} \left[-\mathbf{X} (2\mathbf{X}^\top \mathbf{X} + 2\eta \mathbf{I})^{-1} (-2\mathbf{X}^\top) \right] = dP.$$

C Cas d'un échantillon identiquement distribué mais non indépendant

Pour terminer, pour la généralisation au cas non-indépendant mais identiquement distribué, pour démontrer (7), nous avons :

$$\begin{aligned} \mathbb{E}(d_\Sigma(\hat{\mathbf{y}}, \boldsymbol{\mu})) &= \mathbb{E}(d_\Sigma(\hat{\mathbf{y}}, \mathbf{y})) + \mathbb{E}(d_\Sigma(\mathbf{y}, \boldsymbol{\mu})) \\ &\quad + 2\mathbb{E}((\hat{\mathbf{y}} - \mathbf{y})^\top \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})), \end{aligned}$$

or, en utilisant la linéarité cyclique de la trace et la linéarité de l'espérance :

$$\begin{aligned} \mathbb{E}(d_\Sigma(\mathbf{y}, \boldsymbol{\mu})) &= \mathbb{E}(\text{Tr}[(\mathbf{y} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})]) \\ &= \text{Tr}(\Sigma^{-1} \mathbb{E}[(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^\top]) \\ &= \text{Tr}(\Sigma^{-1} \Sigma) = N, \end{aligned}$$

et, de manière similaire :

$$\begin{aligned} &\mathbb{E}[(\hat{\mathbf{y}} - \mathbf{y})^\top \Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu})] \\ &= \mathbb{E}[\hat{\mathbf{y}}^\top \Sigma^{-1} \mathbf{y} - \hat{\mathbf{y}}^\top \Sigma^{-1} \boldsymbol{\mu}] - \mathbb{E}[\mathbf{y}^\top \Sigma^{-1} \mathbf{y} - \mathbf{y}^\top \Sigma^{-1} \boldsymbol{\mu}] \\ &= \text{Tr}(\Sigma^{-1} [\mathbb{E}(\hat{\mathbf{y}}^\top \mathbf{y}) - \mathbb{E}(\hat{\mathbf{y}}^\top) \boldsymbol{\mu}]) \\ &\quad - \text{Tr}(\Sigma^{-1} [\mathbb{E}(\mathbf{y}^\top \mathbf{y}) - \mathbb{E}(\mathbf{y}^\top) \boldsymbol{\mu}]) \\ &= \text{Tr}(\Sigma^{-1} \text{cov}(\hat{\mathbf{y}}, \mathbf{y})) - \text{Tr}(\Sigma^{-1} \Sigma) \\ &= df - N. \end{aligned}$$

D Calcul de la Hessienne du perceptron une couche

Nous considérons le perceptron multi-couche décrit par la Figure 1. Il présente un vecteur d'entrées \mathbf{x} et un scalaire en sortie \hat{y} , une fonction (non-linéaire) en couche cachée notée f_h et une fonction de transformation de la sortie (quelconque), notée f_o . Le vecteur de paramètres θ est ordonné de la manière suivante :

$$\theta = \left\{ \bigcup_{i=1}^d \mathbf{w}_i^{(1)} \right\} \cup \{ \mathbf{w}^{(o)} \} \cup \{ \mathbf{b}^{(1)} \} \cup \{ b^{(2)} \}, \quad (1)$$

où $\mathbf{w}_i^{(1)}, \mathbf{b}^{(1)} = [b_1^{(1)}, \dots, b_d^{(1)}]^\top$ sont les poids et les biais de la couche cachée appliquées à l'entrée i . Les sorties après la non-linéarité de la couche cachée sont représentés par le vecteur $\mathbf{o}^{(1)} = [o_1^{(1)}, \dots, o_d^{(1)}]^\top$. $\mathbf{w}^{(o)}$ sont les poids de la couche de sortie avec un biais $b^{(2)}$. Nous considérons aussi le cas d'un échantillon $y \in \mathbf{y}$, car la matrice Hessienne sur l'ensemble des N échantillons est la somme des N matrices Hessiennes.

De plus, nous avons la décomposition suivante pour une fonction de coût \mathcal{L} étant la somme des erreurs quadratiques, pour l'échantillon $y \in \mathbf{y}$:

$$\mathbf{H} = \nabla_\theta \hat{y} (\nabla_\theta \hat{y})^\top + (\hat{y} - y) \mathbf{H}^{yy}, \quad (2)$$

où \mathbf{H}^{yy} est la hessienne de la sortie du modèle et $\nabla_\theta \hat{y}$ le gradient en sortie du modèle. A noter que nous ne prenons pas en compte la constante de la fonction de coût ($1/N$ pour une Erreur Quadratique Moyenne par exemple) car cette constante s'annule dans la Proposition 1 (entre la matrice Hessienne et la matrice Jacobienne du gradient (la matrice Jacobienne du modèle ne dépend pas de la fonction de coût). Par ailleurs, pour des raisons de stabilité numérique, nous utilisons comme constante l'inverse de la norme de Froebenius, reportée à la matrice Jacobienne du gradient.

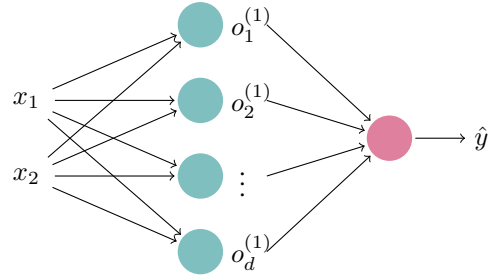


FIGURE 1 : Perceptron avec $d > 1$ neurones.

Le gradient en sortie est exprimé par :

$$\begin{cases} \frac{\partial y}{\partial b^{(2)}} = \alpha_o \\ \nabla_{\mathbf{w}^{(o)}} y = \alpha_o \mathbf{o}^{(1)} \\ \nabla_{\mathbf{b}^{(1)}} y = \alpha_o \boldsymbol{\alpha}_h^{(1)} \odot \mathbf{w}^{(o)} \\ \nabla_{\mathbf{w}_i^{(1)}} y = \alpha_o \alpha_{h,i} w_i^{(o)} \mathbf{x}, \quad i = 1, \dots, d; \end{cases} \quad (3)$$

en introduisant $\boldsymbol{\alpha}_h^{(1)} = [\alpha_{h,1}^{(1)}, \dots, \alpha_{h,d}^{(1)}]^\top$, tel que $\alpha_{h,i}^{(1)} = f'_h(b_i^{(1)} + \langle \mathbf{w}_i^{(1)}, \mathbf{x} \rangle)$, $i = 1, \dots, d$ et $\alpha_o = f'_o(b^{(2)} + \langle \mathbf{w}^{(o)}, \mathbf{o}^{(1)} \rangle)$. Par ailleurs, \odot dénote le produit de Hadamard.

Nous définissons aussi, pour la suite, les quantités $\beta_{h,i}^{(1)} = f''_h(b_i^{(1)} + \langle \mathbf{w}_i^{(1)}, \mathbf{x} \rangle)$, $i = 1, \dots, d$, le vecteur $\boldsymbol{\beta}_h^{(1)} = [\beta_{h,1}^{(1)}, \dots, \beta_{h,d}^{(1)}]^\top$ et $\beta_o = f''_o(b^{(2)} + \langle \mathbf{w}^{(o)}, \mathbf{o}^{(1)} \rangle)$. La hessienne, en θ (1), se décompose alors suivant les blocs représentés dans la Figure 2.

Pour $i, j = 1, \dots, d$, nous retrouvons les quantités suivantes en dérivant les dérivées (3) par rapport à l'autre variable d'in-

$$\mathbf{H}^{yy} = \begin{pmatrix} \mathbf{H}_{\mathbf{w}_1^{(1)} \mathbf{w}_1^{(1)}}^{yy} & \cdots & \mathbf{H}_{\mathbf{w}_1^{(1)} \mathbf{w}_d^{(1)}}^{yy} & \mathbf{H}_{\mathbf{w}_1^{(1)} \mathbf{w}^{(o)}}^{yy} & \mathbf{H}_{\mathbf{w}_1^{(1)} \mathbf{b}^{(1)}}^{yy} & \mathbf{h}_{\mathbf{w}_1^{(1)} b^{(2)}}^{yy} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{H}_{\mathbf{w}_d^{(1)} \mathbf{w}_1^{(1)}}^{yy} & \cdots & \mathbf{H}_{\mathbf{w}_d^{(1)} \mathbf{w}_d^{(1)}}^{yy} & \mathbf{H}_{\mathbf{w}_d^{(1)} \mathbf{w}^{(o)}}^{yy} & \mathbf{H}_{\mathbf{w}_d^{(1)} \mathbf{b}^{(1)}}^{yy} & \mathbf{h}_{\mathbf{w}_d^{(1)} b^{(2)}}^{yy} \\ (\mathbf{H}_{\mathbf{w}_1^{(1)} \mathbf{w}^{(o)}}^{yy})^\top & \cdots & (\mathbf{H}_{\mathbf{w}_d^{(1)} \mathbf{w}^{(o)}}^{yy})^\top & \mathbf{H}_{\mathbf{w}^{(o)} \mathbf{w}^{(o)}}^{yy} & \mathbf{H}_{\mathbf{w}^{(o)} \mathbf{b}^{(1)}}^{yy} & \mathbf{h}_{\mathbf{w}^{(o)} b^{(2)}}^{yy} \\ (\mathbf{H}_{\mathbf{w}_1^{(1)} \mathbf{b}^{(1)}}^{yy})^\top & \cdots & (\mathbf{H}_{\mathbf{w}_d^{(1)} \mathbf{b}^{(1)}}^{yy})^\top & (\mathbf{H}_{\mathbf{w}^{(o)} \mathbf{b}^{(1)}}^{yy})^\top & \mathbf{H}_{\mathbf{b}^{(1)} \mathbf{b}^{(1)}}^{yy} & \mathbf{h}_{\mathbf{b}^{(1)} b^{(2)}}^{yy} \\ (\mathbf{h}_{\mathbf{w}_1^{(1)} b^{(2)}}^{yy})^\top & \cdots & (\mathbf{h}_{\mathbf{w}_d^{(1)} b^{(2)}}^{yy})^\top & (\mathbf{h}_{\mathbf{w}^{(o)} b^{(2)}}^{yy})^\top & (\mathbf{h}_{\mathbf{b}^{(1)} b^{(2)}}^{yy})^\top & h_{b^{(2)}} \end{pmatrix},$$

FIGURE 2 : Matrice Hessienne de la sortie du modèle par bloc.

$$\mathbf{H}^{yy} = c \begin{pmatrix} \text{diag}(\beta_h^{(1)} \odot \mathbf{w}^{(o)}) \otimes \mathbf{x} \mathbf{x}^\top & \text{diag}(\alpha_{h,1}^{(1)}, \dots, \alpha_{h,d}^{(1)}) \otimes \mathbf{x} & \text{diag}(\beta_h^{(1)} \odot \mathbf{w}^{(o)}) \otimes \mathbf{x} & \mathbf{0}_{dp} \\ & \mathbf{O}_d & \text{diag}(\alpha_{h,1}^{(1)}, \dots, \alpha_{h,d}^{(1)}) & \mathbf{0}_d \\ & & \text{diag}(\beta_h^{(1)} \odot \mathbf{w}^{(o)}) & \mathbf{0}_d \\ & & & 0 \end{pmatrix},$$

FIGURE 3 : Matrice Hessienne de la sortie du modèle, par bloc, pour une sortie linéaire.

térêt :

$$\begin{aligned} \mathbf{H}_{\mathbf{w}_i^{(1)} \mathbf{w}_j^{(1)}}^{yy} &= \left(\beta_o (\alpha_{h,i}^{(1)})^2 (\mathbf{w}^{(o)})_i (\mathbf{w}^{(o)})_j + \alpha_o \beta_{h,i}^{(1)} (\mathbf{w}^{(o)})_i \right) \mathbf{x} \mathbf{x}^\top \\ &\quad \text{si } i = j; \\ &= \beta_o \alpha_{h,i}^{(1)} \alpha_{h,j}^{(1)} (\mathbf{w}^{(o)})_i (\mathbf{w}^{(o)})_j \mathbf{x} \mathbf{x}^\top \text{ sinon,} \\ \mathbf{H}_{\mathbf{w}^{(o)} \mathbf{w}^{(o)}}^{yy} &= \beta_o \mathbf{O}^{(1)} \mathbf{O}^{(1)\top}, \\ \mathbf{H}_{\mathbf{b}_i^{(1)} \mathbf{b}_j^{(1)}}^{yy} &= \beta_o (\alpha_{h,i}^{(1)})^2 (\mathbf{w}^{(o)})_i^2 + \alpha_z \beta_{h,i}^{(1)} (\mathbf{w}^{(o)})_i \text{ si } i = j \\ &= \beta_o \alpha_{h,i}^{(1)} \alpha_{h,j}^{(1)} (\mathbf{w}^{(o)})_i (\mathbf{w}^{(o)})_j \text{ sinon,} \end{aligned}$$

et, pour le dernier terme diagonal :

$$h_{b^{(2)}} = \beta_o.$$

Les autres termes sont :

$$\begin{aligned} \mathbf{H}_{\mathbf{w}_i^{(1)} (\mathbf{b}^{(1)})_j}^{yy} &= \frac{\partial}{\partial \mathbf{b}_j^{(1)}} \nabla_{\mathbf{w}_i^{(1)}} y \\ &= \beta_o (\alpha_{h,i}^{(1)})^2 (\mathbf{w}^{(o)})_i^2 \mathbf{x} + \alpha_o \beta_{h,i}^{(1)} (\mathbf{w}^{(o)})_i \mathbf{x} \\ &\quad \text{si } i = j; i, j = 1, \dots, d \\ &= \beta_o \alpha_{h,j}^{(1)} (\mathbf{w}^{(o)})_j \alpha_{h,i}^{(1)} (\mathbf{w}^{(o)})_i \mathbf{x} \text{ sinon,} \\ \mathbf{H}_{\mathbf{w}_i^{(1)} (\mathbf{w}^{(o)})_j}^{yy} &= \frac{\partial}{\partial \mathbf{w}_j^{(o)}} \nabla_{\mathbf{w}_i^{(1)}} y \\ &= \left(\beta_o \alpha_{h,i}^{(1)} (\mathbf{w}^{(o)})_i o_i^{(1)} + \alpha_o \alpha_{h,i}^{(1)} \right) \mathbf{x} \\ &\quad \text{si } i = j \\ &= \beta_o \alpha_{h,i}^{(1)} (\mathbf{w}^{(o)})_i o_j^{(1)} \mathbf{x} \text{ sinon,} \\ (\mathbf{H}_{\mathbf{w}^{(o)} \mathbf{b}^{(1)}}^{yy})_{i,j} &= \frac{\partial^2 y}{\partial \mathbf{w}_i^{(o)} \partial \mathbf{b}_j^{(1)}} \\ &= \beta_o (\mathbf{w}^{(o)})_i \alpha_{h,i}^{(1)} o_i^{(1)} + \alpha_o \alpha_{h,i}^{(1)} \\ &\quad \text{si } i = j \\ &= \beta_o (\mathbf{w}^{(o)})_j \alpha_{h,j}^{(1)} o_i^{(1)} \text{ sinon.} \end{aligned}$$

Finalement les dernières quantités valent :

$$\begin{aligned} \mathbf{h}_{\mathbf{w}_i^{(1)} b^{(2)}}^{yy} &= \beta_o \alpha_{h,i}^{(1)} (\mathbf{w}^{(o)})_i \mathbf{x}, \\ \mathbf{h}_{\mathbf{w}^{(o)} b^{(2)}}^{yy} &= \beta_o \mathbf{O}^{(1)}; \\ (\mathbf{h}_{\mathbf{b}^{(1)} b^{(2)}}^{yy})_i &= \beta_o \alpha_{h,i}^{(1)} (\mathbf{w}^{(o)})_i. \end{aligned}$$

En utilisant l'hypothèse de linéarité de la couche de sortie

Nous considérons maintenant le problème de regression, c'est-à-dire $f_o(x) = cx, \forall x \in \mathbb{R}$, on a alors $\beta_o = 0$ et $\alpha_o = c$.

Les dérivées secondes (matrice symétrique) est donné dans la Figure 3 avec, pour rappel, respectivement \odot et \otimes les produits d'Hadamard et de Kronecker respectivement.

Il ne reste alors plus qu'à recombinaer (3) et les équations de la Figure 3 dans (2).