

CONCLUSION

Dans ce mémoire, nous avons montré l'intérêt que présentent les réseaux de neurones dans le cadre du contrôle non destructif et de la commande du procédé de soudage par points. Il en résulte une méthodologie, clairement définie, pour obtenir, tester et éventuellement améliorer un modèle de prévision du diamètre de bouton valable dans les conditions désirées.

Cette étude nous a amené à orienter nos recherches vers deux aspects théoriques concernant la mise en œuvre des réseaux de neurones, et plus généralement de tout modèle statistique :

1. *comment sélectionner le bon modèle parmi une série de modèles candidats ?* N'ayant pas d'idée a priori sur le nombre minimal de soudures à fournir pour obtenir de bons résultats, nous avons décidé d'axer nos recherches sur le leave-one-out, méthode de validation croisée réputée pour donner de bons résultats avec peu d'exemples,
2. *quel est le domaine de validité d'un modèle ?* Ceci nous a conduit à examiner la notion d'intervalle de confiance sur la sortie d'un modèle, linéaire ou non.

Finalement, il s'avère que ces deux aspects peuvent être abordés dans un même contexte, en utilisant le même outil, à savoir un développement de Taylor de la sortie du modèle au voisinage d'un point de l'espace des paramètres.

Nous avons montré que ce développement de Taylor permet d'estimer l'effet du retrait d'un exemple de la base d'apprentissage sur la solution des moindres carrés, la sortie du modèle, l'intervalle de confiance sur la prédiction de cet exemple, etc. Toutes ces estimations sont obtenues à partir de la grandeur h_{ii} , comprise entre 0 et 1, et qui représente l'influence de l'exemple i sur l'estimation des paramètres du modèle.

Nous avons ensuite étudié l'estimation E_p des performances de généralisation d'un modèle, fondée sur les erreurs de prédiction calculées précédemment. À architecture donnée, nous montrons que la sélection de modèles sur la base de E_p est meilleure que celle résultant d'une mise en œuvre classique du leave-one-out, pour plusieurs raisons : elle évite la sélection de modèles surajustés, ne nécessite aucune hypothèse de stabilité algorithmique, et est bien plus rapide à effectuer. De plus, puisqu'il n'est plus nécessaire d'effectuer autant d'apprentissages qu'il y a d'exemples, cette méthode ne se limite plus aux problèmes où les exemples disponibles sont peu nombreux.

Pour comparer des modèles correspondant à différentes architectures, nous avons montré qu'il est judicieux de considérer la moyenne des intervalles de confiance sur la prédiction des exemples d'apprentissage, et nous avons défini un critère normalisé μ . Nous sommes arrivés à la conclusion que le modèle définitif devait être choisi en fonction de l'objectif recherché : s'il s'agit de trouver le meilleur modèle, sachant que l'on ne dispose que de ces N exemples d'apprentissage, il convient de choisir l'architecture présentant le critère μ le plus grand possible. En revanche, si l'on a l'intention et surtout la possibilité d'améliorer les performances du modèle, le choix d'un modèle légèrement surajusté, associé à l'utilisation des intervalles de confiance, permettra de compléter de manière ciblée la base d'apprentissage et d'augmenter localement la confiance du modèle.

Enfin, nous avons proposé une adaptation de l'algorithme de Levenberg-Marquardt, de manière à pouvoir minimiser directement la fonction de coût correspondant à l'erreur E_p . Nous avons montré qu'il est avantageux de rechercher un minimum de cette nouvelle fonction de coût, après la minimisation classique de la fonction de coût des moindres carrés.

Finalement, l'étude théorique du leave-one-out présentée dans ce mémoire nous a conduit à définir une méthode originale de sélection de modèles, qui considère directement l'origine du surajustement, à savoir la trop grande influence de certains exemples sur l'estimation des paramètres d'un modèle. L'utilisation conjointe de cette méthode et des intervalles de confiance permet à l'ingénieur de maîtriser à la fois la complexité et le domaine de validité de ses modèles, et ainsi d'en envisager l'amélioration progressive.

Une étude détaillée du procédé de soudage par points nous a permis de développer une application de cette approche au problème de la prédiction du diamètre de bouton. Nous avons tout d'abord analysé, d'un point de vue physique, les grandeurs susceptibles d'être pertinentes pour l'objectif de modélisation que nous nous étions fixé. Afin de sélectionner les quantités les plus pertinentes parmi les grandeurs candidates résultant de notre analyse, nous avons utilisé plusieurs méthodes de sélection qui ont donné des résultats concordants. Nous avons ensuite appliqué notre méthodologie de conception de modèles à deux aciers revêtus, et avons obtenu des résultats très proches de l'optimum : des prédictions du diamètre du bouton dont la précision est de l'ordre de l'incertitude sur la mesure de ce diamètre. Il convient de noter que, compte tenu de la nature de ce procédé, et en particulier de la différence entre paramètres de soudage et caractéristiques de la qualité du point, l'utilisation des intervalles de confiance a été déterminante dans le succès de l'application.

Les perspectives industrielles offertes par un modèle de contrôle non destructif de la qualité de soudures par points sont nombreuses. Il reste désormais à vérifier que cette approche pourra s'accommoder des contraintes en milieu de production.

Notre travail nous a donc permis de proposer une méthodologie originale, et d'usage très général, pour la conception de modèles non linéaires. Nous en avons montré l'efficacité lors de son application, dans le cas des réseaux de neurones, à un problème industriel réel dont l'enjeu est important.

REFERENCES BIBLIOGRAPHIQUES

[Anders 99]

U. ANDERS & O. KORN

"Model Selection in Neural Networks"

Neural Networks 12, pp.309-323, 1999

[Antoniadis 92]

A. ANTONIADIS, J. BERRUYER & R. CARMONA

"Régression non linéaire et applications"

Paris : Economica, 1992

[Aro 94]

"Micro 2x16 II Type "Bas-Plus" Logiciel 7B-8B"

Notice d'utilisation, ARO S.A., 1994

[Barhen 93]

J. BAHREN, B. CETIN & J. BURDICK

"Overcoming Local Minima in Neural Learning"

Proceedings de la 6^{ème} Conférence Internationale Neuro-Nîmes, 1993

[Bartlett 97]

P.L. BARLETT

"For Valid Generalization, the Size of the Weights is more Important than the Size of the Network"

M.C. Mozer, M.I. Jordan & T. Petsche (eds), Advances in Neural Information Processing Systems 9, Cambridge, MA : the MIT Press, pp. 134-140, 1997

[Bates 88]

D.M. BATES & D.G. WATTS

"Nonlinear Regression Analysis and its Applications"

Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, 1988

[Bishop 93]

C.M. BISHOP

"Curvature-Driven Smoothing : a Learning Algorithm for Feedforward Networks"

IEEE Transactions on Neural Networks 4 (5), pp. 882-884, 1993

[Bishop 97]

C.M. BISHOP

"Neural Networks for Pattern Recognition"

Third Edition, Clarendon Press, Oxford, 1997

[Bottou 97]

"La mise en œuvre des idées de V.N. Vapnik"

Chapitre 16 de STATISTIQUE ET METHODES NEURONALES, S. Thiria, Y. Lechevallier, O. Gascuel & S. Canu éditeurs, DUNOD, 1997

- [Breiman 96]
L. BREIMAN
"Heuristics of Instability and Stabilization in Model Selection"
Annals of Statistics 24, pp. 2350-2383, 1996
- [Brown 97]
J.D. BROWN, M.G. RODD & N.T. WILLIAMS
"Application of Artificial Intelligence to Resistance Spot Welding"
International Institute of Welding, Doc. No. III - 1092, 1997
- [Brugge 95]
M.H. BRUGGE, W.J. JANSEN, J.A.G. NIJHUIS & L. SPAANENBURG
"Non-destructive Test of Spot Welds Using a Neural Network"
Proceedings ICAN '95, Paris, 1995
- [Cazes 93]
R. CAZES
"Le soudage par résistance"
Les techniques de l'ingénieur, novembre 1993
- [Chen 89]
S. CHEN, S.A. BILLINGS & W. LUO
"Orthogonal Least Squares Methods & their Application to Nonlinear System Identification"
International Journal of Control, Vol. 50, n° 5, pp. 1873-1896, 1989
- [Cybenko 89]
G. CYBENKO
"Approximation by Superpositions of a Sigmoidal Function"
Mathematics of Control, Signals and Systems, Vol. 2, pp. 303-314, 1989
- [De Vaux 98]
R.D. DE VAUX, J. SCHUMI, J. SCHWEINSBERG & L.H. UNGAR
"Predictions Intervals for Neural Networks via Nonlinear Regression"
Technometrics 40, pp. 273-282, 1998
- [Devroye 79]
L.P. DEVROYE & T.J. WAGNER
"Distribution-free Inequalities for the Deleted and Holdout Error Estimates"
IEEE Transactions on Information Theory, IT-25 (2), pp. 601-604, 1979
- [Dilthey 97]
U. DILTHEY & J. DICKERSBACH
"Quality Assurance of Resistance Spot Welding by Employment of Neural Networks"
International Institute of Welding, Doc. No. III - 1093, 1997
- [Dreyfus 97]
G. DREYFUS, L. PERSONNAZ & G. TOULOUSE
"Perceptrons, Old and New"
Enciclopedia Italiana, in press.

[Dupuy 98]

T. DUPUY

"La dégradation des électrodes lors du soudage par points de tôles d'acier zinguées"

Thèse de Doctorat de Ecole des Mines de Paris, 1998

[Efron 93]

B. EFRON & R.J. TIBSHIRANI

"An Introduction to the Bootstrap"

New-York : Chapman & Hall, 1993

[Euvrard 93]

G. EUVRARD

"Estimation d'une régression non linéaire par réseaux de neurones ; application à un problème de robotique mobile"

Thèse de Doctorat de l'Université Paris VI, 1993

[Funahashi 89]

K. FUNAHASHI

"On the Approximate Realization of Continuous Mappings by Neural Networks"

Neural Networks 4, pp. 349-360, 1991

[Gallinari 97]

P. GALLINARI

"Heuristiques pour la généralisation"

Chapitre 14 de STATISTIQUE ET METHODES NEURONALES, S. Thiria, Y. Lechevallier, O. Gascuel & S. Canu éditeurs, DUNOD, 1997

[Geman 92]

S. GEMAN, E. BIENENSTOCK & R. DOURSAT

"Neural Networks and the Bias / Variance Dilemma"

Neural Computation 4, pp. 1-58, 1992

[Gobez 94]

P. GOBEZ

"Soudage des tôles revêtues"

Les Techniques de l'ingénieur, B 7771, 1994

[Hansen 96]

L.K. HANSEN & J. LARSEN

"Linear Unlearning for Cross-Validation"

Advances in Computational Mathematics 5, pp. 269-280, 1996

[Hornik 94]

K. HORNİK, M. STINCHCOMBE, H. WHITE & P. AUER

"Degree of Approximation Results for Feedforward Networks Approximating Unknown Mappings and their Derivatives"

Neural Computation, Vol. 6, c°6, pp. 1262-1275, 1994

[Ivezic 99]

N. IVEZIC, J.D. ALLEN & T. ZACHARIA

"Neural Network-Based Resistance Spot Welding Control and Quality Prediction"

Proceedings of the Second International Conference on Intelligent Manufacturing and Processing of Materials (IPMM '99), Honolulu, 1999

[Jou 94]

M. JOU

"An Intelligent Control System for Resistance Spot Welding Using Fuzzy Logic and Neural Networks"

Thèse, Rensselaer Polytechnic Institute, Troy, 1994

[Jou 95]

M. JOU, R.W. MESSLER & C.J. LI

"An Intelligent Control System for Resistance Spot Welding Using a Neural Network and Fuzzy Logic"

IEEE Transaction, 1995

[Kearns 97]

M. KEARNS, D. RON

"Algorithmic Stability and Sanity-Check Bounds for Leave-One-Out Cross Validation"

Submitted to the Tenth Annual Conference on Computational Learning Theory

[Levenberg 44]

K. LEVENBERG

"A Method for the Solution of Certain Non-linear Problems in Least Squares"

Quarterly Journal of Applied Mathematics II (2), pp. 164-168, 1944

[MacKay 92]

D.J.C. MACKAY

"A Practical Bayesian Framework for Backpropagation Networks"

Neural Computation 4 (4), pp. 448-472, 1992

[Marquardt 63]

D.W. MARQUARDT

"An Algorithm for Least-squares Estimation of Non-linear Parameters"

Journal of the Society of Industrial and Applied Mathematics 11 (2), pp. 431-441, 1963

[Matsuyama 97]

K.I. MATSUYAMA

"Nugget Size Sensing of Spot Weld Based on Neural Network Learning"

International Institut of Welding, Doc. No. III - 1081, 1997

[McQuarrie 98]

A.D.R. MCQUARRIE & C.L. TSAI

"Regression and Time Series Model Selection"

Singapore : World Scientific, 1998

[Messler 94]

R.W. MESSLER, M. JOU & C.J. LI

"A Fuzzy Logic Control System for Resistance Spot Welding Based on a Neural Network Model"

Proceedings of Sheet Metal Welding Conference N° 6, AWS, 1994

[Moody 94]

J. MOODY

"Prediction Risk and Neural Network Architecture Selection"

From Statistics to Neural Networks : Theory and Pattern Recognition Applications, V. Cherkassky, J.H. Friedman, and H. Wechsler (eds), Springer-Verlag, 1994

[Nash 90]

J.C. NASH

"Compact Numerical Methods for Computers : Linear Algebra and Function Minimisation"

Ed. Adam Hilger, 1990

[Nelson 91]

M.C. NELSON & W.T. ILLINGWORTH

"A Practical Guide to Neural Nets"

Reading, MA : Addison Wesley, 1991

[NF A 87-001]

"Caractérisation de la soudabilité par résistance par points de produits plats revêtus ou non"

Comité de Normalisation de la Soudure - AFNOR, décembre 1994

[Osman 95]

K.A. OSMAN, A.M. HIGGINSON, H.R. KELLY, C.J. NEWTON & D.R. BOOMER

"Monitoring of Resistance Spot-Welding Using Multi-Layer Perceptrons"

Proceedings Autotech '95, Birmingham, 1995

[Osman 96]

K.A. OSMAN, A.M. HIGGINSON, H.R. KELLY, C.J. NEWTON & P. SHEASBY

"Prediction of Aluminium Spot Weld Quality Using Artificial Neural Networks"

Proceedings AWS '96, 1996

[Oukhellou 98]

L. OUKHELLOU, P. AKNIN, H. STOPPIGLIA & G. DREYFUS

"A new Decision Criterion for Feature Selection : Application to the Classification of Non Destructive Testing Signature"

European Signal Processing Conference (EUSIPCO'98), Rhodes, 1998

[Oussar 98]

Y. OUSSAR

"Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus"

Thèse de Doctorat de l'Université Paris VI, 1998

[Plutowski 94]

M. PLUTOWSKI, S. SAKATA & H. WHITE

"Cross-Validation Estimates IMSE"

Advances in Neural Information Processing Systems 6, San Mateo, CA: Morgan Kaufmann Publishers, 1994

[Powell 76]

M.J.D. POWELL

"Somme Global Convergence Properties of a Variable Metric Algorithm for Minimisation without Line Searches"

Nonlinear Programming, London, 1986 SIAM-AMS Proceedings 9, R.W. Cottle & C.E. Lemke, Eds. Providence RI, 1976

[Press 92]

W.H. PRESS, S.A. TEUKOLSKY, W.T. VETTERLING & B.P. FLANNERY

"Numerical Recipes in C : The Art of Scientific Computing"

Second Edition, Cambridge University Press, 1992

[Rivals 98]

I. RIVALS, L. PERSONNAZ

"Construction of Confidence Intervals in Neural Modeling Using a Linear Taylor Expansion"

International Workshop on Advanced Black-Box Techniques for Nonlinear Modeling : Theory and Applications, Leuven-Belgium, 1998

"Construction of Confidence Intervals for Neural Networks Based on Least Squares Estimation"

Neural Networks, à paraître

[Rumelhart 86]

D.E. RUMELHART, G.E. HINTON & R.J. WILLIAMS

"Learning Internal Representations by Error Propagation"

Parallel Distributed Processing, MIT Press, Cambridge MA, pp. 318-362, 1986

[Saarinen 93]

S. SAARINEN, R. BRAMLEY & G. CYBENKO

"Ill-Conditioning in Neural Network Training Problems"

SIAM J. Sci. Stat. Comp. 14, pp. 693-714, 1993

[Saporta 90]

G. SAPORTA

"Probabilités, analyse des données et statistique"

Editions Technip, Paris, 1990

[Satoh 97]

T. SATOH, H. ABE & S. SUZUYAMA

"A Trial of Quality Assurance in Resistance Spot Welding by Aid of Neural Network and Fuzzy Reasoning - Accomplished with Detection of Top Diameter of Electrode"

International Institut of Welding, Doc. No. III - 1083, 1997

[Sauvage 94]

F. SAUVAGE, G. KAPLAN

"Le soudage"

Chapitre 38 du Livre de l'Acier, Technique & Documentation - Lavoisier, 1994

[Seber 89]

G.A.F. SEBER & C.J. WILD

"Nonlinear regression"

Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons (eds), New York, 1989.

[Sjöberg 92]

J. SJÖBERG & L. LJUNG

"Overtraining, Regularization and Searching for Minimum in Neural Networks"

Technical report LiTH-ISY-I-1297, Department of Electrical Engineering, Linköping University, S-591 93 Linköping, <http://www.control.isy.liu.se>

[Sorensen 96]

P.H. SORENSEN, M. NORGDARD, L.K. HANSEN & J. LARSEN

"Cross-Validation with LULOO"

Proceedings of the International Conference on Neural Information Processing - ICONIP '96, 1996

[Stone 74]

M. STONE

"Cross-Validatory Choice and Assessment of Statistical Predictions"

Journal of the Royal Statistical Society B 36, pp. 111-147, 1974

[Stoppiglia 97]

H. STOPPIGLIA

"Méthodes statistiques de sélection de modèles neuronaux ; applications financières et bancaires"

Thèse de Doctorat de l'Université Paris VI, 1997

[Thiria 97]

S. THIRIA, Y. LECHEVALLIER, O. GASCUEL & S. CANU

"Statistique et méthodes neuronales"

2^{ème} cycle école d'ingénieur, S. Thiria, Y. Lechevallier, O. Gascuel & S. Canu éditeurs, DUNOD, 1997

[Tibshirani 96]

R.J. TIBSHIRANI

"A Comparaison of Some Error Estimates for Neural Models"

Neural Computation 8, pp. 152-163, 1996

[Thièblemont 92]

E. THIEBLEMONT

"Modélisation du soudage par résistance par points"

Thèse de doctorat, Institut National Polytechnique de Lorraine, 1992

[Vapnik 82]

V.N. VAPNIK

"Estimation of Dependences Based on Empirical Data"

Springer Verlag, New-York, 1982

[Waddel 97]

W. WADDEL & N. WILLIAMS

"Control of Resistance Spot Welded Quality Using Multi-parameter Derived Algorithms for Zinc-Coated Sheets"

Rapport technique EUR 17859 EN du contrat N° 7210-MB/805, Steel Research, 1997

[Waschkies 97]

E. WASCHKIES

"Prüfen des Widerstandspunktschweißprozesses mit Ultraschall"

Schweissen und Schneiden 1/97, pp. 17-19, 1997

[Wesgate 86]

S.A. WESTGATE

"An Evaluation of Ultrasonic and Electrical Resistance Methods for the Non-destructive Testing (NDT) of Resistance Spot Welds in Low Carbon Steel Sheet"

The Welding Institute, 1986

[Wolfe 69]

P. WOLFE

"Convergence Conditions for Ascent Methods"

S.I.A.M. Review 11, pp. 226-235, 1969

[Zhou 98]

G. ZHOU & J. SI

"A Systematic and Effective Supervised Learning Mechanism Based on Jacobian Rank Deficiency"

Neural Computation 10, pp.1031-1045, 1998

ANNEXE 1 : CALCUL DES h_{ii}

Soit Z une matrice de dimensions (N, q) (avec $N \geq q$), matrice jacobienne d'une fonction $f(X, \theta)$ réalisée par exemple par un réseau de neurones à q coefficients, ou par une fonction linéaire à $q - 1$ régresseurs (dans ce cas, la matrice Z est égale à la matrice des observations).

Définissons $Z = [z^1, \dots, z^N]$, avec $z^i = \left. \frac{\partial f(x^i, \theta)}{\partial \theta} \right|_{\theta = \theta^*}$. On cherche à calculer les termes diagonaux de la matrice de projection $H = Z ({}^t Z Z)^{-1} {}^t Z$, définis comme :

$$h_{ii} = {}^t z^i ({}^t Z Z)^{-1} z^i \quad (\text{A1.1})$$

En tant qu'éléments diagonaux d'une matrice de projection orthogonale, les termes $\{h_{ii}\}_{i=1, \dots, N}$ ne sont définis que dans le cas où Z est de rang plein, c'est-à-dire si ${}^t Z Z$ est inversible. Dans ce cas, ils vérifient les propriétés suivantes :

$$\forall i \in [1, \dots, N] \quad 0 \leq h_{ii} \leq 1 \quad (\text{A1.2})$$

$$\text{trace}(H) = \sum_{i=1}^N h_{ii} = \text{rang}(Z) \quad (\text{A1.3})$$

On obtient Z soit par calcul direct, à partir de la forme analytique de la fonction f , soit - dans le cas de réseaux de neurones non bouclés - en rétropropageant une erreur égale à -1.

Une première méthode de calcul des h_{ii} consiste à calculer la matrice ${}^t Z Z$, à l'inverser par une méthode classique (Cholesky, décomposition LU, ...), puis à la multiplier à droite et à gauche par les vecteurs z^i et ${}^t z^i$. Cette méthode ne donne cependant de bons résultats que si la matrice ${}^t Z Z$ est suffisamment bien conditionnée pour que son inversion se déroule sans problème. Dans le cas contraire, ce calcul donne des valeurs supérieures à 1, voire négatives, c'est-à-dire ne satisfaisant pas à la relation (A1.2).

La solution que nous proposons consiste à décomposer la matrice Z sous la forme :

$$Z = U W {}^t V \quad (\text{A1.4})$$

avec :

- U matrice (N, q) telle que ${}^t U U = I$,
- W matrice (q, q) diagonale, dont les termes diagonaux, appelés valeurs singulières de Z , sont positifs ou nuls et classés par ordre décroissant,
- V matrice (q, q) telle que ${}^t V V = V {}^t V = I$.

Cette décomposition, connue sous le nom de décomposition en valeurs singulières ou décomposition SVD (Singular Value Decomposition), est précise et très robuste, même si la matrice Z est mal conditionnée ou de rang inférieur à q (cf. [Press 92]).

On obtient donc :

$${}^tZZ = V W {}^tU U W {}^tV = V W^2 {}^tV \quad (\text{A1.5})$$

puis

$$({}^tZZ)^{-1} = V W^{-2} {}^tV \quad (\text{A1.6})$$

Cette décomposition permet donc le calcul direct de la matrice $({}^tZZ)^{-1}$, dont les éléments s'écrivent :

$$({}^tZZ)^{-1}_{lj} = \sum_{k=1}^q \frac{V_{lk} V_{jk}}{W_{kk}^2} \quad (\text{A1.7})$$

On peut alors calculer l'expression de h_{ii} sous la forme :

$$h_{ii} = {}^t\mathbf{z}^i ({}^tZZ)^{-1} \mathbf{z}^i = \sum_{l=1}^q \sum_{j=1}^q Z_{il} Z_{ij} ({}^tZZ)^{-1}_{lj} \quad (\text{A1.8})$$

En utilisant (A1.7) puis en inversant l'ordre des sommes, il vient :

$$h_{ii} = \sum_{k=1}^q \left(\frac{1}{W_{kk}} \sum_{j=1}^q Z_{ij} V_{jk} \right)^2 \quad (\text{A1.9})$$

Cette méthode permet ainsi le calcul des $\{h_{ii}\}_{i=1, \dots, N}$ sans avoir à calculer explicitement les termes de la matrice $({}^tZZ)^{-1}$, ce qui est important au niveau de la précision du calcul, dans le cas de matrices mal conditionnées. D'un point de vue numérique, étant donné que les valeurs singulières de Z sont classées par ordre décroissant, il est conseillé de calculer les $\{h_{ii}\}_{i=1, \dots, N}$ en faisant varier k de q à 1 et non pas de 1 à q .

Cette méthode de calcul fournit des termes systématiquement positifs ou nuls, ce qui permet de s'assurer en partie de la condition (A1.2).

ANNEXE 2 : SURAJUSTEMENT ET REDONDANCE DE COEFFICIENTS

La méthode présentée au paragraphe 2.4.2 pour détecter le surajustement part du constat suivant : à partir d'une certaine taille d'architecture, la matrice Z - dont le rang est normalement égal au nombre de coefficients ajustables du modèle - n'est plus de rang plein, traduisant ainsi un surajustement. Cette interprétation de la déficience du rang de Z suppose toutefois que l'architecture utilisée ne présente aucune redondance parmi ses coefficients. Nous précisons ce point dans la présente annexe.

Considérons par exemple le réseau représenté sur la figure A2.1, réalisant une fonction $y = f(x, \mathbf{w})$, dans laquelle \mathbf{w} est un vecteur de 6 paramètres. À partir de la forme algébrique de f , on montre la relation suivante :

$$\nabla(x, \mathbf{w}) \quad w_{43} \frac{\partial f(x, \mathbf{w})}{\partial w_{43}} = w_{30} \frac{\partial f(x, \mathbf{w})}{\partial w_{30}} + w_{31} \frac{\partial f(x, \mathbf{w})}{\partial w_{31}}$$

Cette relation traduit une dépendance entre les colonnes correspondantes de la matrice Z , qui - quel que soit le vecteur des paramètres - sera de rang inférieur ou égal à 5.

La relation précédente peut également se comprendre de la façon suivante : en multipliant le paramètre w_{43} par un réel quelconque k et en divisant w_{30} et w_{31} par ce même facteur, la fonction f reste inchangée. On pourrait ainsi fixer arbitrairement un de ces trois poids à une valeur non nulle (par exemple à 1) sans modifier la famille de fonctions paramétrées engendrée par ce réseau. Parmi les 6 poids utilisés, il n'y en a donc que 5 réellement utiles : l'un d'eux est redondant.

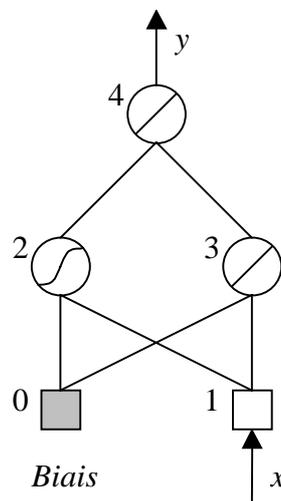


Figure A2.1 : Exemple de modèle statique avec redondance de coefficients

Dans le cas d'un modèle statique, ce cas ne se produit que lorsqu'on utilise plusieurs neurones linéaires en cascade, ce qui ne présente effectivement aucun intérêt. Il est donc difficile de ne pas se rendre compte de la présence de redondances dans l'architecture d'un réseau de neurones classique utilisé comme modèle statique.

En revanche, lorsque ces modèles sont bouclés de façon à modéliser des phénomènes dynamiques, ceci est moins évident à constater. Le modèle représenté sur la figure A2.2 permet de s'en convaincre. Il s'agit d'un modèle, comportant une entrée, une sortie et une variable d'état, réalisant une fonction du type :

$$\begin{cases} y(p+1) = f(x(p), y(p), s(p), \mathbf{w}) \\ s(p+1) = g(x(p), y(p), s(p), \mathbf{w}) \end{cases}$$

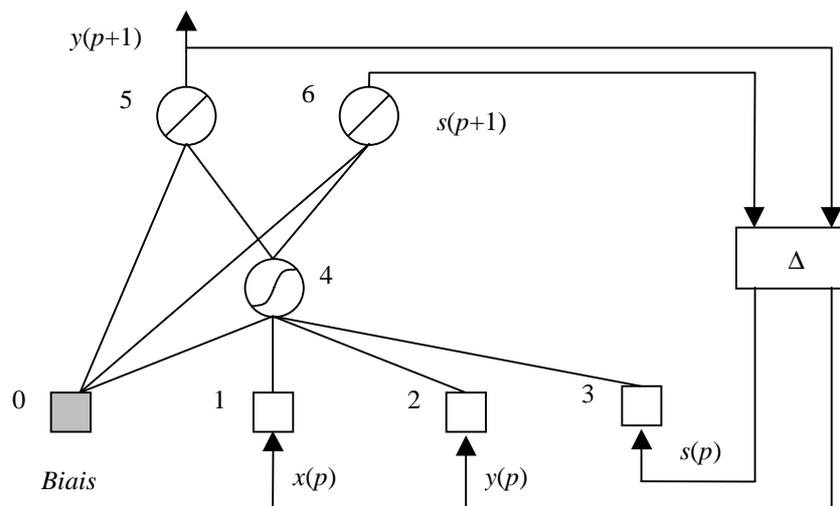


Figure A2.2 : Exemple de modèle dynamique avec redondance de coefficients

Lorsque l'on considère cette fonction sur un horizon fini, et que l'on en fixe l'état initial (c'est-à-dire $y(0)$ et $s(0)$), on peut écrire $y(p+1) = F(x(0), \dots, x(p), \mathbf{w})$.

En l'absence de valeurs désirées pour la sortie d'état, et en appliquant le même type de raisonnement que dans le cas statique, on peut montrer que la matrice jacobienne de ce modèle, qui possède 8 coefficients, est toujours de rang inférieur ou égal à 5.

Pour ce faire, considérons les trois transformations linéaires suivantes, qui modifient la fonction g tout en laissant la fonction f invariante :

$$\begin{pmatrix} w_{60} \rightarrow w_{60} + k \\ w_{40} \rightarrow w_{40} - kw_{43} \end{pmatrix} \Leftrightarrow s \rightarrow s + k : \text{on peut donc fixer arbitrairement } w_{60} \text{ à } 0,$$

$$\begin{pmatrix} w_{60} \rightarrow k'w_{60} \\ w_{64} \rightarrow k'w_{64} \\ w_{43} \rightarrow w_{43} / k' \end{pmatrix} \Leftrightarrow s \rightarrow k's : \text{on peut donc fixer arbitrairement } w_{64} \text{ à } 1,$$

$$\begin{pmatrix} w_{60} \rightarrow w_{60} + k''w_{50} \\ w_{64} \rightarrow w_{64} + k''w_{54} \\ w_{42} \rightarrow w_{42} - k''w_{43} \end{pmatrix} \Leftrightarrow s \rightarrow s + k''y : \text{on peut donc fixer arbitrairement } w_{42} \text{ à } 0.$$

Pour chacune de ces transformations, il existe une relation linéaire entre les colonnes de la matrice jacobienne de F . Par exemple, pour la première de ces transformations, il existe une

relation linéaire, valable quels que soient $x(0), \dots, x(p)$, entre $\frac{\partial y(p+1)}{\partial w_{40}}$, $\frac{\partial y(p+1)}{\partial w_{43}}$ et $\frac{\partial y(p+1)}{\partial w}$. L'expression de cette relation, dont les coefficients dépendent de $y(0)$, $s(0)$ et du vecteur w , n'est pas aussi simple que dans l'exemple du modèle statique présenté plus haut.

Il y a donc 3 coefficients redondants dans l'architecture de la figure A2.2. Le modèle dynamique réalisant la même fonction f , sans aucune redondance, est représenté sur la figure A2.3 :

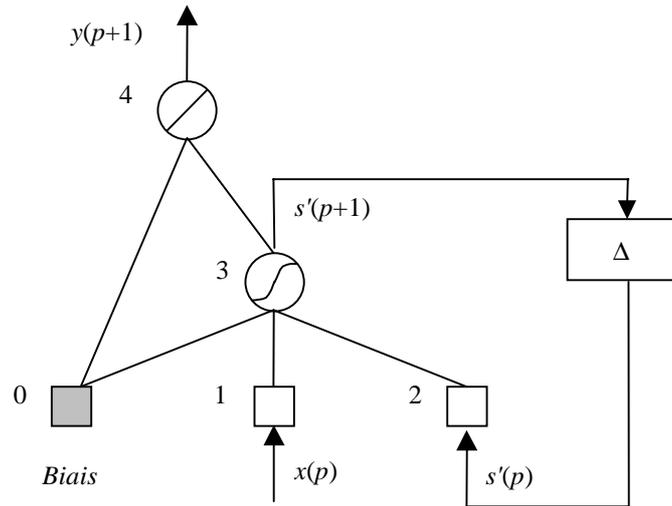


Figure A2.3 : Modèle dynamique équivalent, sans coefficients redondants

Cet exemple montre que le bouclage d'une sortie, ou d'une variable d'état, provenant d'un neurone linéaire peut conduire à des redondances de coefficients et ainsi à une déficience systématique dans le rang de Z . Il n'a cependant qu'une valeur illustrative : notre propos est simplement de mettre en garde contre certaines redondances "cachées" dans les modèles dynamiques.

La mise au point de règles concernant la définition d'architectures de modèles bouclés et de formes canoniques sans redondances demanderait sans doute une approche un peu plus poussée, qui sort du cadre de la présente thèse.

ANNEXE 3 : DEMONSTRATION DE LA RELATION (3.8)

Nous présentons ici une démonstration de la relation (3.8), qui se fonde sur deux lemmes d'inversion matricielle. Elle est inspirée de [Antoniadis 92].

Considérons la matrice de dimensions $(q+1, q+1)$ $M = \begin{pmatrix} {}^tZ Z & z^i \\ {}^t z^i & 1 \end{pmatrix}$ supposée inversible et

notons son inverse $M^{-1} = \begin{pmatrix} A & \mathbf{b} \\ \mathbf{b} & c \end{pmatrix}$. La matrice A , le vecteur \mathbf{b} et le réel c vérifient donc les

équations suivantes :

$${}^tZ Z A + z^i \mathbf{b} = I \quad (\text{A3.1})$$

$${}^tZ Z \mathbf{b} + z^i c = 0 \quad (\text{A3.2})$$

$$\mathbf{b} + {}^t z^i A = 0 \quad (\text{A3.3})$$

$$c + {}^t z^i \mathbf{b} = 1 \quad (\text{A3.4})$$

Lemme 1 :

$$A = \left({}^tZ^{(-i)} Z^{(-i)} \right)^{-1} \quad (\text{A3.5})$$

Démonstration :

À partir de (A3.1) et (A3.3), on a ${}^tZ Z A - z^i {}^t z^i A = I$, soit $A = \left({}^tZ Z - z^i {}^t z^i \right)^{-1} = \left({}^tZ^{(-i)} Z^{(-i)} \right)^{-1}$

Lemme 2 :

$$A = \left({}^tZ Z \right)^{-1} + \frac{\left({}^tZ Z \right)^{-1} z^i {}^t z^i \left({}^tZ Z \right)^{-1}}{1 - h_{ii}} \quad (\text{A3.6})$$

Démonstration :

En résolvant le système linéaire constitué des équations (A3.1) à (A3.4), on obtient

$c = \frac{1}{1 - h_{ii}}$, $\mathbf{b} = -\frac{\left({}^tZ Z \right)^{-1} z^i}{1 - h_{ii}}$ et la relation (A3.6). Ceci est un cas particulier d'un lemme

d'inversion matricielle plus général.

Démonstration de la relation (3.8) :

En combinant (A3.5) et (A3.6), on obtient :

$$\left({}^tZ^{(-i)} Z^{(-i)}\right)^{-1} = \left({}^tZ Z\right)^{-1} + \frac{\left({}^tZ Z\right)^{-1} \mathbf{z}^i {}^t\mathbf{z}^i \left({}^tZ Z\right)^{-1}}{1 - h_{ii}} \quad (\text{A3.7})$$

Puis, à partir de la relation (3.3), restreinte au premier ordre, et de (3.6), il vient :

$$\boldsymbol{\theta}_{LS} - \boldsymbol{\theta}_{LS}^{(-i)} = \left({}^tZ^{(-i)} Z^{(-i)}\right)^{-1} {}^tZ^{(-i)} \left(\mathbf{y}_p^{(-i)} - \mathbf{f}^{(-i)}(X, \boldsymbol{\theta}^*)\right) - \left({}^tZ Z\right)^{-1} {}^tZ \left(\mathbf{y}_p - \mathbf{f}(X, \boldsymbol{\theta}^*)\right) \quad (\text{A3.8})$$

De plus, on peut facilement montrer que :

$${}^tZ^{(-i)} \left(\mathbf{y}_p^{(-i)} - \mathbf{f}^{(-i)}(X, \boldsymbol{\theta}^*)\right) = {}^tZ \left(\mathbf{y}_p - \mathbf{f}(X, \boldsymbol{\theta}^*)\right) - \mathbf{z}^i \left(y_{pi} - f(\mathbf{x}^i, \boldsymbol{\theta}^*)\right), \quad (\text{A3.9})$$

où y_{pi} est la $i^{\text{ème}}$ composante du vecteur \mathbf{y}_p .

En combinant les équations (A3.7), (A3.8) et (A3.9), on démontre alors que :

$$\boldsymbol{\theta}_{LS}^{(-i)} - \boldsymbol{\theta}_{LS} = \frac{\left({}^tZ Z\right)^{-1} \mathbf{z}^i}{1 - h_{ii}} \left\{ \left(y_{pi} - f(\mathbf{x}^i, \boldsymbol{\theta}^*)\right) - {}^t\mathbf{z}^i \left({}^tZ Z\right)^{-1} {}^tZ \left(\mathbf{y}_p - \mathbf{f}(X, \boldsymbol{\theta}^*)\right) \right\} \quad (\text{A3.10})$$

Le second terme à l'intérieur des accolades est la $i^{\text{ème}}$ composante de la projection de $\left(\mathbf{y}_p - \mathbf{f}(X, \boldsymbol{\theta}^*)\right)$ sur le sous-espace des solutions ; par conséquent, la quantité à l'intérieur des accolades est la $i^{\text{ème}}$ composante du vecteur des résidus, c'est-à-dire le résidu de l'exemple i . Ceci prouve la relation (3.7).

ANNEXE 4 : PROCÉDE D'ORTHONORMALISATION DE GRAM-SCHMIDT

Nous présentons ici l'algorithme de Gram-Schmidt modifié, décrit en détail par exemple dans [Chen 89].

Le problème que nous considérons est le classement de q entrées par ordre d'importance dans l'optique d'une régression linéaire par rapport aux paramètres.

Notons N le nombre d'observations, $X = \begin{pmatrix} x_1^1 & \dots & x_q^1 \\ \vdots & \dots & \vdots \\ x_1^N & \dots & x_q^N \end{pmatrix} = (\mathbf{x}_1 \dots \mathbf{x}_q)$ la matrice des observations, avec $\mathbf{x}_i = \begin{pmatrix} x_i^1 \\ \vdots \\ x_i^N \end{pmatrix}$ le vecteur de la $i^{\text{ème}}$ entrée et $\mathbf{y} = \begin{pmatrix} y^1 \\ \vdots \\ y^N \end{pmatrix}$ le vecteur de la sortie du processus.

On suppose par ailleurs que les vecteurs \mathbf{y} et $\{\mathbf{x}_i\}_{i=1, \dots, N}$ sont centrés, c'est-à-dire qu'on leur a appliqué une transformation affine de telle sorte que la moyenne de leurs composantes soit nulle.

On se place dans l'espace à N dimensions engendré par les vecteurs \mathbf{y} et $\{\mathbf{x}_i\}_{i=1, \dots, N}$.

La première itération consiste à chercher le vecteur d'entrée qui explique le mieux, au sens des moindres carrés, le vecteur de sortie. Pour cela, on se sert du carré du cosinus des angles entre le vecteur de sortie et les différents vecteurs d'entrée :

$$\cos^2(\mathbf{x}_i, \mathbf{y}) = \frac{({}^t\mathbf{x}_i \mathbf{y})^2}{({}^t\mathbf{x}_i \mathbf{x}_i) \cdot ({}^t\mathbf{y} \mathbf{y})} \quad (\text{A4.1})$$

On sélectionne le vecteur d'entrée pour lequel cette quantité est maximale. Ensuite, on élimine la contribution de l'entrée sélectionnée en projetant le vecteur de sortie, et tous les vecteurs d'entrée restants, sur le sous-espace orthogonal au vecteur sélectionné.

La procédure se poursuit en choisissant, une nouvelle fois, le vecteur d'entrée projeté qui explique le mieux la sortie projetée. Elle se termine lorsque tous les vecteurs d'entrée ont été ordonnés.

Dans le cas d'un espace à 3 dimensions (c'est-à-dire avec 3 exemples) et 2 entrées, le principe de la première itération est représenté sur la figure A4.1.

Sur cet exemple, l'angle α_1 entre \mathbf{y} et \mathbf{x}_1 étant plus petit que l'angle α_2 entre \mathbf{y} et \mathbf{x}_2 , l'entrée \mathbf{x}_1 est celle qui explique le mieux la sortie \mathbf{y} . On la sélectionne donc en premier et l'on projette les vecteurs \mathbf{y} et \mathbf{x}_2 sur le plan perpendiculaire à \mathbf{x}_1 , nommé P_1 . On recommence alors la procédure à partir des projections $\mathbf{p}_1(\mathbf{y})$ et $\mathbf{p}_1(\mathbf{x}_2)$ de ces vecteurs et \mathbf{x}_2 est alors naturellement la deuxième et dernière entrée sélectionnée.

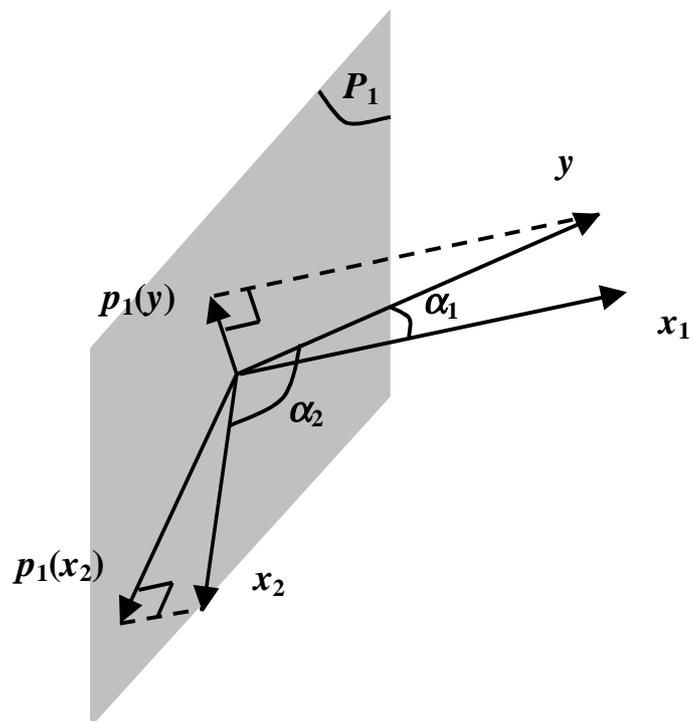


Figure A4.1 : Principe de la méthode de Gram-Schmidt dans un espace à 3 dimensions

ANNEXE 5 : CARACTERISTIQUES DES PRODUITS UTILISES

Les caractéristiques et repérages des tôles d'acier revêtues utilisées dans ce travail sont résumés dans les tableaux suivants :

Repère	Étude	Usine de provenance	N° de bobine	Nature de la tôle	Épaisseur nominale de la tôle	Nature du revêtement	Épaisseur nominale du revêtement
GZ 1	97058	Montataire	366105/2	IF Ti	0,8 mm	Galvanisé	2 x 10 µm
GZ 2	98222	Mardyck	Plusieurs (lot 98177)	IF Ti	0,7 mm	Galvanisé	2 x 10 µm
GA	99088	Mardyck	Plusieurs (lot 98977)	IF Ti	1,2 mm	Galvanisé-allié	2 x 10 µm

Tableau A5.1 : Repérages et caractéristiques des produits

Pour les produits GZ 2 et GA, nous donnons ci-dessous la composition chimique et les caractéristiques mécaniques relevées sur une seule bobine, les autres bobines étant considérées comme ayant des propriétés comparables.

Repère	C	Mn	P	S	Si	Al	Ni	Cr	Cu	Mo	Sn	Nb	V	Ti	B	Ca	N2
GZ 1	1,9	143	11	8,4	16	39	17	17	7	<1	3	2	2	55	<0,3	<0,3	3,7
GZ 2	2,4	94	6	6,7	5	30	20	15	8	<1	1	<1	2	55	<0,3	<0,3	2,8
GA	1,7	95	12	8,5	6	39	18	18	10	<1	1	<1	2	55	0,4	<0,3	2,2

Tableau A5.2 : Composition chimique des échantillons (teneur massique 10⁻³ %)

Repère	Limite élastique à 0,2 % (Mpa)		Contrainte à rupture (Mpa)		Allongement à rupture (%)	
	Sens long	Sens travers	Sens long	Sens travers	Sens long	Sens travers
GZ 1	170	152	304	299	42,8	44,2
GZ 2	166	170	296	293	46,6	46,0
GA	182	187	310	308	44,4	39,8

Tableau A5.3 : Caractéristiques mécaniques des échantillons

Repère	Effort (daN)	Durée de soudage (périodes)	Durée de forgeage (périodes)
GZ 1	230	10	10
GZ 2	210	9	9
GA	300	14	14

Tableau A5.4 : Paramètres de soudage suivant [NF A 87.001]