وزارة الـتـعـلـيـم الـعـالـي و الـبـحـث الـعـلـمي

BADJI MOKHTAR ANNABA UNIVERSITY
UNIVERSITE BADJI MOKHTAR ANNABA

جـامـعـة باجي مختار عـنـابـة

Faculté des Sciences
Département de Physique

Année 2005

MEMOIRE
Présenté en vue de l'obtention du diplôme de

MAGISTER

ETUDE PAR RESEAUX DE NEURONES DU COMPORTEMENT MECANIQUE D'UN MATERIAU DUCTILE. -APPLICATION AU HDPE-

Option : Matériaux Avancés

Par

Wafia GHABECHE

Jury:

PRESIDENT :	Seddik	BOURAS	M.C.	U. Annaba
EXAMINATEURS :	Hadj Ahmed	ABASSI	Pr	U.Annaba
	Nacereddine	ZEGHIB	M.C.	U.Annaba
MEMBRES INVITES :	Abdel Krim	MOUSSAOUI	C.C.	U.Guelma
	Noureddine	GUERSI	C.C.	U.Annaba
DIRECTEUR DE THESE :	Abdelhamid	SOUAHI	M.C.	U. Annaba

Résumé

Les propriétés mécaniques des matériaux sont d'un intérêt considérable puisqu'elles conditionnent non seulement tous les problèmes de leur mise en œuvre et leur mise en forme mais aussi leur comportement en service dans des applications extrêmement diversifiées.

Pour un problème déterminé, le choix d'un matériau dépendra de ces propriétés mécaniques telles que la résistance, la dureté ou la ductilité. Il est donc nécessaire de les mesurer avec un certain nombre d'expériences que l'on appelle les essais mécaniques.

Dans ce travail, nous voulons effectuer une programmation boite noire (black box), pour modélisé un essai de traction uniaxial.

Dans la première partie, on établit une base de données conséquente relative à ce sujet. Les paramètres pris en compte sont les contraintes en fonction des déformations pour des vitesses de déformation prises entre 10^{-5} - 10^{-3} s^{-1}, d'un polyéthylène de haute densité (HDPE), corrigées par la méthode de Bridgman.

Le réseau choisi est un perceptron à une couche cachée, type feedforward, reconnu pour sa capacité d'approximer n'importe quelle fonction non linéaire. En premier, nous scindons notre ensemble de données en deux sous-ensembles : Le premier pour l'apprentissage, le deuxième pour la validation. Ensuite nous entraînons notre réseau afin qu'il apprenne bien le comportement du HDPE, enfin nous validions et nous testons sa réponse pour des nouvelles valeurs de vitesse de déformation.

Abstract

The mechanical properties of materials are of a considerable interest since they condition not only all the problems of their working implementation but also their behavior in service in extremely diversified applications.

For a determined problem, the choice of a material will depend on these mechanical properties such as resistance, hardness or ductility. It is thus necessary to measure them with a certain number of experiments which one calls the mechanical tests.

In this work, we want to approach an effective law of behavior by the artificial neural networks.

In the first part, we establish a data base consequent relating to this subject. The parameters taken into account are the strains according to the stress for the speeds of deformation taken between 10^{-5} - 10^{-3} S^{-1}, of a high-density polyethylene (HDPE), corrected by the Bridgman's method.

The selected network is a single hidden-layer feedforward, recognized for its capacity to approximate any nonlinear function. First, we divide our whole of data into two subsets: the first for the training, the second for the validation. Then we involve our network so that it learns the behavior well from the HDPE, finally we validate and we test his response for new values speed of deformation.

الملخص

الخصائص الميكانيكية للمواد تمثل أهمية كبرى لأنها لا تحدد فقط كل إشكاليات التصنيع وما قبله، بل أيضا سلوكها أثناء الاشتغال، في استعمالات جد متنوعة.

من أجل تطبيق معين، اختيار مادة ما يتحدد بخصائصها الميكانيكية مثل المقاومة، الصلابة أو الليونة إذا من الواجب قياسها بعدة تجارب ندعوها التجارب الميكانيكية. في هذا البحث، نريد تقريب قانون سلوك حقيقي بواسطة الشبكة العصبية الآلية.

في الجزء الأول ،نحضر قاعدة من المعطيات الخاصة بهدا العمل ،المقاييس المأخوذة بعين الاعتبار هي ، لمتعدد الايثيلان ذو 10^{-1} ثا$^{-3}$ ،10^{-5}تغير الأجهادات مع التشوهات من أجل سرعة تشوه مأخوذة بين 10 كثافة عالية(HDPE) مصححة بطريقة BRIDGMAN

الشبكة المختارة هي عصبون ذو طبقة مخبأة ، نوع الانتشار نحو الأمام معروف بقدرته على تقريب أي دالة خطية ، في البداية ، نجزئ معطياتنا إلى مجموعتين ، الأولى للتعليم والثانية للتحقق ثم ندرب الشبكة لكي تتعلم جيدا السلوك متعدد الأيثيلان ذو كثافة عالية ، وأخيرا نسجل ونجرب استجابته من أجل قيم جديدة للسرعات.

SOMMAIRE

INTRODUCTION GENERALE

ETUDE BIBLIOGRAPHIQUE

I- Polymères en générale, Polyéthylène haute densité (HDPE) en particulier

Introduction 5

I-1 Chimie des polymères 5

I-1.1 Définitions générale 5

I-1.1.1 Monomère 5

I-1.1.2 Macromolécule 6

I-1.1.3 Homopolymère 7

I-1.1.4 Copolymère 8

I-1.1.5 Structure de base macromoléculaire 8

I-1.2 Architecture des macromolécules 9

I-1.3 Classification des polymères selon leur comportementmécanique 9

I-2 Le polyéthylène haut densité (HDPE) 10

I-2.1 Propriétés 13

I-2.1.1 Structure chimique et cristallinité 14

I-2.1.2 Masse molaire 14

I-2.1.3 Distribution des masses molaires 15

I-2.1.4 Rigidité 15

I-2.1.5 Comportement au fluage 16

I-2.1.6 Dureté 16

I-2.1.7 Température de fusion 17

I-2.2 Utilisations les plus courantes 17

Conclusion 18

Références 19

II- Essai de traction

II-1 Aperçu historique 21
II-2 Rappel théorique 22
II-3 Critique des essais conventionnels 26
II-4 Correction de BRIDGMAN 27

Conclusion 31

Références 33

III- Réseaux De neurones : De la biologie à la physique

Introduction 36

III-1 Bref aperçu historique 36
III-2 Le neurone biologique 39
III.2.1 Le potentiel d'action 40
III-3 Le passage 40
III-4 Le neurone formel (artificiel) 43
III.4.1 Fonction d'activation 44
III-5 Réseaux de neurones Artificiels (RNA) 47
III-5.1 Topologie de connexion des neurones 48
III-5.1.1 Réseaux statiques non bouclés 48
III-5.1.2 Réseaux à une couche cachée 49
III-5.2 Apprentissage 51
III-5.2.1 la règle de WIDROW-HOFF 53
III-5.2.2 Algorithme de rétropropagation du gradient 53
III-5.2.3 Algorithme de Levenberg-Marquardt 54
III-5.3 Validation 55
III-5.3.1 Fonction d'autocorrélation des résidus 55
III-5.3.2 Fonction d'intercorrélation 55
III-6 Conception d'un réseau de neurone 56
III-6.1 Choix des échantillons 56
III-6.2 Elaboration de la structure du réseau 56
III-6.3 Apprentissage 56
III-6.4 Validation 56

III-7 Applications 57

III-8 Comparaison avec d'autres méthodes 57

Conclusion 58

Références 59

RESULTATS ET DISSCUSSION

IV- Un réseau pour la prédiction d'une loi de comportement effective.

Objectif 64

IV-1 Construction du réseau 64

IV-1.1 Choix de l'architecture 65

IV-1.2 Choix de la fonction d'activation 67

IV-1.2 Choix de nombre de neurones dans la couche cachée 67

IV-1.2.1 Premier paquet 67

IV-1.2.2 Deuxième paquet 73

IV-1.2.3 Troisième paquet 78

IV-1.2.4 Quatrième paquet 83

IV-2 La méthode Early Stopping 88

Conclusion 92

Références 94

CONCLUSION GENERALE

ANNEXES

ANNEXE1 : Calcul du facteur de triaxialite d'après Bridgman

ANNEXE2 : Algorithme de retro-propagation du gradient

ANNEXE3 : Source expérimentale des données de simulation

NOMENCLATURE

a	Activation du neurone
a, **R**	Rayon de l'éprouvette
b	Biais du neurone
C	Fonction dépendante du matériau
C_j^p	Activation
D, D_0	Diamètres minimums (instantané, initial)
d	Symbole d'incrémentation infinitésimale
E, e	Erreur entre les valeurs actuelles et celles prédites.
E (**w**)	Fonction de coût usuelle : Erreur quadratique moyenne
E	Module d'élasticité ou d'Young
e_x, e_r, e_θ	Déformations
F	Charge ou force
F_T	Facteur de triaxialité
f ()	Fonction de transfert ou activation
h, h'	Demi-hauteurs
H	Hessienne
HDPE	Polyéthylène haute densité
I_j^p	La $j^{ème}$ unité cachée à l'entrée résultante
I_W	Indice de *polymolécularité*
J( )	*Jacobian* de la performance mesurée en fonction des variables de poids et biais.
L, L_0	Longueurs
$\overline{M}_N$	Masse molaire moyenne en nombre.
$\overline{M}_W$	Masse molaire moyenne en poids
N	Nombre des données dans l'ensemble des données,
o_k	Activation de la $k^{ème}$ unité de sortie
p	Potentiel
R_e	Limite d'élasticité
R_m	Résistance à la traction
$R_{0.2}$	Limite conventionnelle d'élasticité
Rc	Rayon de la courbure du profil au centre de la striction

R, **x**	Coordonnées cylindriques
RNA	Réseau de neurones
RMSE	Ecart type de l'erreur de prédiction
RP	Rétropropagation de l'erreur
S, S_0	Sections
s	Somme pondérée
t_k	Activation désirée de la $k^{ème}$ unité de sortie
t	Indice d'itération
U	Entrées de système
v_{kj}	Poids d'une connexion entre la $j^{ème}$ unité cachée et la $k^{ème}$ unité de sortie
W	Vecteur poids
W1, W2, W3, Wn	Coefficients synaptiques ou poids synaptiques
w_{ji}	Poids d'une connexion entre la $i^{ème}$ entrée et la $j^{ème}$ unité cachée
X	Vecteur d'entrée
x_k	l'entrée avec laquelle le poids w est associé à l'instant k.
x_1, x_2, x_3, x_n	Entrées ou excitations
x_i^p	La $i^{ème}$ entrée lorsque le $p^{ème}$ exemple est présenté au réseau
Y_d, y	Sortie désirée (réelle)
$\overline{x}$	La Moyenne
$\hat{r}_{u\varepsilon}$	Fonction d'intercorellation des résidus
$\hat{r}_{\varepsilon\varepsilon}$	Fonction d'autocorélation des résidus
$\hat{y}$	Sortie estimée,

Symboles grecs

α	Angle entre la génératrice du profil et l'axe de l'éprouvette
α	Terme de moment, pas d'adaptation, taux d'apprentissage
β	Fonction dépendante des déformations
δ	Rayon
φ, φ'	Angles
ε	Erreur de prédiction
ε_R	Déformation à la rupture

$\dot{\varepsilon}$	Vitesse de déformation
ε_{rr}	Composante radiale de la déformation
ε_{zz}	Composante axiale de la déformation
ε_{eff}	Déformation effective
$\dot{\varepsilon}_{eff}$	Vitesse de déformation effective
ε_N	Déformation nominale
ρ	Déplacement radial
θ	Coordonnées cylindriques
θ	Paramètres du réseau, poids et biais.
σ	Contrainte
σ_N	Contrainte nominale
σ_r	La contrainte réelle
σ_{zz}	Composante axiale ou contrainte de KIRCHOFF
σ_{rr}	Composante radiale de la contrainte
$\sigma_{\theta\theta}$	Composante circonférentielle de la contrainte
$\sigma_{r\theta}$, $\sigma_{x\theta}$, σ_{rx}	Composantes orthogonales
μ	Coefficient de moment
ΔL	Allongement
Δ_k	Différence entre la sortie attendue et la sortie effective du neurone à l'instant k
Δv_{kj} , Δw_{ij}	Termes d'adaptation des poids au cours de l'apprentissage
η	Taux d'apprentissage

LISTE DES FIGURES

Fig.	Titre	Page
I.1	Représentation d'une macromolécule linéaire du (polyéthylène) : (a) conformation aléatoire d'une molécule de 7000 unités CH_2 et d'une masse de 98000 $gmol^{-1}$ à l'échelle de 10 nm ; (b) et(c) segment de 4 unités CH_2-CH_2 à l'échelle subnanométrique ; (d) distribution du potentiel électrostatique du même segment.	7
I 2	Répartition des thermoplastiques dans le monde	12
I.3	Procédé Basse Pression : Catalyse de Ziegler Natta	13
I.4	Comparaison entre la macromolécule d'un polyéthylène haute densité (HDPE), polyéthylène moyenne densité (MDPE) et d'un polyéthylène basse densité (LDPE)	14
I 5	Schéma de principe de l'enduction	16
II.1	Expérience de traction sur fils par Leonardo DA VINCI	21
II.2	Eprouvettes d'essai de traction	22
II.3	Courbes contrainte-allongement	23
II.4	Courbe conventionnelle de traction	23
II.5	Courbe de déformation nominale	25
II.6	Eprouvette axisymétrique en forme de sablier	27
II.7	Courbe dynamométrique	28
II.8	$\sigma_{zz} = f(\varepsilon)$, au centre de la striction	29
II.9	Variation de la contrainte effective σ_{eff} en fonction de la déformation vraie au centre de la striction	31
III.1	Le neurone biologique	39
III.2	Les mécanismes à l'origine du potentiel d'action	40

III.3	Le neurone formel de McCulloch et Pitts	42
III.4	Fonction de transfert binaire	43
III.5	Mise en correspondance neurone biologique / neurone artificiel	43
III.6	Le neurone artificiel	44
III.7	Fonction logique OU et le neurone correspondant	45
III.8	Réseau de neurones artificiels	47
III.9	Un Perceptron multicouche	49
III.10	Réseau à une couche cachée	49
VI.1	Diagramme de la procédure	65
VI.2	schéma bloc de l'algorithme de prédiction	66
VI.3	Type de fonctions d'activation utilisées	67
VI.4, 11, 18, 25, 32	Architecture du réseau choisi	68, 73, 78, 83, 89.
VI.5, 12, 19, 26.	Comparaison entre les valeurs expérimentales et ceux obtenus après apprentissage	70, 75, 80, 85
VI.6, 13, 20, 27.	Erreur due à l'apprentissage	70, 75, 80, 85.
VI.7, 14, 21, 28.	Test de validation.	71, 76, 81, 86.
VI.8, 15, 22, 29.	Erreur due à la validation	71, 76, 81, 86.
VI.9, 16, 23, 30.	Teste de généralisation pour une vitesse de déformation $\dot{\varepsilon}_T = 3.5\ 10^{-4}\ s^{-1}$	72, 77, 82, 87.
VI.10, 17, 24, 31.	Erreur commise sur l'ensemble de test.	72, 77, 82, 87.
VI.33	Evolution de l'erreur.	90

VI.34	Comparaison entre les valeurs expérimentales et ceux obtenues après simulation pour une vitesse de déformation $\dot{\varepsilon}_T = 3.5\ 10^{-4}\ s^{-1}$	91
VI.35	Erreur commise sur l'ensemble d'apprentissage	91
1.1	Influence du profil sur la contrainte effective dans le cas de la traction	
1.2	Diagramme pour l'analyse géométrique approximative des conditions sur les contraintes dans le voisinage d'une instabilité plastique : striction.	
1.3	Equilibre des contraintes dans le plan médian d'une éprouvette soumise à un étirage.	
2.1	Définition des notions pour un MLP.	
3.1	Variation de la contrainte en fonction de la déformation pour des essais effectués avec des différentes vitesses de déformation	

LISTE DES TABLEAUX

Tab.	Titre	Page
I.1	Schéma de classification des polymères synthétiques considérant la régularité et rigidité de leurs chaînes et la nature de liaisons intermoléculaires	11
I.2	Propriétés et application du HDPE	17
II.1	Caractéristiques du HDPE utilisé	28
VI.1	Paramètres choisis.	65
VI.2, 3, 4, 5, 6.	Choix du nombre de neurones dans la couche cachée	68, 73, 78, 83, 89,

Introduction générale

Les propriétés mécaniques des polymères sont d'un intérêt considérable puisqu'elles conditionnent non seulement tous les problèmes de leur mise en œuvre en amont et de leur mise en forme en aval mais aussi leur comportement en service dans des applications extrêmement diversifiées.

Pour un problème déterminé, le choix d'un matériau polymère dépendra de ces propriétés mécaniques telles que la résistance, la dureté ou la ductilité. Il est donc nécessaire de les mesurer avec un certain nombre d'expériences que l'on appelle les essais mécaniques.

L'expérimentation n'est pas toujours accessible, faute de moyen ou difficultés dans la réalisation des expériences, alors on passe à la simulation soit comme soutien à l'expérimentation pour des raisons purement économique ou pour des raisons de sécurité.

La simulation nous permet de prédire le comportement des matériaux et leurs propriétés dans des conditions multiples. Parmi les moyens de prédiction dans un environnement non linéaire (ce qui est le cas dans notre travail) on distingue les réseaux de neurones artificiels.

Dans le premier chapitre, il est introduit quelques définitions et généralités sur les polymères. Une description succincte du polyéthylène (matériau de notre étude), étant donnée.

Nous avons présenté, dans le deuxième chapitre, un essai très largement utilisé pour l'étude de comportement mécanique des matériaux : L'essai de traction ; nous avons présenté son aspect théorique, suivi par des critiques des essais conventionnels : Nous avons introduit les principales raisons pour lesquelles nous effectuons des corrections lors d'une traction uniaxiale : La correction de BRIDGMAN en est une.

Lors de l'émergence d'une nouvelle technique, le chercheur se demande naturellement en quoi elle peut lui être utile. Si elle est dotée d'un nom plus métaphorique que scientifique – ce qui est évidemment le cas pour les réseaux de neurones –, la réponse à cette question doit être particulièrement précise et motivée.

C'est la raison pour laquelle le troisième chapitre est consacré à un exposé des arguments théoriques et pratiques qui justifient l'utilisation des réseaux de neurones. La mise en œuvre des réseaux de neurones est généralement simple ; La tentation peut être grande d'appliquer

cette technique d'une manière irréfléchie ou inadaptée, ce qui ne peut conduire qu'à des déceptions.

Nous commençons par introduire quelques définitions essentielles et généralités sur le neurone biologique et formel et les réseaux de neurones artificiels.

Nous introduisons la propriété fondamentale des réseaux de neurones à apprentissage supervisé : L'approximation non linéaire parcimonieuse. Nous montrons que cette propriété fait des réseaux de neurones d'excellents outils de modélisation non linéaire. La notion d'apprentissage apparaît alors de manière très naturelle, et nous insistons sur le fait que ce terme doit être utilisé avec précaution, car l'apprentissage des réseaux de neurones n'a rien à voir avec l'apprentissage au sens de la psychologie ou de l'intelligence artificielle : La mise en œuvre des réseaux de neurones se révélera, en grande partie, comme constituant une extension non linéaire de méthodes statistiques bien connues, ce qui suffit à justifier la très grande étendue de leur champ d'application.

Dans le chapitre VI, nous exposons notre travail qui consiste à approcher une loi de comportement effective d'un polyéthylène de haute densité en traction uniaxiale. Nous devisons nos données en deux sous ensembles : Le premier est pour l'apprentissage, le deuxième est utilisé pour la validation ; Nous testons chaque fois la réponse du réseau.

Enfin, nous allons clôturer cette étude par une conclusion générale, dans laquelle nous reprenons quelques définitions et résultats. Nous citerons ensuite quelques perspectives.

La frontière que l'homme tente de définir dépend essentiellement de son degré de connaissance de la matière et des controverses.

INTRODUCTION

Les polymères sont une classe de matériaux qui, dans l'ensemble, sont récents. Ils n'ont pas connu de développement industriel qu'à partir de la deuxième moitié du vingtième siècle. Par conséquent il existe pour ces produits un important potentiel de progrès.

I-1 CHIMIE DES POLYMERES

De nombreux matériaux sont non cristallins : Leurs structures ne sont pas faites d'un réseau régulier tridimensionnel d'atomes. Ces solides ne peuvent pas donc, comme les solides cristallins réels, se décrire en terme de défauts. Cependant, même si ces solides ne sont pas cristallins, ils montrent un ordre local géométrique ou chimique. Ils n'ont pas d'ordre à grande distance comme les cristaux, vu que ces groupements élémentaires sont liés au hasard dans l'espace solide.

Les solides non cristallins étaient souvent appelés *amorphes* jusqu'à ce que les rayons X mirent en évidence un ordre local à l'échelle du groupement élémentaire : Les atomes sont arrangés à l'intérieur de ces unités d'une manière ordonnée alors que dans un solide, réellement amorphe, l'arrangement se fait au hasard.

Bien que leurs propriétés soit très variés, tous les polymères sont faits de longues molécules avec un squelette covalent d'atomes de carbone. Ces molécules sont liées les unes aux autres par les liaisons hydrogènes ou Van Der Waals (liaisons secondaires) ou par des ponts covalents supplémentaires.

La température de fusion d'une liaison secondaire est très basse, proche de la température ambiante. Ainsi on utilise ces matériaux très près de la température absolue de fusion de leurs liaisons secondaires, cependant loin de la fusion du squelette covalent qui est bien résistant.

I-1.1 Définitions générales

I-1.1.1 Monomère

Les monomères sont des entités constitutives qui s'associent par l'intermédiaire de liaison chimique pour former la chaîne polymère. La plus petite unité qui se répète régulièrement dans

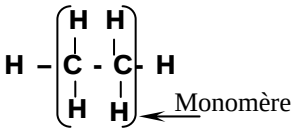
le polymère est appelée unité constitutive répétitive (pour le polyéthylène : Monomère « $CH_2=CH_2$ », l'unité monomère « $-CH_2-CH_2$ »)

Elle peut être plus grosse ou plus petite que l'unité monomère, elle peut être formée à partir de plusieurs monomères comme seulement d'une fraction d'un monomère.

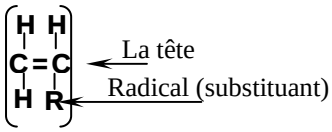
Pour le POLYAMIDE 66 synthétisé + { HEXAMETHYLENEDIMINE ($H_2N-(CH_2)_6-NH_2$) ; ACIDE ADIPIQUE ($HOOC-(CH_2)_4-COOH$)

↳ ($-CO-(CH_2)_4-CO-NH-(CH_2)_6-NH-$)

La plupart des polymères ne sont pas symétriques comme le polyéthylène. Les composés vinyliques sont des exemples de monomères dissymétriques :



Polyéthylène PE
(Symétrique)



Monomère vinylique
(Dissymétrique)

I-1.1.2 Macromolécule

Ce n'est que dans les années 1920 que la notion de macromolécule a commencé à être largement acceptée par la communauté scientifique. Hermann Staudinger, Prix Nobel de chimie en 1953, en a été l'un des principaux artisans. Il a démontré que les matières polymères comme la cellulose, la soie et le caoutchouc étaient formées de longues chaînes moléculaires, qualifié de macromolécules. Cette idée a ouvert la voie à une recherche plus systématique, tant chimique que physique, sur la synthèse, la structure et le comportement des polymères solides.

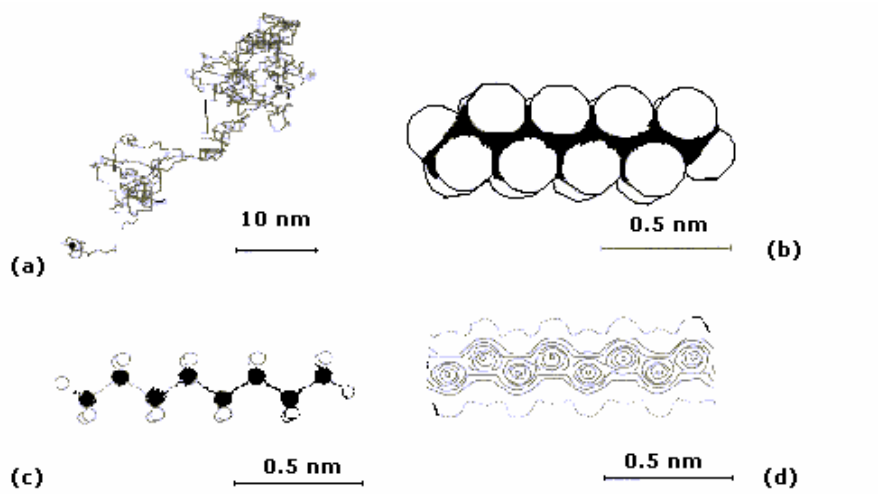
Les macromolécules sont typiquement formées de plusieurs milliers de motifs structuraux simples appelés *mères*, du grec *meros* qui signifie « partie », liés entre eux par des liaisons primaires, liaisons covalentes, à l'échelle d'une centaine de nanomètres, la plupart des macromolécules synthétiques se présentent de façon semblable, sous forme de fils longs et souples (figureI.1(a)) Indépendamment de leur structure chimique détaillée, ces ficelles montrent les trois caractéristiques physiques fondamentales communes à toute macromolécule :

- Grande longueur
- Forte anisotropie des segments ;
- Souplesse moléculaire importante.

Schématiquement, ces trois caractéristiques permettent d'expliquer tous les phénomènes physiques spécifiques aux matériaux polymères. Elles conditionnent aussi leurs propriétés uniques, qui représentent leur atout principal :

- Facilité de mise en œuvre ;
- Adaptabilité et grande déformabilité.

Un zoom sur ces ficelles révèle la géométrie d'une chaîne et sa structure chimique. Les figureI.1 (b) et (c) montrent un détail d'une molécule de polyéthylène (PE) à l'échelle du nanomètre. La distribution du potentiel électrostatique a également été indiquée (figureI.1 (d)) On peut apprécier facilement la mère du polyéthylène, les groupes CH_2, et la grande anisotropie de distribution du potentiel associé à cette molécule



FigureI.1. Représentation d'une macromolécule linéaire du (polyéthylène) : (a) conformation aléatoire d'une molécule de 7000 unités CH_2 et d'une masse de 98000 $gmol^{-1}$ à l'échelle de 10 nm ; (b) et (c) segment de 4 unités CH_2 - CH_2 à l'échelle subnanométrique ; (d) distribution du potentiel électrostatique du même segment.

Une chaîne de polyéthylène contient normalement entre 2000 et plus de 20 000 mères.

C'est la répétition d'un grand nombre d'unités structurales liées entre elles par des liaisons covalentes qui donne lieu aux macromolécules linéaires, anisotropes et flexibles. Pratiquement tous les polymères de commodités sont constitués selon ce principe.

La masse moyenne des molécules d'un polyéthylène linéaire tel qu'il est employé dans la tuyauterie est typiquement 100 000 $gmol^{-1}$, ce qui correspond à une longueur étendue de 900 nm. Par comparaison, le segment de polyéthylène représenté dans la figureI.1 (b) et (c) est composé de quatre unités $–[CH_2–CH_2]$-ce qui lui donne une masse molaire de 112 $gmol^{-1}$ et une longueur

étendue de 0.88 nm. A l'échelle de la figureI.1 (b), une molécule d'une masse molaire de 100 000 $gmol^{-1}$ aurait une longueur étendue de 42.9 m !

L'anisotropie des molécules est due au fait que la connexion le long de l'axe de la chaîne est assurée par des liaisons covalentes fortes, qui résultent du partage des électrons par le recouvrement des orbitales atomiques des éléments qui la composent, tandis que la cohésion latérale entre des segments moléculaires est assurée par des liaisons secondaires faibles, aussi appelées forces de Van Der Waals. Cette anisotropie conduit non seulement à une grande anisotropie de la résistance et de la rigidité lorsque les segments d'un polymère ont une orientation préférentielle (dans un cristal par exemple), mais aussi à une anisotropie d'autre propriétés physiques, comme les propriétés optiques, diélectriques et thermiques.

La souplesse importante des chaînes résulte de la facilité avec laquelle les groupements et segments moléculaires peuvent effectuer des rotations autour des liaisons covalentes. Le comportement dynamique des chaînes, leur arrangement spatial et la viscoélasticité des matériaux polymères en dépendent directement.

I-1.1.3 Homopolymère

Si un type seulement est impliqué dans une réaction de polymérisation, le produit synthétisé est un homopolymère.

I-1.1.4 Copolymère

Si la polymérisation est réalisée à partir d'un minimum de deux types de monomères, le polymère obtenu est un copolymère

- Copolymère alterné : Différents monomères sont associés en alternance selon une séquence régulière.
- Copolymère statique : Les motifs sont associés selon une distribution irrégulière mais obéit à des lois statistiques (les copolymères aléatoires sont un cas particulier, ils sont tels qu'il n'existe pas de distribution définie des différents unités monomères)
- Copolymère en blocs : Ce sont des polymères dont les chaînes contiennent de longue séquence d'unités monomères du même type, ils sont intéressants en tant que compatibilisants des mélanges d'homopolymères correspondants. Ils se placent à l'interface et chacune des séquences s'enchevêtre avec l'homopolymère parent à fin d'assurer une bonne adhésion.

I-1.1.5 Structure de base macromoléculaire

La raison pour laquelle il existe autant de composés carbonés est la capacité du carbone à former des liaisons covalentes avec lui-même et donc de former de nombreuses chaînes de différentes longueurs, ainsi que des cycles de différentes tailles. La plupart des composés organiques sont fort sensibles à la chaleur et se décomposent généralement en dessous de 300°C. Ils ont tendance à être insolubles dans l'eau, en tout cas moins solubles que les sels inorganiques. Par contre, et à l'inverse de tels sels, ils ont tendance à être solubles dans les solvants organiques tels que l'éther ou l'alcool.

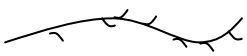
Les aspects physico-chimiques déterminent très largement le comportement physique des polymères : La composition élémentaire, la disposition des atomes et liaisons, l'ordre d'enchaînement et les aspects topologiques définissent la structure de base d'une macromolécule : Elle dépend de la nature chimique des unités monomères et des paramètres de synthèse (température, pression, nature de l'amorceur et des catalyseurs, rapport de réactivité des monomères, milieu réactionnel, etc.)

I-1.2 Architecture des macromolécules

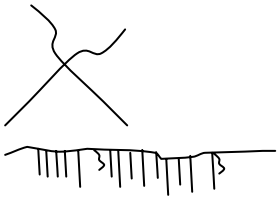
La majorité des polymères produits aujourd'hui est constituée de macromolécules linéaires. Toute-fois la chimie permet aussi la réalisation des architectures macromoléculaires non linéaires très variées qui ont un intérêt industriel certain. Selon son architecture, une macromolécule peut être :



Linéaire (Polyéthylène de haute densité : HDPE)



Homopolymère branché (Polyéthylène de basse densité : LDPE)



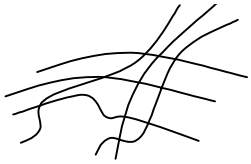
Homopolymère à branches courtes (Polyéthylène linéaire de basse densité : LDPE)

Polymère en étoile de f =4 (Styrène- butadiène bloc copolymère)

Polymère en peigne (Polymacromonomère de styrène)



Polymère cyclique (Polycarbonate cyclique)



Polymère réticulé, une réticulation assurée par des liaisons covalentes

(Résines époxydes, caoutchouc vulcanisé)

I-1.3 Classification des polymères selon leur comportement mécanique

On classe les polymères, selon leur comportement thermomécanique, en trois grandes classes : Les thermoplastiques, les thermodurcissables et les élastomères, cette nomination est d'origine historique.

Thermoplastique signifie littéralement que l'échantillon devient plastique par chauffage, les thermoplastiques sont constitués de chaînes plus au moins linéaires liées entre elles par des forces de van der Waals, selon la régularité de leur chaînes, ces polymères peuvent cristalliser (thermoplastes semicristallins) ou non (thermoplastes amorphes). Etant donné la faiblesse des forces van der Waals, ils ramollissement ou fondent dans une plage de température sise entre 80 et 225°C. Par contre un thermodurcissable durcit et prend une forme sous l'effet de la chaleur ou de la lumière. Toutefois, de nombreux polymères ne se conforment pas à ces dénominations strictes. Ainsi, des thermoplastiques comme le polytetrafluoroéthylène (PTFE) sont difficiles, voire impossibles à mettre en oeuvre par thermoformage, alors que des thermodurcissables comme les époxydes, certains polyuréthannes et silicones durcissent sans apport de chaleur, à la température ambiante ou sous-ambiante.

La véritable distinction entre ces deux classes de polymères provient du fait que dans le premier cas, les macromolécules restent chimiquement séparées les unes des autres après la mise en œuvre, ce qui permet un remoulage à discrétion. Dans le deuxième cas, des petites molécules ou chaînes linéaires sont réticulées pendant le durcissement. Elles sont toutes reliées entre elles par des liaisons primaires et stables. L'échantillon forme ainsi une seule super-macromolécule (un réseau 3D) et il ne peut plus être remoulé. Comme conséquence pratique de cette différence, un thermodurcissable peut se gonfler, mais ne se dissout dans aucun solvant. En fracturant un échantillon réticulé, on casse automatiquement des liaisons covalentes avec création des radicaux libres. Les polymères de la troisième famille, celle des élastomères, se distinguent par les faibles valeurs des forces d'attraction entre les chaînes et de leur modules élastiques (à température ambiante) et par leur capacité élevée de déformation (quelques centaines de pour cent)

Le schéma des différentes familles des polymères solides est complété par les élastomères thermoplastiques et par les polymères à cristaux liquides (*liquid crystalline polymers* ou LCP) Les caractéristiques fondamentales de ces cinq grandes classes sont résumées dans le tableauI.1.

TableauI.1. Schéma de classification des polymères synthétiques considérant la régularité et rigidité de leurs chaînes et la nature de liaisons intermoléculaires.

Classe	**Régularité des chaînes (absences des groupes latéraux encombrants)**	**Souplesse des chaînes**	**Connexion entre molécules principalement par**
Thermoplastiques amorphes	Chaînes irrégulières	Moyenne	Liaisons secondaires
et semicristallins	Chaînes linéaires régulières	Moyenne ou élevée	Liaisons secondaires
Elastomères	Chaînes linéaires	Elevée	Liaisons primaires (réticulation moyenne)
Elastomères thermoplastiques	Copolymères en bloc élastomère-thermoplastique	Selon composant	Liaisons secondaires dans les domaines thermoplastiques
Polymères à cristaux liquides	Chaînes linéaires avec ou sans segments latéraux	Chaînes ou segments relativement rigides	Liaisons secondaires
Thermodurcis	Segments réguliers	Restreintes par réticulation	Liaisons primaires (densité de réticulation moyenne ou forte)

Les matériaux thermoplastiques sont des solides généralement solubles, formés des chaînes distinctes bien compactées, liées entre elles par les seules liaisons secondaires (forces de van der Waals ou dipolaires, liaisons d'hydrogènes, etc.) et mis en forme (même de façon répétitive) par chauffage et refroidissement.

Les matériaux caoutchoutiques ou élastomères sont des solides souples et gonflables formés de chaînes peu compactées et (à température ambiante) très mobiles ; Pour assurer une stabilité de forme indispensable, diminuer leur propension au fluage et prévenir une souplesse excessive, les caoutchoucs sont généralement réticulés, c'est-à-dire que leurs chaînes sont liées entre elles par des liaisons primaires et stables (en densité faible, de l'ordre du pour cent des sites disponibles à cet effet)

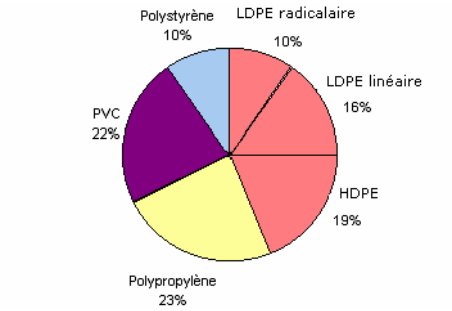
Les élastomères thermoplastiques sont des solides qui ont un comportement semblable à celui des élastomères et une formabilité semblable à celle des thermoplastiques. Ils sont formés de copolymères en bloc, qui réunissent au sein d'une même chaîne des segments d'un thermoplastique rigide (par exemple, du PS) et d'un élastomère (par exemple, du polybutadiène (PBD))

Les polymères à cristaux liquides (LCP) comportent des chaînes rigides et linéaires qui forment des domaines cristallins liquides soit en masse fondue (polymères thermotropes) soit en solution (polymères lyotropes). Le degré d'orientation moléculaire élevé qui en résulte est maintenu, voire augmenté lors de la mise en œuvre et confère aux LCP solides une stabilité de forme exceptionnelle et un coefficient de dilatation thermique faible et comparable à celui des céramiques.

Les matériaux thermodurcissables sont des solides infusibles et insolubles qui forment un réseau tridimensionnel à partir des segments moléculaires liés entre eux par une densité élevée de liaisons primaires ; Les thermodurcis sont mis en forme de façon irréversible par réaction chimique des molécules d'un *durcisseur* avec les petites molécules d'une *résine* et /ou avec des chaînes linéaires. La structure chimique de la résine de départ, le choix de durcisseur et le degré de réticulation influent l'ensemble des propriétés thermomécaniques des thermodurcis (module d'élasticité, température de transition vitreuse, résistance mécanique). Vu leur faible viscosité avant durcissement et les propriétés chimiques et mécaniques finales favorables, les thermodurcis sont largement employés comme matrice dans les matériaux composites structuraux.

Un nombre croissant des systèmes polymères à fonction spécifique est obtenu par combinaison ordonnée (par voie chimique ou physique) des segments d'une origine et longueur très variées. Ils échappent à cette classification.

Le nombre de matières plastiques commercialisées est bien plus élevé, multiplié en plus par les variations quasi infinies de leur composition (mélanges de divers polymères, de charges, agents renforçateurs et additifs utilisés)



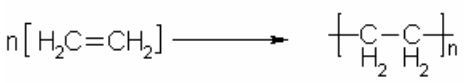
FigureI.2. Répartition des thermoplastiques dans le monde. [SCHULTZ N., 2000]

Le grand nombre des polymères donne également une idée de leur importance économique. En effet, le marché mondial actuel des matières plastiques est de l'ordre de 120 millions de tonnes

par an, en volume ce marché est aujourd'hui équivalent à celui de l'acier. Le polyéthylène haute densité (HDPE) représente 19% des thermoplastiques produits dans le monde.

I-2 LE POLYETHYLENE HAUT DENSITE (HDPE)

Le polyéthylène haute densité est un polymère thermoplastique de grande diffusion, il a des propriétés qui lui confèrent une malléabilité à chaud et une thermoplasticité réversible. Il fait partie de la famille des polyoléfines, au même titre que les polyéthylènes basse ou moyenne densité ou bien le propylène. Il est issu de la polymérisation de l'éthylène gazeux :

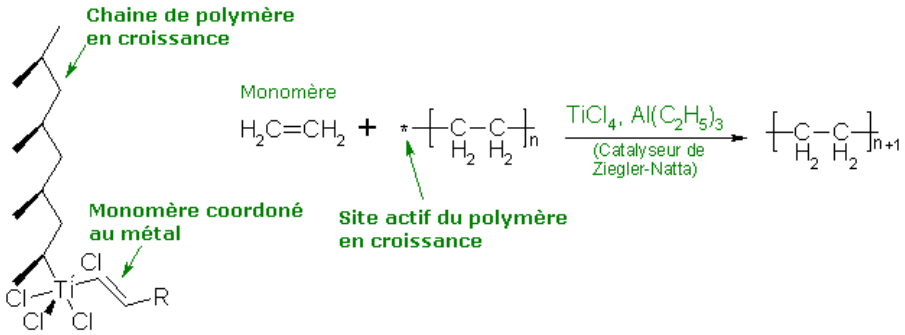


Il est obtenu par des procédés de polymérisation de l'éthylène à basse pression, seul ou avec des comonomères. Les premières unités de production datent du milieu des années cinquante. 1956 était l'année de démarrage de la première unité qui avait utilisé le procédé Ziegler.

Dans les années soixante, des améliorations furent apportées au procédé Ziegler par l'utilisation de catalyseurs superactifs permettant de supprimer la coûteuse opération d'élimination des résidus catalytiques.

Le milieu de polymérisation est tout de même soumis à une pression avoisinant les 50 bars ! C'est uniquement pour le différencier du précédent procédé que celui-ci a été baptisé *basse pression*.

C'est grâce à la présence d'un catalyseur organométallique Ziegler- Natta que va se faire la polymérisation. Le catalyseur permet d'associer autour d'un centre métallique, monomère et site actif de la molécule en croissance.



FigureI.3. Procédé Basse Pression : catalyse de Ziegler Natta. [SCHULTZ N., 2000]

De plus, les carbones asymétriques (s'ils sont associés à différents groupements) auront la même configuration absolue. Ce qui permet d'avoir des polyéthylènes isotactiques.

Industriellement, le procédé le plus couramment employé est celui qui consiste à faire réagir catalyseur et monomère quand ils sont en suspension pour obtenir la plus grande surface de contact disponible en employant des réacteurs plus classiques.

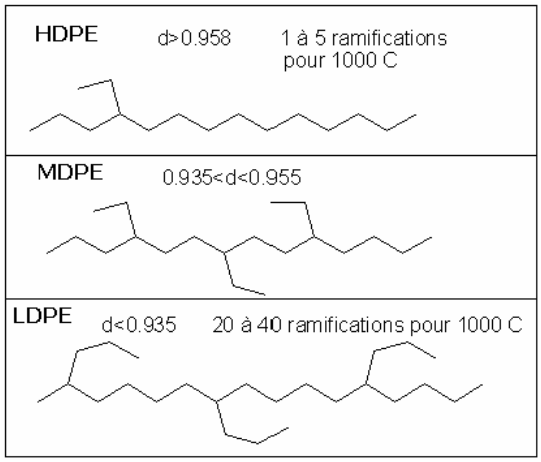
I-2.1 Propriétés

Un objet fait en HDPE a des propriétés qui dépendent de :

- Ses paramètres structuraux : Cristallinité, masse volumique et distribution des masses molaires ;
- De la conception de la pièce ;
- Des conditions de mise en œuvre ;
- De paramètre d'utilisation comme le type de la durée des contraintes, la température.

I-2.1.1 Structure chimique et cristallinité

Le HDPE est un polymère semi cristallin. L'introduction de branchements courts dans la chaîne carbonée, favorisant l'encombrement stérique, abaisse la masse volumique du HDPE et sa cristallinité (figureI.4).



FigureI.4. Comparaison entre la macromolécule d'un polyéthylène haute densité (HDPE), polyéthylène moyenne densité (MDPE) et d'un polyéthylène basse densité (LDPE) [SCHULTZ N., 2000]

Le HDPE est d'autant plus cristallin que :

- Le nombre de branchements courts est plus faible : En fonction du taux de comonomère incorporé, la masse volumique des polyéthylènes haute densité peut varier de 940 kg/m^3 à plus de 965 kg/m^3. Lors de la cristallisation du polymère fondu, les chaînes se replient dans des plans pour former des cristallites. Entre ces cristallites, le polymère bien moins ordonné constitue la phase amorphe. Les cristallites s'organisent dans une superstructure : Les sphérolites ;
- La masse molaire est plus basse : La masse volumique d'un HDPE homopolymère d'indice de fluidité de 5 g/min (à 190°C sous une charge de 2.16 kg) es d'environ 965 kg/m^3. Celle d'un homopolymère d'indice de fluidité de 2 g/10min (à 190°C sous une charge de 21.6 kg) est inférieure à 954 kg/m^3 ;
- La vitesse de refroidissement est plus lente : Par exemple, en fonction du temps de refroidissement, la masse volumique d'une résine d'indice de fluidité sous 2.16 kg égal à 5 g/10 min peut varier de 955 kg/m^3 (refroidissement rapide) à 967 kg/m^3 (refroidissement lent)

I-2.1.2 Masse molaire

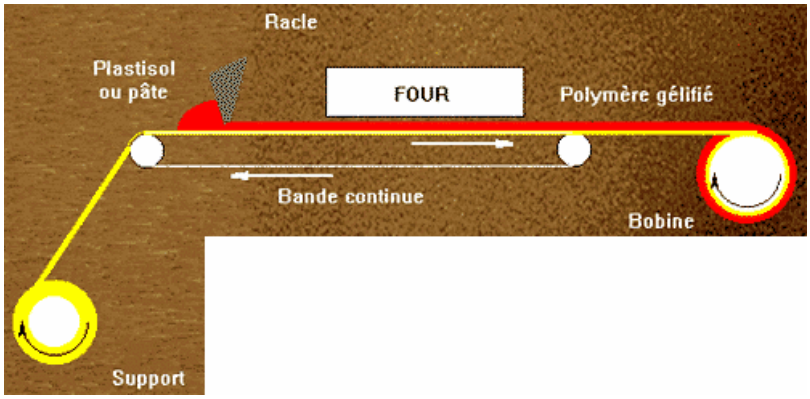
Lorsque la longueur de la chaîne hydrocarbonée augmente, la masse molaire augmente : Le polymère est plus visqueux à l'état fondu donc plus difficile à mettre en œuvre, mais ses propriétés mécaniques sont meilleures à l'état solide. La longueur des chaînes est généralement comprise entre 2 000 et 50 000 unités monomères.

I-2.1.3 Distribution des masses molaires

Le HDPE est constitué de macromolécules de longueurs différentes donc de masses molaires différentes. La largeur de la distribution des masses molaires (appelée indice de *polymolécularité* (I_W) ou de *polydispersité*) peut être définie comme le rapport de la masse molaire moyenne en poids ($\overline{M_W}$) à la masse molaire moyenne en nombre ($\overline{M_n}$).

Dans la pratique, on utilisera un HDPE ayant une distribution étroite de masses molaires pour :

- Le moulage par injection : C'est pour que les pièces aient un minimum de distorsion et de tensions internes, préjudiciables à la résistance au choc et à la résistance à la fissuration sous contrainte
- Le moulage par rotation
- L'enduction (figureI.5)



FigureI.5. Schéma de principe de l'enduction [Plasturgie CD-ROM, 1999]

Au contraire, une distribution plus large de masses molaires sera appréciée pour l'extrusion-soufflage de corps creux, l'extrusion de films et de tuyaux, ainsi que pour le gainage des câbles. En effet, quand la distribution des masses molaires est plus large :

- La mise en œuvre est plus facile ;
- Les extrudats ont un meilleur aspect ;
- L'étirabilité à l'état fondu est améliorée ;

Mais les tensions internes sont plus élevées.

I-3.1.4 Rigidité

La rigidité d'une pièce en HDPE est d'autant plus élevée que la cristallinité de polymère est importante ; Dans une moindre mesure, elle baisse quand la masse molaire augmente.

I-3.1.5 Comportement au fluage

Le fluage caractérise l'évolution au cours du temps (allongement, fléchissement, écrasement) d'une pièce soumise à une charge constante : C'est une manifestation de la viscoélasticité des plastiques.

Il faut tenir compte de cette propriété pour le dimensionnement de pièces soumises à des contraintes de longue durée. Le comportement au fluage peut être évalué au cours d'essais mécaniques simples de traction monoaxiale ou de flexion ; Il est d'autant meilleur que la masse molaire est importante.

I-3.1.6 Dureté

Le HDPE a une dureté relativement faible, elle varie peu d'une résine à l'autre et augmente avec la cristallinité.

I-3.1.7 Température de fusion

La température de fusion, comprise entre 125 et 135 °C, et l'enthalpie de fusion du HDPE sont mesurées par analyse thermique différentielle : La température de fusion correspond à l'abscisse du sommet du pic de fusion, l'enthalpie est obtenue par intégration de la surface du pic.

Ces caractéristiques ainsi que le point de ramollissement Vicat sont d'autant plus élevés que le taux de cristallinité est plus élevé ; De plus, la température de fusion dépend aussi de la structure moléculaire induite par le procédé de polymérisation [DOSSOGNE R., 1996]

I-3.2 Utilisations les plus courantes

- Casiers de manutention (bouteilles, poissons, industries)
- Articles ménagers, jouets, sièges, luminaires
- Isolation électrique, réservoirs, chaudronnerie anticorrosion
- Mono- multifilaments, fils plats étirés, corderies, sacs

Le P.E basse pression est toujours de haute densité, et le P.E haute pression est toujours de basse densité.

TableauI.2. Propriétés et applications du HDPE

PROPRIETES	APPLICATIONS	
Barrière eau, gaz, UV Inertie chimique Opacité Rigidité Moulabilité Résistance aux chocs Résistance à l'abrasion, glisse Tenue à la pression Tenue aux températures	Accessoires de salles de bain Bacs de rangement Bidons (huile moteur, produits chimiques) Bonbonnes, cuves Bouteilles de lait Caillebotis Caisses, casiers Canalisations (gaz, eau) Conteneurs Coques de chaussures de ski Coques de planches à voile Corbeilles à papier Cubitainers Flacons pour cosmétiques	Flacons ménagers Gainages de câbles Jouets Kayaks, canoës Mandrins, bobinages Palettes Poubelles Réservoirs d'essence, d'eau Revêtement et semelles de skis Seaux Serres Tubes, Tuyaux

CONCLUSION

Avec des taux de croissance annuels supérieurs aux taux moyens de l'industrie, les matériaux polymères ont vu leur champ d'application s'étendre, notamment dans des milieux où les notions de fiabilité et de coût prévalent, la maîtrise de leur comportement mécanique devient très important.

Les essais mécaniques à l'état solide revêtent une importance accrue dans la mesure où ils définissent la pertinence des données sur lesquelles les industriels s'appuient pour la conception des pièces nécessaires à des applications diverses.

REFERENCES

DOSSOGNE R.
Polyéthylène haute densité ;
(1996) Technique de l'ingénieur, traité des matériaux non métalliques, pp. A 3 315.

KAUSCH H.H., HEYMANS N., PLUMER C. J., P. DECROLY
Matériaux polymères : Propriétés mécaniques et physiques, principe de mise en œuvre ;
(2001) Traité des matériaux, 14 (19), pp. 34-38.

Plasturgie CD-ROM
Caoutchouc et composite ;
(1999) Edité par IMPACT - Cabinet conseil en plasturgie.

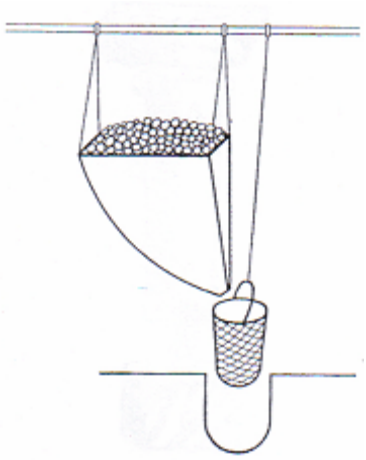
SCHULTZ N.
(2000) 2ème Année DUT Chimie.

II-1 APERÇU HISTORIQUE

Chaque sollicitation peut donner lieu à un essai particulier, qui s'effectue le plus souvent sur des éprouvettes extraites du matériau que l'on veut étudier [BASQUIN R., LEMASSON G., 1964]

Cette méthodologie d'expérimentation n'a eu naissance qu'au 16ème siècle. Avant, les données mécaniques viennent de la pratique : *Ça passe ou ça casse !*

Léonard de Vinci (1452-1519) fait figure de pionnier « La mécanique est le paradis des mathématiques car c'est là où on recueille les fruits », il est le premier à réaliser des expériences sur des fils en tension (figureII.1). Il montre que leur résistance diminue lorsque leur longueur augmente : *Je sais donc je pense.*



FigureII.1. Expérience de traction sur fils par Leonardo DA VINCI

Depuis, l'expérimentation n'a cessé de progresser, Galileo Galilei Linceo (1564-1642), avec ses travaux sur la tension et la flexion des poutres, invente sans le nommer la notion de contrainte en remarquant que la résistance des colonnes en traction est proportionnelle à leur section.

Il ne semble pas que les constructions anciennes aient fait l'objet d'études prévisionnelles concernant leur résistance. L'absence de souci d'économie de matière, le sens élevé de l'esthétique. Une forme esthétique est souvent une forme optimale vis-à-vis de la résistance, des connaissances empiriques fondées sur l'expérience ont permis la réalisation d'ouvrages durables. En 1678, Robert Hooke énonce les bases de la théorie de l'élasticité linéaire (réversibilité et proportionnalité des déformations aux efforts), qui rend compte des petites déformations de la plupart des corps solides. Elle est utilisée peu après par Edme Mariotte et Jean Bernoulli pour résoudre des problèmes de flexion de poutres.

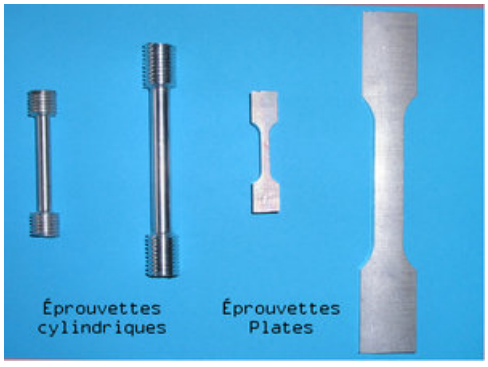
Après les travaux de Charles Augustin Coulomb, Henri Navier, Augustin-Louis Cauchy, entre autres, au milieu du 19ème siècle, la résistance des matériaux est devenue une discipline des sciences appliquées. Son développement rapide, dû aux travaux des ingénieurs du 20ème siècle, a conduit à l'élaboration de nombreuses méthodes de calcul analytique qui ont pu être érigées en règles ou normes à l'usage des bureaux d'étude. L'avènement des ordinateurs a rendu possible et a facilité l'exploitation de méthodes numériques générales qui permettent de résoudre les problèmes posés par les structures complexes. Les recherches sont, depuis les années 1970, orientées vers le développement de ces méthodes, vers l'étude des petites et grandes déformations permanentes des matériaux, des phénomènes de rupture, de la résistance aux environnements complexes (efforts évolutifs, hautes et basses températures) et vers l'utilisation de matériaux nouveaux (superalliages, polymères, matériaux composites, céramiques) [LEMAITRE J., 2001]. L'un des essais les plus importants pour l'expérimentation est l'essai de traction.

II-2 RAPPEL THEORIQUE

L'essai de traction est l'essai le plus classique. En théorie il consiste à exercer sur une éprouvette des forces croissantes (effort de traction) dans la direction des génératrices qui vont la déformer progressivement puis la rompre [REDJEL B., BENMEDAKHENE S., 1991]

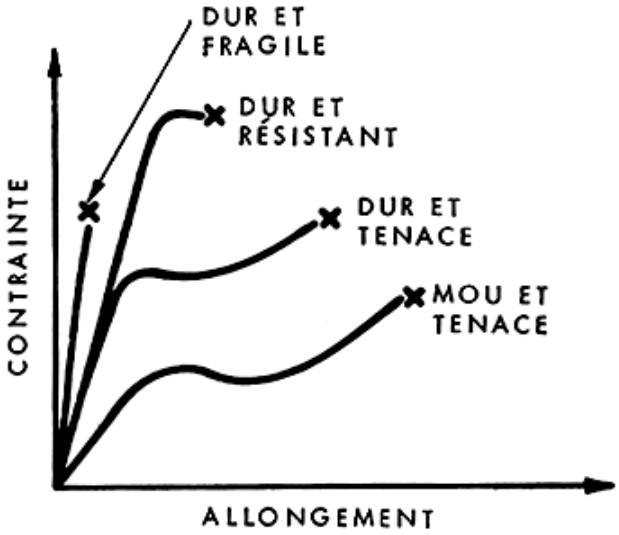
La traction s'exerce lentement et cause l'allongement de l'éprouvette, à un instant donné, une région de la partie médiane se ternit ; L'éprouvette s'étrangle et le matériau semble s'écouler comme un liquide très visqueux. C'est un phénomène d'instabilité plastique dit : Striction.

L'éprouvette est constituée par un barreau cylindrique rectifié, de section rigoureusement constante (voir figureII.2) Les dimensions sont normalisées. En cas d'impossibilité, il est nécessaire d'utiliser une éprouvette géométriquement semblable, afin que les résultats soient comparables.



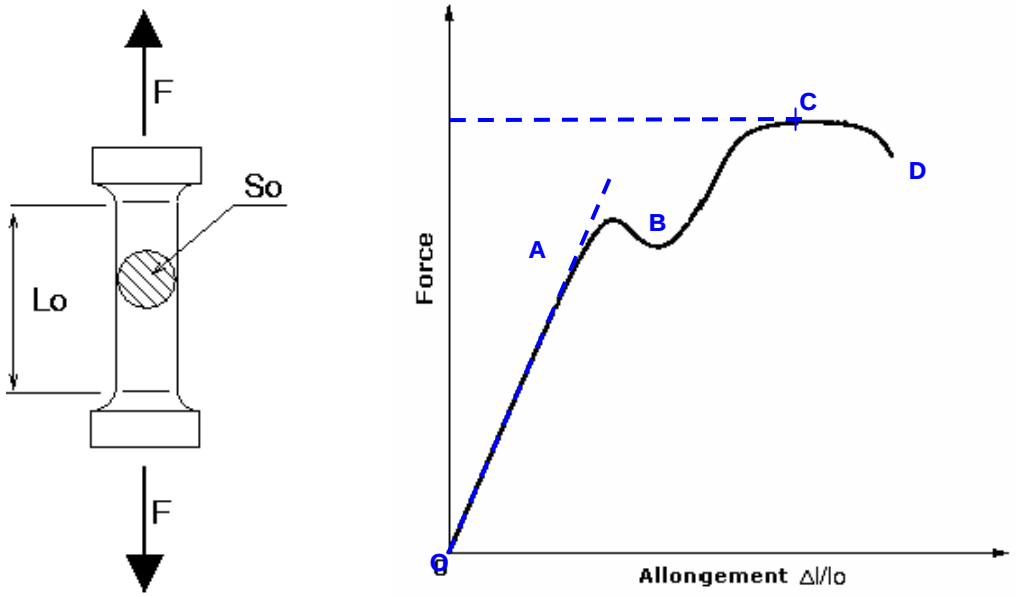
FigureII.2. Eprouvettes d'essai de traction

L'essai dure quelques minutes, la vitesse de mise en charge ne doit pas dépasser 1 daN/mm² par seconde pendant la période de déformation élastique, ensuite elle peut être doublée et même triplée. Suivant la nature du matériau [KERGUIGNAS M., CAIGNAERT G., 1977], le diagramme affecte l'aspect de l'un des diagrammes de la figureII.3.



FigureII.3. Courbes contrainte-allongement.

Examinons le diagramme enregistré au cours d'un essai de traction sur une éprouvette cylindrique en un matériau ductile ; la courbe **OABCD** montre comment varie l'allongement en fonction de la force (charge) :



FigureII.4. Courbe conventionnelle de traction

La partie **OA** est une droite peu inclinée sur l'axe O*X*. Elle correspond à une période de déformation élastique, les allongements très petits, sont directement proportionnels aux charges.

En effet, si, au cours d'essai, on cesse de tirer sur l'éprouvette, celle-ci reprend sa forme initiale. La contrainte de traction σ (prononcée sigma) dans le matériau de l'éprouvette est donnée très simplement par la relation :

$$\sigma = \frac{\text{Charge } F}{\text{Section de l'éprouvette } S} \qquad \textbf{(II. 1)}$$

Au point **A** correspond la fin de la proportionnalité de l'allongement à la force ; Pratiquement, ce point marque aussi la fin de l'élasticité.

La limite d'élasticité R_e est, par définition, le quotient de la charge correspond au point **A**, par la section de l'éprouvette.

La partie **ABC** est une courbe dont la forme varie suivant la température, la vitesse, le temps d'étirage pour un même matériau avec les mêmes dimensions. Elle correspond à une période plastique de grands allongements qui croissent plus vite que les charges. Si au cours de cette période, l'effort de traction est supprimé, l'éprouvette demeure plus ou moins allongée.

Le point **C** marque le début de la striction : l'éprouvette, dont la section diminue, continue à s'allonger alors que la charge décroît. La striction s'accentue de plus en plus et la rupture a lieu ; elle correspond au point **D** du diagramme. L'abscisse du point **C** définit la charge maximale que supporte l'éprouvette au moment où la striction s'amorce. Nous en déduisons la résistance à la rupture du matériau, celle-ci est égale au quotient de la charge maximale par la section initiale de l'éprouvette.

Après la rupture, rapprochons les deux parties de l'éprouvette rompue et mesurons l'allongement total, par convention, cet allongement est supposé uniformément réparti sur la longueur L de l'éprouvette : Le confinement de la déformation au centre de l'éprouvette pendant la striction prouve le contraire.

Les courbes de traction permettent de définir un certain nombre de caractéristiques mécaniques du matériau : *Limite élastique, module d'Young*, *résistance à la traction, allongement à la rupture.*

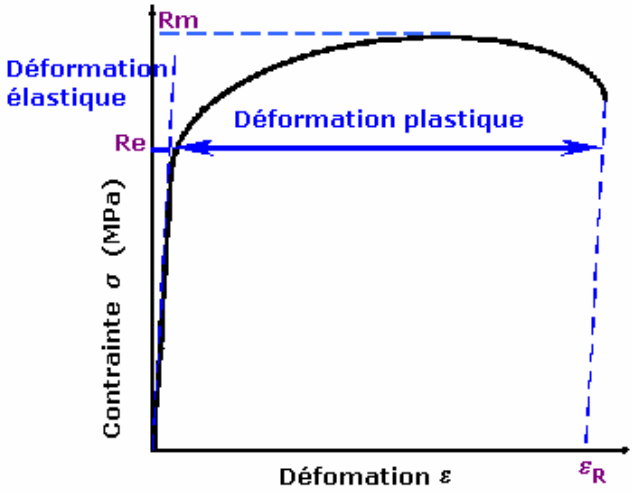
Afin de pouvoir utiliser les courbes de traction brutes, on doit les modifier de façon que les résultats obtenus ne soient fonction que du matériau étudié et non de la géométrie de l'éprouvette. On définit ainsi :

La Contrainte nominale : $\sigma_N = \frac{F}{S_0}$, unité : Pa **(II. 2)**

La Déformation nominale : $\varepsilon_N = \frac{\Delta L}{L_0}$ **(II. 3)**

S_0 et L_0 sont les caractéristiques initiales de l'éprouvette, ΔL est son allongement.

On trace alors la courbe : Contrainte en fonction de la déformation (figureII.5).



FigureII.5. Courbe de déformation nominale.

La courbe comporte deux parties :

- Le domaine de déformation élastique : L'éprouvette a un comportement élastique linéaire et suit la loi de HOOKE :

$$\sigma = E\,\varepsilon \qquad \textbf{(II. 4)}$$

E Représente le module d'élasticité ou module de Young qui est théoriquement la pente de la partie élastique de la courbe de traction $E = \sigma/\varepsilon$. En réalité, à cause de la déformabilité des machines de traction, le module d'Young n'est représenté par la pente de la partie linéaire qu'à un coefficient près. Aussi le mesure-t-on par d'autres méthodes en utilisant, par exemple, les vibrations ultrasonores de petites éprouvettes. Le module d'Young sert à caractériser la rigidité des matériaux.

- Le domaine de déformation plastique :

Dans ce domaine, la déformation est irréversible : Si l'on supprime la charge le matériau garde une déformation permanente. La courbe de traction passe par un maximum au-delà duquel la contrainte diminue bien que l'allongement continue de croître. La déformation

n'est alors plus homogène, elle est localisée dans la zone de striction (zone la plus mince de l'éprouvette)

II-3 CRITIQUE DES ESSAIS CONVENTIONNELS

A partir des courbes de déformation nominales, $\sigma_N = f(\varepsilon_N)$, on ne peut déduire que des informations très partielles sur le comportement plastique du matériau. De plus, comme nous allons le voir, ces courbes ne permettent pas, en général, de déduire la loi de comportement plastique du matériau aux grandes déformations. Essentiellement, on détermine quatre grandeurs caractéristiques (comme on vient de le citer précédemment) : Le module d'élasticité E, la limite d'élasticité R_e qui donne la valeur de la contrainte nominale théorique à partir de laquelle le matériau commence à se déformer plastiquement, la résistance à la traction R_m qui est définie par la contrainte nominale supportée par l'éprouvette et la déformation à la rupture ε_R qui correspond à la déformation plastique nominale à la rupture en traction de l'éprouvette. Comme la déformation plastique apparaît souvent progressivement, la limite d'élasticité est difficile à déterminer. On adopte alors une limite conventionnelle d'élasticité $R_{0.2}$ qui est la contrainte nominale correspondant à une déformation permanente de 0.2 %. Parmi ces caractéristiques, la limite d'élasticité R_e joue un rôle très important, car elle détermine la contrainte limite qu'il ne faut pas dépasser pour ne pas induire des déformations permanentes.

En général, les résultats des essais de traction sont exprimés en fonction des contraintes et déformations nominales. Idéalement, ce sont les contraintes et déformations réelles qui devraient être prises en compte. Mais, en pratique, leur détermination est rendue très difficile en raison du phénomène de striction. La contrainte réelle σ_r est donnée par le rapport entre la force F et la section réelle à tout instant de l'échantillon. Elle est parfois calculée à partir de la contrainte nominale moyennant des hypothèses simplificatrices telles que la conservation du volume de l'échantillon lors de l'étirage. La relation utilisée est alors :

$$\sigma_r = \sigma_N (1 + \varepsilon_N) \qquad \textbf{(II. 5)}$$

Quant à la déformation réelle ε_r, elle s'exprime en considérant l'accroissement infinitésimal de l'éprouvette de L à L + ΔL. on trouve que la déformation réelle ε_r est reliée à la déformation nominale [BAHADUR, 1973 ; PAMPILLO et DAVIS, 1972] :

$$\varepsilon_r = \ln (1 + \varepsilon_N) \qquad \textbf{(II. 6)}$$

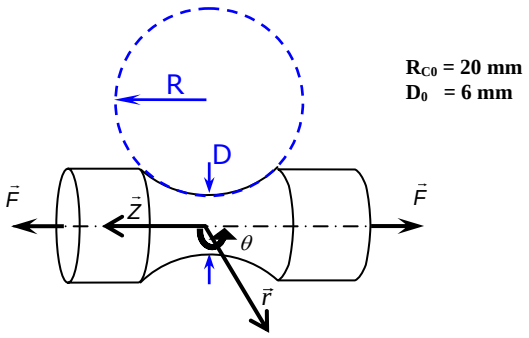
En réalité, ces calculs ne sont valables que dans les cas où la déformation est homogène c'est à dire pour une réponse viscoélastique (petites déformations) et dans le cas d'un comportement caoutchoutique. A partir de 0,1 environ, l'écart entre ε_r et ε_N devient de plus en plus important [KURZ et al., 1987]. De plus, on constate que le maximum de la contrainte nominale ne correspond pas à un maximum de la valeur réelle de la résistance intrinsèque du matériau. En réalité la résistance à la traction du matériau augmente jusqu'à sa rupture. Le maximum de la courbe de traction en contrainte nominale résulte seulement du phénomène de striction qui entraîne une diminution importante de la section de l'éprouvette [SOUAHI, 1986]. La force de traction nécessaire à la déformation de l'éprouvette devient alors plus faible en dépit de l'augmentation de la résistance du matériau, la résistance à la traction R_m est une mesure de la résistance de l'éprouvette et non pas celle du matériau. Le maximum de force observé ne correspond donc pas à un adoucissement réel du matériau mais simplement à la diminution rapide de la section de l'éprouvette [SOUAHI A., 1993]

Nous avons opté - afin d'effectuer notre simulation - pour des données prises des travaux faits sur la correction du phénomène de triaxialité des contraintes [SOUAHI A., 1986] lors d'une traction uniaxiale sur une éprouvette axisymétrique (en HDPE) en forme de sablier, en corrigeant la triaxialité par la méthode de BRIDGMAN.

II-4 CORRECTION DE BRIDGMAN

Eprouvette de traction :

Les éprouvettes sont axisymétriques en forme de sablier (figureII.6) ayant un rayon de courbure de 20 mm. Cette géométrie a été choisie suite à un ensemble de travaux antérieurs [G'SELL et JONAS, 1981] car elle impose une localisation de la déformation au centre de la partie utile et minimise les risques de cisaillement oblique en traction.



FigureII.6. Eprouvette axisymétrique en forme de sablier.

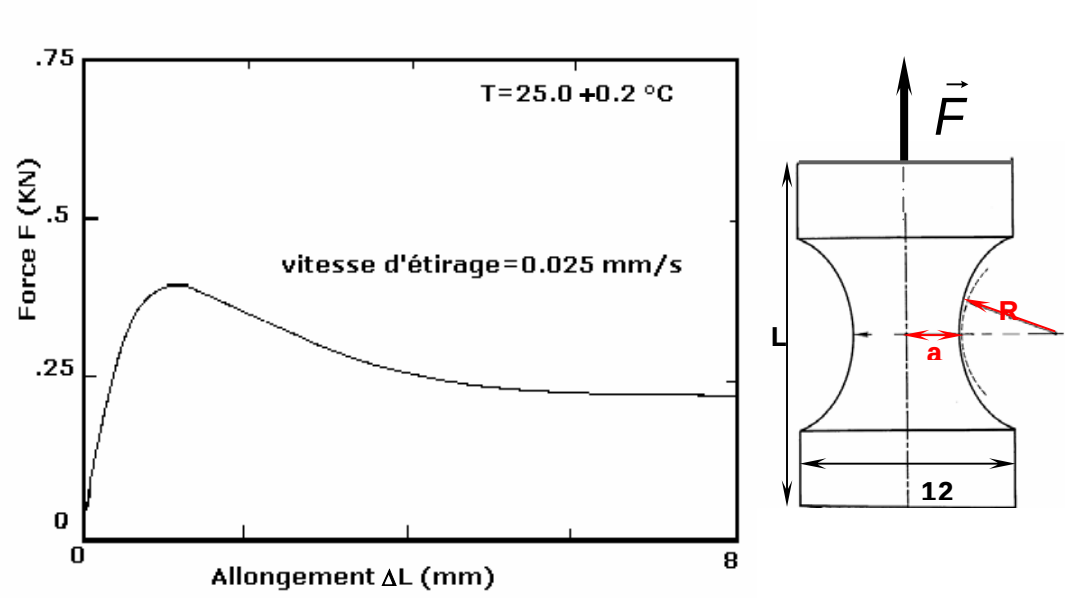
Le matériau est le polyéthylène produit par Union Carbide ltd, Canada, sous les références DFDY 6130, Natural 77.Selon les caractéristiques :

TableauII.1. Caractéristiques du HDPE utilisé

Caractéristiques	valeur	unité
Masse spécifique	0,956	g/cm^3
Indice de fluidité à l'état fondue	0,15	g/10mn
Masse moléculaire moyenne pondérale=M_w	192.200	g/mole
Masse moléculaire moyenne en nombre= M_n	14.500	g/mole
Taux de cristallinité en masse	70%	

L'étude quantitative du comportement mécanique d'un matériau en traction s'appui sur la détermination de la loi phénoménologique (analytique ou numérique) qui exprime les variations de la contrainte σ subie par le matériau en fonction de sa déformation ε et de sa vitesse de déformation $\dot{\varepsilon}$.

La traction uniaxiale sur une éprouvette -des quels nous avons tiré les données de simulation- donne la courbe dynamométrique force-allongement (voir figureII.7) qui commence par un stade viscoélastique initial à pente décroissante, un maximum de la force d'étirage pour un allongement voisin de 1mm, ensuite la force connaît une décroissance progressive.



FigureII.7. Courbe dynamométrique

En ce qui concerne le champ de déformation; Sa composante radiale a été déterminée par la mesure de la réduction locale du diamètre dans la section la plus faible de l'éprouvette :

$$\varepsilon_{rr} = \log\left(\frac{D}{D_0}\right) \qquad \textbf{(II. 7)}$$

Où D_0 est le diamètre minimum de l'éprouvette non déformée. Sous l'hypothèse que le matériau soit transversalement isotrope et se déforme à volume constant, la composante axiale de la déformation se déduit directement de la composante radiale par la relation :

$$\varepsilon_{zz} = -2\,\varepsilon_{rr} \qquad \textbf{(II. 8)}$$

En sachant que les effets de triaxialité de contrainte dans le plan de symétrie sont d'une amplitude de second ordre, leur influence sur la distribution locale des déformations est négligée. Alors la déformation effective locale s'exprime simplement par :

$$\varepsilon_{eff} = \varepsilon_{zz} = 2\,\log\left(\frac{D}{D0}\right) \qquad \textbf{(II. 9)}$$

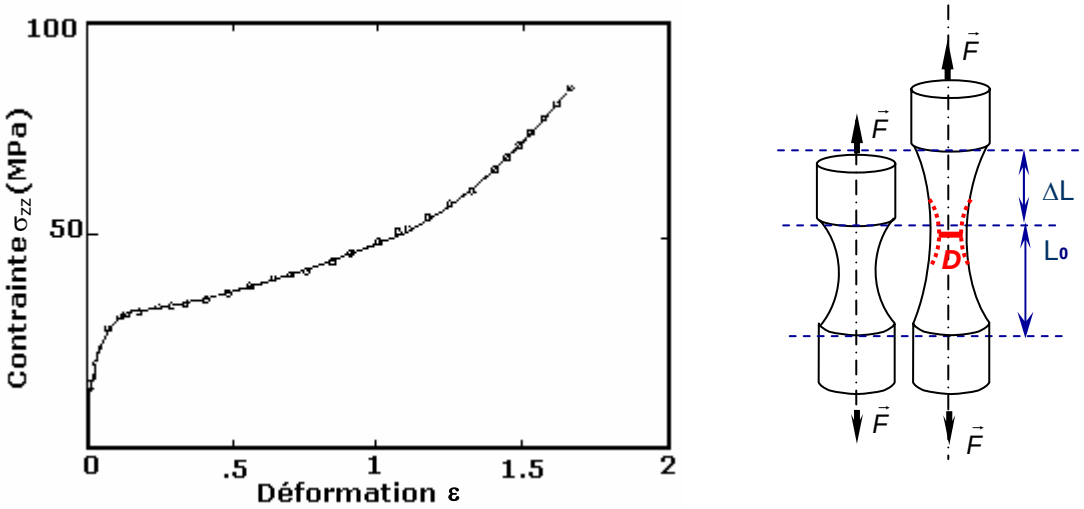
En conséquence, la vitesse de déformation effective s'écrit

$$\dot{\varepsilon}_{eff} = \frac{d\,\varepsilon_{eff}}{dt} = -\frac{2}{D}\frac{dD}{dt} \qquad \textbf{(II. 10)}$$

Dans une section droite d'une éprouvette de traction, la composante majoritaire du champ de contrainte est la composante axiale ou contrainte de KIRCHOFF. Elle s'exprime par l'équation :

$$\sigma_{zz} = \frac{4F}{\pi D^2} \qquad \textbf{(II. 11)}$$

D est le diamètre local de l'échantillon déformé (centre de la striction)



FigureII.8. $\sigma_{zz} = f(\varepsilon)$, au centre de la striction

On peut remarquer la limite élastique progressive à la fin du stade viscoélastique, suivie d'un long stade plastique caractérisé par un durcissement progressif. Aucun crochet de traction ne subsiste dans cette courbe, ce qui montre bien que le maximum de force observé auparavant, sur la figureII.8, ne correspondait pas à un adoucissement réel du matériau mais simplement à la diminution rapide de la section de l'éprouvette.

Différents travaux antérieurs ont montré que pour une éprouvette de section non uniforme (résultant de l'usinage initial ou de la formation d'une striction), les composantes de la contrainte radiale (σ_{rr}) et circonférentielle ($\sigma_{\theta\theta}$) ne sont plus négligeables [ALY HELAL, 1982 ; G'SELL et al., 1983 et 1985] Cet effet de triaxialité est d'autant plus sensible que la courbure locale du profil est plus marquée. Il s'exprime par le facteur de triaxialité, $F_T = \sigma_{eff} / \sigma_{zz}$, où σ_{eff} est la contrainte effective qui, en traction uniaxiale pure, produirait une vitesse de déformation égale à la vitesse de déformation effective $\dot{\varepsilon}_{eff}$ observée dans l'éprouvette réelle.

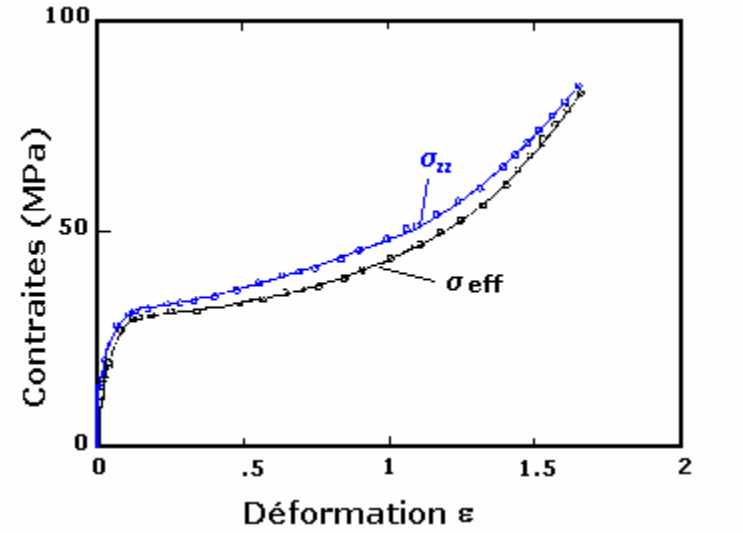
En se fondant sur une loi d'écoulement déduite du critère de VON MISES, BRIDGMAN [1944, Annexe1] a dérivé une expression du facteur de triaxialité :

$$F_T = \frac{1}{\left[\left(1 + 2Rc / a\right) \log \left(1 + a / 2Rc\right)\right]} \qquad \textbf{(II. 12)}$$

F_T dépend principalement de la géométrie du profil au niveau de la section considéré, a est le rayon de l'éprouvette, Rc est celui de la courbure du profil au centre de la striction de diamètre courant D. (voire la figureII.8) L'analyse de BRIDGMAN est la plus couramment utilisée. Sa validité opérationnelle a été confirmée dans un travail précédent [G'SELL et al., 1985] et malgré les approximations qu'elle suppose, elle était retenue pour cette étude.

Ainsi, il faut soumettre la contrainte axiale σ_{zz} à une correction de triaxialité pour tenir compte de l'effet des composantes transversales positives du tenseur des contraintes sur la déformation plastique du matériau.

A partir des valeurs de F_T, la contrainte σ_{zz} est corrigée en contrainte effective $\overline{\sigma_{eff}}$ dont les variations avec la déformation sont représentées sur la figureII.9.



FigureII.9. Variation de la contrainte effective σ_{eff} en fonction de la déformation vraie au centre de la striction [SOUAHI A., 1986]

En effet, le facteur de triaxialité corrige la contrainte en abaissant le durcissement apparent dans la zone intermédiaire autour de $\varepsilon \sim 0.7$.

CONCLUSION

Les propriétés mécaniques des polymères sont d'un intérêt considérable puisqu'elles conditionnent non seulement tous les problèmes de leur mise en œuvre en amont et de leur mise en forme en aval mais aussi leur comportement en service dans des applications extrêmement diversifiées.

Pour un problème déterminé, le choix d'un matériau polymère dépendra de ces propriétés mécaniques telles que la résistance, la dureté ou la ductilité, et il est donc nécessaire de les mesurer avec un certain nombre d'expériences que l'on appelle les essais mécaniques.

L'essai de traction est, depuis bien des années, le procédé expérimental le plus largement utilisé dans l'étude du comportement mécanique. Il permet l'étude détaillée des propriétés mécaniques de base, et il est à l'origine de la plupart des connaissances actuelles. En outre, son interprétation est simple lorsqu'il est complété par des observations micrographiques, optiques et électroniques, qui renseignent sur les modes de déformation mis en jeu.

Il est à noter, en outre, que les courbes conventionnelles de traction ne tiennent pas compte, pour l'évaluation de la contrainte, de la diminution de section de l'éprouvette au cours de la traction.

- Il est donc intéressant de déterminer à chaque instant la contrainte vraie $\sigma_v = F/S$ correspondant à une déformation vraie $\varepsilon_v = (\Delta L)/L$.

Où S et L sont la section et la longueur de l'éprouvette, à l'instant considéré.

On montre que :

$$\sigma_v = \sigma\left(1 + \frac{\Delta l}{l_0}\right),$$
$$\varepsilon_v = \lg\left(1 + \frac{\Delta l}{l_0}\right). \qquad \textbf{(II. 13)}$$

- Ou bien d'effectuer une correction, dans le cas précédent et vu la géométrie des éprouvettes, une correction de BRIDGMAN de la triaxialité a été réalisée, cette correction donne des résultats très satisfaisants : De ce fait on a choisi, comme données de simulation, les valeurs de la contrainte en fonction de la déformation pour déférentes vitesses de déformation lors d'un étirage uniaxiale d'un HDPE, corrigées par cette méthode.

REFERENCES

ALY HELAL
Etude et simulation du développement de la striction dans le polyéthylène à haute densité ;
(1982) Thèse de docteur – ingénieur, I.N.P.L., Nancy.

BAHADUR
Strain hardening equation and the prediction of tensile strength of ;
(1973) Rolled polymers, Polym. Eng. Sci, 13, 266.

BASQUIN R., LEMASSON G.
Cours de mécanique : Résistance des matériaux ;
(1964) Librairie DELAGRAVE, PP. 38.

BAUWENS J.-C.
Déformation plastique des hauts polymères vitreux soumis à un système de contraintes quelconque ;
(1967) J. Polym. Sci. A-2, 5, 1145.

BAUWENS J.-C.
Plastic deformation : experimental data and constitutive equations, Plastique deformation of amorphous and semi-cristalline materials, Ed. by ESACAIG B. and G'SELL C. ;
(1982) Les Editions de Physique, les Ulis, 175.

DAVIS L.A. et PAMPILLO C. A.
Kinetics of deformation of PTFE at high pressure;
(1972) J. Appl. Phys., 43 (11), 4285.

G'SELL C., ALY HELAL N. A. et JONAS J.J.
Effect of stress triaxiality on neck propagation during the tensile stretching of solid polymers;
(1983) J. Mater. Sci., 18, 1731.

G'SELL C et GOPEZ J. R.
Plastic banding in glassy polycarbonate under plane simple shear;
(1985) J. Mater. Sci., 20, 3462.

KURZ W., MERCIER J. P. et ZAMELLI G.
Introduction à la science des matériaux ;
(1987) Presse polytechniques romandes, Lausanne.

KERGUIGNAS M., CAIGNAERT G.
Résistance des matériaux ;
(1977) DUNOD UNIVERSITE, pp. 156-157.

LEMAITRE J.
Petite histoire de l'expérimentation en mécanique des solides ;
(2001) Meccanica, pp. 13-35.

REDJEL B., BENMEDAKHENE S.
Pratique des essais destructifs : l'essai de traction ;

(1991) Technique et science, 8, pp. 40-45.

SOUAHI A. 1
Effet de la triaxialité des contraintes pendant la striction diffuse en traction dans le polyéthylène à haute densité ;
(1986) Thèse de Magister, pp. 56-87.

SOUAHI A. 2
Influence de la réticulation sur le comportement mécanique des polymères amorphes au voisinage de la température de transition vitreuse ;
(1993) Thèse de docteur, pp. 40-46.

WARD I. M., RABINOWITZ S. et PARRY J. S. C
The effect of hydrostatic pressure on the shear yield behaviour of polymers;
(1970) J. Mater. Sci., 5, 29.

INTRODUCTION:

Comprendre les mécanismes à l'origine des fonctions supérieures du cerveau est l'objet de recherches au carrefour de la neurobiologie, de la psychologie, de l'informatique et de la physique.

Néanmoins, le terme de *neurones* ne doit pas faire illusion. Les succès récents des réseaux de neurones formels ne doivent rien à la biologie, mais sont dûs exclusivement à une meilleure compréhension des propriétés mathématiques fondamentales de ces réseaux. On est bien éloigné de la réalisation de machines susceptibles de reproduire une partie, même infime, des capacités de calcul des systèmes nerveux les plus simples. A l'heure actuelle, les réseaux de neurones formels doivent être considérés comme des extensions puissantes de techniques statistiques *classiques*.

III.1 BREF APERÇU HISTORIQUE

Les prémices

1890 : W. James, célèbre psychologue américain introduit le concept de mémoire associative, et propose ce qui deviendra une loi de fonctionnement pour l'apprentissage sur les réseaux de neurones connue plus tard sous le nom de loi de HEBB.

Dans la période de 1940 à 1956, les sciences et techniques de la cognition s'élaborèrent peu à peu et de l'ancienne cybernétique naquit les sciences cognitives. La volonté de fonder une science *naturelle* de l'esprit émergea quasiment au même moment à cette époque, en Europe et aux Etats-Unis. Par les efforts conjugués de chercheurs d'origines très diverses (Neurosciences, Intelligence artificielle, Linguistique, Psychologie et Epistémologie principalement), un nouveau champ d'investigation pluridisciplinaire, que sont les sciences cognitives, vit donc le jour.

L'apparition des réseaux de neurones artificiels s'inscrit complètement dans cette genèse des sciences cognitives, et leur origine tient dans la volonté de modéliser, de façon mathématique, les neurones biologiques.

Les pionniers

Les premiers à proposer un modèle, sont deux bio-physiciens de Chicago, Mc Culloch et Pitts, qui en 1943 inventent le premier neurone formel qui portera leurs noms : Neurone de Mc Culloch - Pitts ou automate à seuil.

Quelques années plus tard, en 1949, c'est HEBB qui proposera une formulation du mécanisme d'apprentissage, sous la forme d'une règle de modification des connexions synaptiques (règle de HEBB) Cette règle, basée sur des données biologiques, modélise le fait que si des neurones, de part et d'autre d'une synapse, soient activés de façon synchrone et répétée, la force de la connexion synaptique va aller croissant.

Le fait remarquable est qu'il ne s'agit à l'époque que d'une simple *intuition*, que les techniques expérimentales ne permettent pas encore de corroborer, le phénomène de modification synaptique reste d'ailleurs encore aujourd'hui très mal connu. L'intuition de Donald HEBB, en ajoutant la dimension dynamique au schéma statique dressé par Mc Culloch et Pitts, a constitué *l'étincelle de vie* qui a permis aux réseaux de neurones artificiels de connaître alors un véritable essor. [HEBB D.O., 1949]

Les réseaux de neurones artificiels sont *nés,* il y a une cinquantaine d'années, des efforts combinés de scientifiques issus de divers horizons et aux motivations variées. Leur histoire est galonnée d'un certain nombre de publications, livres ou articles d'intérêt essentiellement historique, associées aux étapes décisives de leur développement.

Le premier RNA

En 1958 ROSENBLATT, combinant les idées de ses prédécesseurs, conçoit le Perceptron : Un réseau de neurones artificiels inspiré du système visuel, possédant une couche de neurones *perceptive* et une couche de neurones *décisionnelle*. Ce réseau, qui parvient à apprendre à identifier des formes simples et à calculer certaines fonctions logiques, constitue le premier système artificiel exhibant une faculté jusque là réservé au vivant : La capacité d'apprendre par l'expérience; Le premier réseau de neurones artificiel proprement dit [ROSENBLATT R., 1962]

L'ombre

Malgré tout l'enthousiasme que soulève le travail de ROSENBLATT dans le début des années 60, la fin de cette décennie sera marquée en 1969, par une critique violente du Perceptron par MINSKY et PAPERT. Ils publient un livre qui, au terme d'une analyse mathématique approfondie, met à jour les limites intrinsèques du Perceptron. Ces conclusions plongent alors la recherche sur les réseaux de neurones artificiels dans une disgrâce qui ne prendra fin que 15 ans plus tard [MINSKY M. et PAPERT S., 1969]

Les RNA multicouches

Il faut attendre le début des années 80 pour voir un regain d'intérêt pour les réseaux de neurones artificiels. Celui-ci s'explique tout d'abord par les résultats des travaux de HOPFIELD qui démontre, en 1982, l'utilité des réseaux complètement connectés. Les réseaux récurrents, avec *feed-back*, qui constituent la deuxième grande classe de réseaux avec les réseaux de type Perceptron, aussi qualifiés de *feedforward*, dans la compréhension et la modélisation des processus de la mémoire et rend manifeste la relation existant, sur le plan formel, entre ce type de réseaux et des systèmes physiques (tels que les verres de spin) pour lesquels la physique statistique fournit un cadre théorique parfaitement approprié [HOPFIELD J.J., 1982]

Parallèlement aux travaux de HOPFIELD, WERBOS conçoit un mécanisme d'apprentissage pour les réseaux multicouches de type perceptron : C'est l'algorithme d'apprentissage par ***Back-propagation*** : Rétro-propagation de l'erreur (*RP*), qui fournit un moyen simple d'entraîner les neurones des couches cachées. Cet algorithme sera réellement popularisé en 1986 par RUMELHART et al. dans un article de Nature et un livre *Parallel Distributed Processing* qui a longtemps constitué la *bible* des connexionnistes. Cet algorithme, et ce livre, ont eu un impact considérable : Disposant d'un moyen simple d'entraîner les neurones cachés, les réseaux de type Perceptron munis d'une ou plusieurs couches cachées, (appelés MLP pour Multilayer Perceptron) qui, contrairement à leur célèbre ancêtre, ne souffrent d'aucune limitation théorique, ont pu être employés avec un succès grandissant pour résoudre toute une panoplie de problèmes complexes rencontrés dans de nombreux domaines à la fois scientifiques et techniques.

Depuis la fin des années 80, l'intérêt pour les réseaux de neurones artificiels ne s'est pas démenti, dans tous les milieux et sur tous les fronts. Du coté théorique, on a pu démontrer rigoureusement un résultat de grande importance affirmant que les réseaux MLP possédant seulement deux couches cachées sont capables d'approximer avec une précision arbitraire n'importe quelle fonction, c'est-à-dire de résoudre de façon optimale n'importe quel problème pouvant se ramener à un mapping d'un vecteur d'entrée vers un vecteur de sortie. En pratique, les réseaux de neurones ont dors et déjà donné lieu à de très nombreuses applications, notamment en physique des particules, où leur utilisation a été envisagée pour la première fois en 1987, par un physicien américain B. DENBY.

Années 90 : Autres modèles et algorithmes ad hoc ont été développé avec l'attention d'étendre les possibilités des réseaux statiques non-bouclés : Introduction progressive d'une mémoire interne et de la récurrence au sein du réseau.

Cette capacité, de modéliser tout ce qui a un sens *approximateur universel*, a été établie rigoureusement : *Théorème de Kolmogorov*. Les réseaux de neurones sont un outil rigoureux aux bases démontrées.

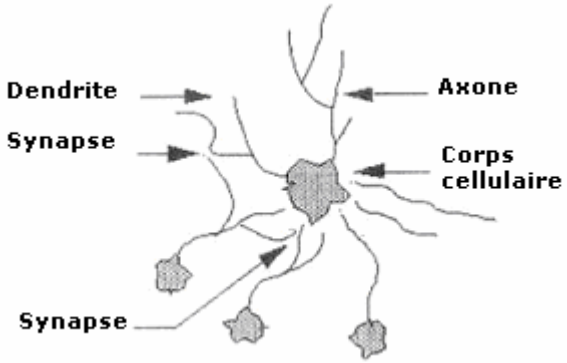
La raison pour laquelle cette méthode se popularise seulement maintenant est la puissance de calcul nécessaire pour la mettre en œuvre : Un apprentissage qui prend 3 mn sur un Pentium II en 1999 aurait demandé 320 h sur un IBM-PC XT en 1982 - 13 jours et 13 nuits.

III-2 LE NEURONE BIOLOGIQUE

Il est intéressant de présenter, succinctement et simplement, le fonctionnement des neurones biologiques, puisqu'ils ont servi de modèle pour les neurones formels.

Les neurones dans le cerveau, forment un ensemble de réseaux et chacun de ceux-ci a une tâche spécifique, dans le système nerveux, ils assurent la communication entre le corps et le cerveau.

Le moyen de communication des neurones fonctionne de façon binaire (tout ou rien) et s'appelle le potentiel d'action.



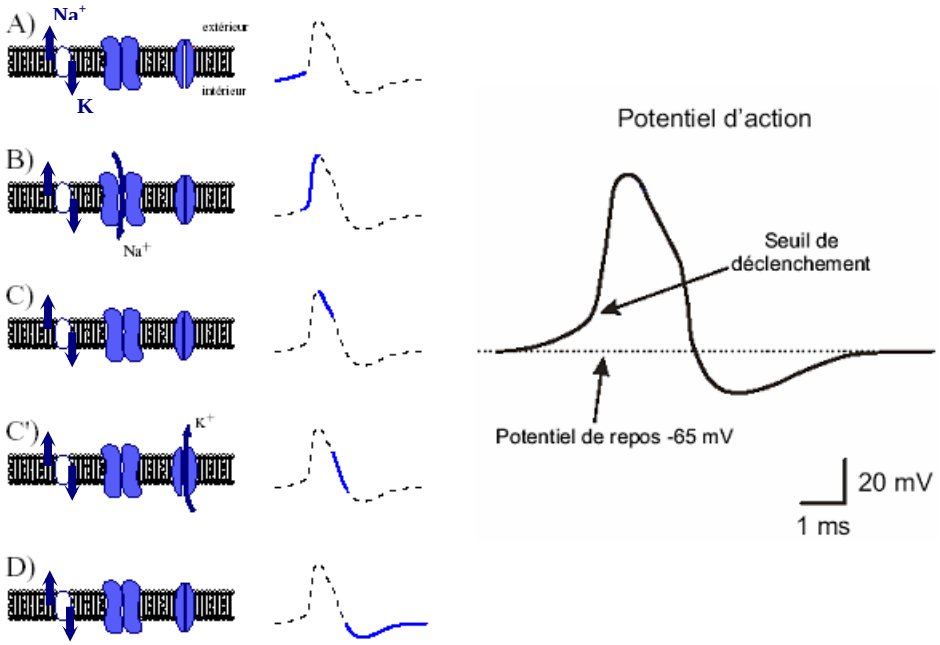
FigureIII.1. Le neurone biologique.

Un neurone est composé des dendrites connectées à des centaines et des centaines d'autres neurones, de corps cellulaire, faisant les sommes des impulsions et des absences d'impulsions qu'il reçoit, de l'axone qui est le prolongement du corps cellulaire, des terminaisons axonales qui sont les subdivisions de l'axone ressemblant un peu aux dendrites mais ayant pour différence de se terminer chacune par une synapse, et finalement des synapses qui envoient l'information du neurone à des centaines et des centaines d'autres neurones (figureIII.1).

III-2.1 Le potentiel d'action

Le potentiel d'action est toujours de nature tout ou rien. Au moment où il est déclenché par le corps cellulaire sur l'axone, il garde toujours son intensité de 100 mV. Elle ne diminue donc jamais, qu'importe la distance avec laquelle le potentiel d'action doit avoir affaire.

Les propriétés physiologiques des neurones ne reflètent pas la complexité de la réalité, seules les propriétés de base utilisées dans le cadre du connexionisme sont décrites, ces derniers sont exploitées dans le modèle formel proposé par McCulloch et Pitts, le neurone est considéré comme un automate à seuil, qui suivant une excitation et un certain seuil, répond ou non par un potentiel d'action.



FigureIII.2. Les mécanismes à l'origine du potentiel d'action [SAÏGHI S., 2004]

III-3 LE PASSAGE

« Pour comprendre les développements contemporains, nous devons d'abord nous mettre dans l'état d'esprit des gens qui, au 19ème siècle et dans les siècles précédents, réfléchissaient sur les processus de la pensée et du langage humain » - Simon H.A., Quelques remarques historiques sur la formation des sciences de la cognition -

Avant, on s'interrogeait comment un organe en chaire et en sang puisse produire de la pensée et avoir des idées ? C'est cette incompatibilité entre la nature de l'esprit (la pensée) et celle du corps (le cerveau) qui pose cette étrangeté. Aussi on voulait comprendre le fonctionnement du

cerveau afin de l'imiter et d'inventer un appareil qui simulera son raisonnement et sa capacité de résoudre les problèmes avec une certaine logique. Les travaux de Boole, VENN et JEVONS, entrevirent l'idée que l'action de penser pouvait être, d'une façon ou d'une autre, concrétisée et mise dans une forme telle que l'on pouvait commencer à en parler avec quelque précision, commencer donc à la caractériser. On peut penser que ce fut là le premier pas vraiment significatif pour répondre aux questions posées ultérieurement : Logique symbolique.

Plusieurs recherches sur le sujet ont été entreprises par plusieurs chercheurs tel que FREGE, RUSSELL, WHITEHEADMAIS qui travaillaient sur des systèmes de traitement de l'information. Certes, c'était que du crayon sur du papier mais on a là des systèmes. Enfin c'est l'article de Claude Shannon, publié en 1938 qui a montré comment des circuits électriques (circuits à relais) pouvaient être analysés en termes d'algèbre booléenne, pouvoir passer d'un dispositif physique à un ensemble d'équations mathématiques pousse à élucidé la nature de la logique, en d'autres termes la nature de la pensée. La question qui c'est posée, alors, était comment construire un appareil physique capable de se comporter selon les lois de la logique. Et est-ce que le cerveau obéît aux lois de la logique booléenne ?

C'est cette idée qui apparut en 1943 dans l'article de McCulloch et Pitts intitulé *A logical calculus of the ideas immanent in nervous activity*, ils décrivaient un cerveau hypothétique à l'aide de neurones considérés comme des mécanismes binaires, susceptibles d'être agencés les uns et les autres de façon à assurer des fonctions mentales d'une complexité définie arbitrairement. Ils proposaient de considérer cela soit comme une métaphore, soit comme une théorie [Le MOIGNE J.L., 2001]

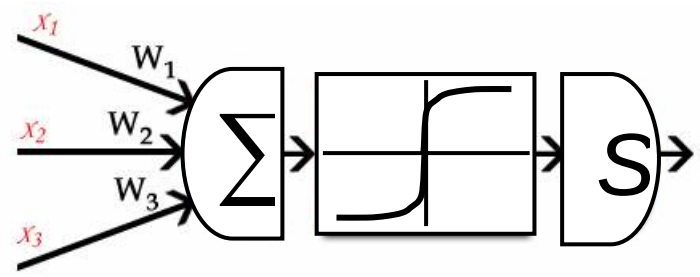
A l'époque, McCulloch étudie la strychnisation du cerveau des chimpanzés. Il se convint que les circuits d'influx nerveux étaient analogues à la propagation logique d'un raisonnement [DAMELINCOURT J., 2000]

Il en déduit la notion de psychon

- Evénement psychique.
- Tout ou rien
- Implique son antécédent temporel et intervient dans la détermination des psychons subséquents.

Avec l'aide du mathématicien Walter Pitts, il va présenter le neurone formel :

- Il est binaire : Il est soi actif soit inactif.
- Chaque neurone possède un seuil de déclenchement fixé.
- Le neurone reçoit des entrées de synapse excitatrices, ayant des poids identiques
- Le neurone peut aussi recevoir des entrées de synapses inhibitrices, dont l'action est absolue, c'est à dire que, si la synapse est active, le neurone ne peut pas être actif.
- Un quantum de temps est nécessaire à l'intégration des entrées synaptiques, basé sur le délai observé physiologiquement.



FigureIII.3. Le neurone formel de Mc Culloch et Pitts

Ils conçurent un calcule booléen des impulsions des réseaux neuronique, dont les éléments (les neurones idéaux) se déclenchent à l'instant t + 1, si la somme des poids des impulsions reçues en entrée à l'instant t dépasse une valeur seuil fixée (et constante dans le temps) [CORDESCHI R.]. De façon plus simple si aucune synapse inhibitrice n'est activée, le neurone ajoute ses entrées synaptiques et vérifie que la somme atteint ou excède son seuil est devient actif, sinon le neurone est inactif [DAMELINCOURT J., 2000].

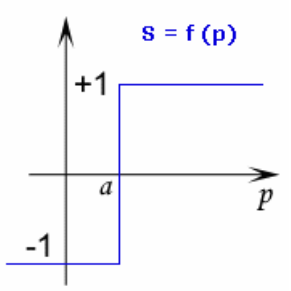
x_1, x_2, x_3, x_n sont les entrées ou encore les excitations (figureIII.3).

W_1, W_2, W_3, W_n sont appelés les coefficients synaptiques ou poids synaptiques.

Le potentiel p est donné par $p = x_1W_1 + x_2W_2 + x_3W_3 + ... + x_nW_n$

La fonction de sortie est s = f (p) où f est la fonction de transfert ou d'activation.

Le neurone crée par McCulloch et Pitts était un neurone utilisant une fonction binaire (figureIII.4).

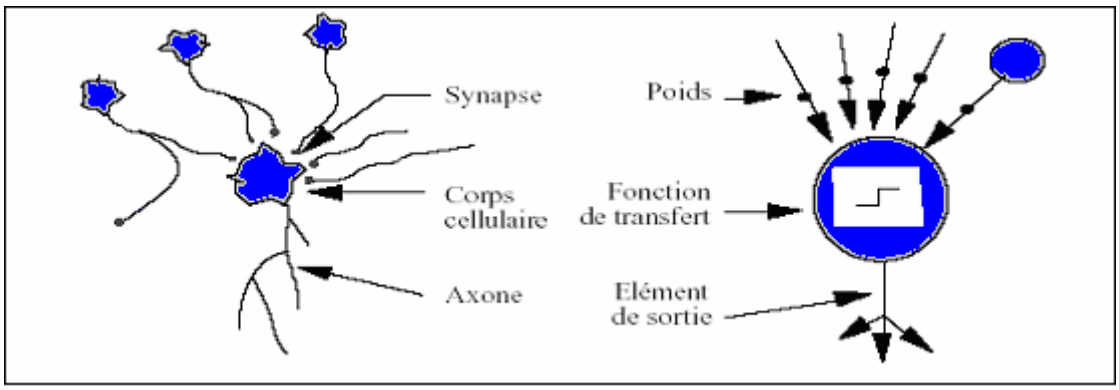


$$\begin{cases} si \quad p > a \quad \text{alors} \quad S = 1 \\ sinon \qquad S = -1 \end{cases}$$

FigureIII.4. Fonction de transfert binaire.

III-4 LE NEURONE FORMEL (ARTIFICIEL)

Qu'il soit simulé par logiciel ou réalisé sous une forme d'une puce électronique, le neurone est avant tout un automate. A partir des signaux d'entrée qu'il reçoit des automates qui lui sont reliés et de son état interne, il produit un signal de sortie qui devient son nouvel état interne [AL ANI T., 2004]



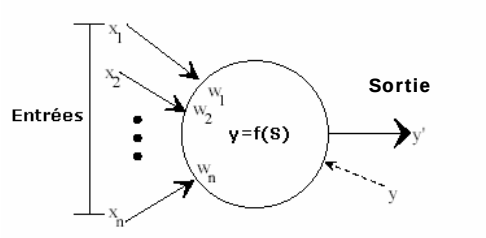
FigureIII.5. Mise en correspondance neurone biologique/neurone artificiel

Le modèle du neurone formel n'a pas pour rôle de représenter toutes les caractéristiques de neurone biologique :

- Le processus de transport, par l'axone et la modulation synaptique, est réduit à une simple multiplication de deux vecteurs XW
- La dynamique électrochimique complexe, à la source de la propagation du signal, est ramenée à sa plus simple expression par une fonction statique f.

Un neurone contient deux éléments principaux :

- Un ensemble de poids associés aux connexions du neurone,
- Une fonction d'activation.



FigureIII.6. Le neurone artificiel.

III-4.1 Fonctions d'activation

f est une fonction non-linéaire de la somme pondéré S des entrées appelée fonction d'activation (fonction de transfert) :

$$s_i = a + b = \sum_{j=1}^{N} \omega_j x_j + b = WX + b \qquad \textbf{(III. 1)}$$

$$y_i = f(s_i) \qquad \textbf{(III. 2)}$$

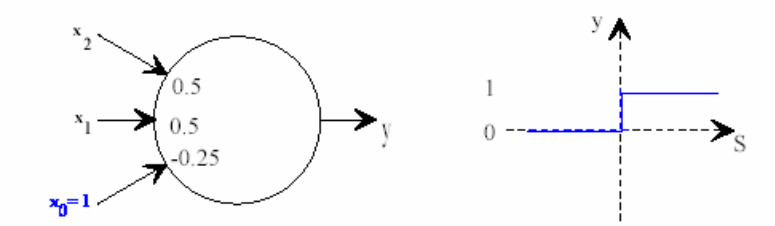
Où :

x_i : Les entrées au neurone qui proviennent de l'environnement externe au réseau ou d'autres neurones. X : Vecteur d'entrée.

ω_{ij} : Poids synaptiques associés à chaque connexion qui peut varier au cours du temps pour certains modèles. W : Vecteur poids.

b : Le biais du neurone= ± β (seuil du signal d'inhibition de neurone)

Le bais sert à déplacer le pas de la fonction d'activation sur l'axe S. La valeur de cette entrée est toujours 1 et le déplacement dépend alors seulement du poids de cette entrée spéciale. Déjà avec un neurone utilisant la fonction d'activation à seuil et une entrée de biais (x_0) il est possible de calculer les fonctions logiques ET et OU (figureIII.7)



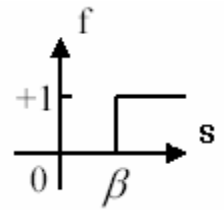
FigureIII.7. Fonction logique OU et le neurone correspondant**.**

a : Activation du neurone i tel que :

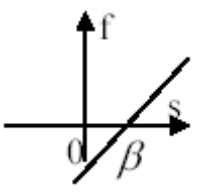
$$a_i(x) = \sum_{j=1}^{N} \omega_j x_j \qquad \textbf{(III. 3)}$$

La somme S devient l'argument de la fonction d'activation, qui est le plus souvent d'une des formes présentées ci-dessous :

f peut grossièrement être représentée par une fonction de Heaviside :

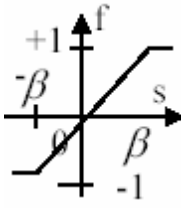


Ou une fonction linéaire sans saturation :

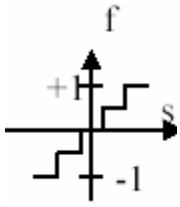


Ou une fonction seuil du type $y_i = f(a_i+b_i)$, $b_i = -\beta_i$:

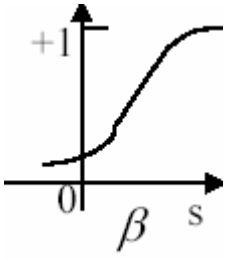
$y_i = 1$ si $a_i \geq \beta_i$, le neurone de l'indice i est activé
$y_i = 0$ si $a_i < \beta_i$, le neurone de l'indice i est inactivé



Fonction à seuil multiple :

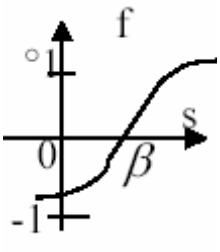


Fonction sigmoïde : $f(s)=\frac{e^{-\beta s}-1}{e^{\beta s}+1}$ **(III. 4)**



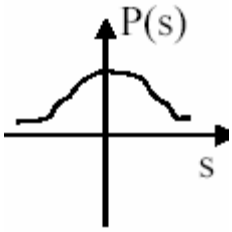
Il est plus commun d'employer une fonction non linéaire bornée et dérivable sur $]-\infty, +\infty[$ plutôt qu'une fonction linéaire par morceaux. Comme La fonction sigmoïde :

$$f(s) = \frac{1}{1 + e^{-\beta s}} \quad \textbf{(III. 5)}$$



Ou la fonction f (s) = tanh (s), qui est une fonction couramment utilisée.

Fonction stochastique : $P(s)=\frac{1}{1+e^{-\frac{s}{T}}}$ **(III. 6)**



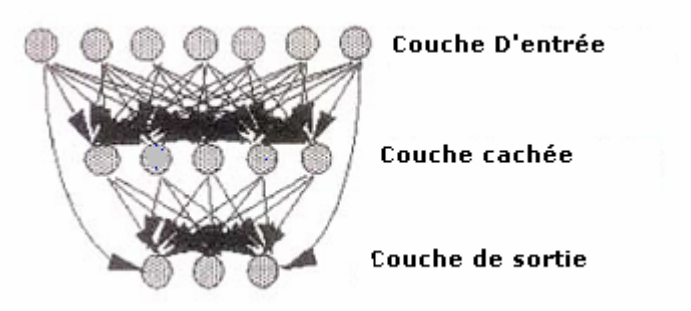
Le choix de la fonction d'activation dépend de l'application. S'il faut avoir des sorties binaires c'est la première fonction que l'on choisit habituellement.

Il n'y a pas de dynamique interne au sein de ce modèle. Pour des poids synaptiques fixes, le neurone formel accomplit une transformation statique. Dans un sens, ce modèle ne capture que les propriétés spatiales du neurone biologique et passe outre les relations temporelles.

La capacité représentative du neurone formel est fortement restreinte en raison de la forme rudimentaire de la fonction f utilisée.

III-5 RESEAUX DE NEURONES ARTIFICIELS (RNA)

Un réseau de neurones est un outil d'analyse statistique permettant de construire un modèle de comportement à partir de données qui sont des exemples de ce comportement. Il est constitué d'un graphe pondéré orienté dont les nœuds symbolisent les neurones [BALMISSE G., 2002]



FigureIII.8. Réseau de neurones artificiels

La capacité de traitement de ce réseau est stockée sous forme de poids d'interconnexions obtenus par un processus d'adaptation ou d'apprentissage à partir d'un ensemble d'exemples d'apprentissage. Rumelhart donne huit composants principaux d'un réseau de neurones [Rumelhart et al., 1986] :

- Un ensemble de neurones
- Un état d'activation pour chaque neurone (actif, inactif)
- Une fonction de sortie pour chaque neurone (f (S))
- Un modèle de connectivité entre les neurones (chaque neurone est connecté à tous les autres, par exemple)
- Une règle de propagation pour propager les valeurs d'entrée à travers le réseau vers les sorties
- Une règle d'activation pour combiner les entrées d'un neurone (très souvent une somme pondérée)
- Une règle d'apprentissage
- Un environnement d'opération (le système d'exploitation, par exemple)

Le comportement d'un réseau et les possibilités d'application, dépendent complètement de ces huit facteurs, le changement d'un seul d'entre eux peut changer le comportement du réseau complètement.

III-5.1 Topologie de connexion des neurones

La topologie du réseau est la façon dont les neurones sont reliés, elle se décline de deux manières : Dans la première chaque neurone communique avec tous les autres, et même avec lui-même (réseau totalement connecté) Dans la seconde les neurones sont organisés en couches, chaque couche reçoit alors des signaux de la couche précédente et en envoie à la couche suivante, mais les neurones appartenant à la même couche ne communiquent pas entre eux. Lorsque les signaux ne se propagent qu'en avant, on parle de réseau *feedforward,* et lorsqu'ils peuvent revenir en arrière on parle de réseau *feedback.*

On distingue deux grands types d'architectures de réseaux de neurones : Les réseaux de neurones non bouclés et les réseaux de neurones bouclés [DREYFUS G., 1998]

Dans ce qui suit nous nous parlerons que des réseaux pour lesquels nous avons opté pour notre application.

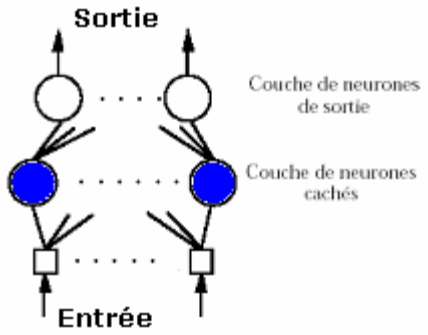
III-5.1.1 Réseaux statiques non bouclés [BISHOP C., 1995]

Un réseau de neurones non bouclé réalise une (ou plusieurs) fonctions algébriques de ses entrées, par composition des fonctions réalisées par chacun de ses neurones.

Il est représenté graphiquement par un ensemble de neurones *connectés* entre eux, l'information circulant des entrées vers les sorties sans *retour en arrière.* Si l'on représente le réseau comme un graphe dont les nœuds sont les neurones et les arêtes les *connexions* entre ceux-ci, le graphe d'un réseau non bouclé est acyclique. Le terme de *connexions* est une métaphore : Dans la très grande majorité des applications, les réseaux de neurones sont des formules algébriques dont les valeurs numériques sont calculées par des programmes d'ordinateurs, non des objets physiques (circuits électroniques spécialisés), néanmoins, le terme de connexion, issu des origines biologiques des réseaux de neurones, est passé dans l'usage, car il est commode quoique trompeur. Il a même donné naissance au terme de connexionisme.

La figureIII.9 représente un réseau de neurones non bouclé qui a une structure particulière, très fréquemment utilisée : Il comprend des entrées, une couche de neurones *cachés* et des neurones

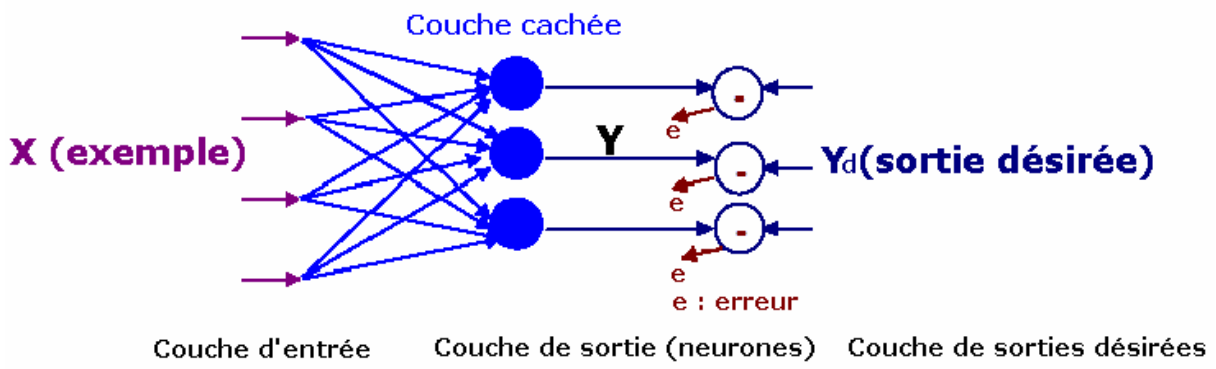
de sortie. Les neurones de la couche cachée ne sont pas connectés entre eux. Cette structure est appelée Perceptron multicouche.



FigureIII.9. Un perceptron multicouche

Les réseaux de neurones non bouclés sont des objets statiques : Si les entrées sont indépendantes du temps, les sorties le sont également. Ils sont utilisés principalement pour effectuer des tâches d'approximation de fonction non linéaire, de classification ou de modélisation de processus statiques non linéaires.

III-5.1.2 Réseaux à une couche cachée



FigureIII.10. Réseau à une couche cachée

Le perceptron à une couche cachée (Single Layer Feedforward Networks, figureIII.10) est une forme réduite du perceptron multicouches (MLP, multilayer perceptron), il est caractérisé par :

- Entrée et sortie vectorielles
- Une fois l'entrée X (exemple d'apprentissage) est présentée aux réseaux, les activations de neurones sont calculées une fois les réseaux sont stabilisés.
- Les erreurs ($e = Y - Y_d$) sont calculées.
- Les erreurs (e) sont minimisées par un algorithme d'apprentissage.

Parmi les nombreux modèles non linéaires, neuronaux ou non, qui peuvent être utilisés pour estimer une relation non linéaire, le perceptron à une couche cachée présente des caractéristiques intéressantes, que l'on peut résumer en quelques mots. C'est un modèle « *boite noire* » non linéaire, à capacité d'*approximation universelle flexible* et *parcimonieuse*.

1. Approximation universelle

La propriété d'approximation universelle des réseaux de neurones peut s'énoncer comme suit : Pour toute fonction déterministe suffisamment régulière, il existe au moins un réseau de neurones non bouclé, possédant une couche de neurones cachés et un neurone de sortie linéaire, qui réalise une approximation de cette fonction et de ses dérivées successives, au sens des moindres carrés, avec une précision arbitraire. Ce théorème ne s'applique que si les fonctions d'activation des neurones cachés possèdent certaines propriétés [CYBENCO G., 1989]; les fonctions sigmoïdes et les fonctions à base radiale remplissent les conditions d'application du théorème.

2. Parcimonie

Il va de soi que la propriété d'approximation universelle n'est pas spécifique aux réseaux de neurones : Les polynômes, les séries de Fourier, les fonctions Splines, possèdent cette même particularité. Ce qui différencie les réseaux de neurones des autres approximateurs universels usuels, c'est leur parcimonie : Pour obtenir une approximation d'une précision donnée, les réseaux de neurones utilisent moins de paramètres que les approximateurs usuels c'est-à-dire impliquant des paramètres ajustables à estimer, nécessitant moins de paramètres que les développements en série de fonctions fixes pour atteindre un même degré d'approximation [BARRON A. R., 1993]. En particulier, le nombre de paramètres varie essentiellement de manière linéaire en fonction du nombre de variables de la fonction que l'on cherche à approcher, alors qu'il varie beaucoup plus rapidement (voire exponentiellement) avec la dimension de l'espace des entrées dans le cas des approximateurs usuels.

Cet avantage est particulièrement intéressant du point de vue du temps de calcul, et surtout du point de vue de la quantité d'information nécessaire pour le calcul des coefficients :

- Du point de vue du temps de calcul, l'estimation des coefficients du réseau (l'apprentissage), est d'autant plus rapide que le nombre de paramètres à calculer est petit ;

- Lorsque la fonction que l'on cherche à approcher n'est pas connue analytiquement, mais seulement par l'intermédiaire de l'ensemble d'apprentissage, la taille de l'échantillon- donc le nombre d'exemples- nécessaire croît avec le nombre de poids. Le fait d'utiliser moins de coefficients que les méthodes classiques de régression permettent donc une économie en nombre d'exemples, ce qui peut être particulièrement important lorsque l'acquisition des exemples est coûteuse ou lente [HORNIK K. et al., 1994]

3. Flexibilité

Ces modèles sont flexibles : Plus la relation (non linéaire) à modéliser est complexe, plus nombreux seront les neurones et les paramètres dans le modèle neuronal correspondant. Cela signifie que leur complexité interne peut être facilement accrue, sans changer la forme globale du modèle. Ils appartiennent à la classe générale des modèles non-paramétriques qui ne font pas d'hypothèses sur la forme paramétrique de la fonction à modéliser. Ils constituent donc des outils de régression flexibles.

III-5.2 Apprentissage

L'apprentissage d'un réseau de neurones signifie qu'il change son comportement de façon à lui permettre de se rapprocher d'un objectif défini. Ce but est normalement l'approximation d'un ensemble d'exemples ou l'optimisation de l'état du réseau en fonction de ses poids pour atteindre l'optimum d'une fonction objective fixée a priori.

Il existe deux types d'apprentissages principaux. Ce sont l'apprentissage supervisé et l'apprentissage non-supervisé [Hecht-Nielsen, 1990].

On parle d'apprentissage supervisé quand le réseau est alimenté avec la bonne réponse pour les exemples d'entrées donnés. Le réseau a alors comme but d'approximer ces exemples aussi bien que possible et de développer à la fois la bonne représentation mathématique qui lui permet de généraliser ces exemples pour ensuite traiter des nouvelles situations qui n'étaient pas présentes dans les exemples.

Dans le cas de l'apprentissage non-supervisé le réseau décide lui-même quelles sont les bonnes sorties. Cette décision est guidée par un but interne au réseau qui exprime une configuration idéale à atteindre par rapport aux exemples introduits. Les cartes auto-organisatrices de KOHONEN sont un exemple de ce type de réseau [KOHONEN, 1984]

Il existe plusieurs règles d'apprentissage pour chaque type d'apprentissage. L'apprentissage supervisé est le type d'apprentissage le plus utilisé. Pour ce type d'apprentissage la règle la plus utilisée est celle de WIDROW-HOFF. D'autres règles d'apprentissage sont par exemple la règle de HEBB, la règle du perceptron, la règle de GROSSBERG etc. [FOGELMAN-SOULIE, 1988; RUMELHART et al., 1986; HECHT-NIELSEN, 1990]

Les méthodes d'apprentissage des poids, pour les MLP, sont très nombreuses et peuvent être présentées en trois classes. Dans la première, on trouve les méthodes exploitant l'architecture particulière, en succession de couches, de ces réseaux [PARISI R. et al., 1996]

La seconde classe comprend les différentes méthodes locales du premier ou du second ordre, basées sur le calcul du gradient [SHEPERD A. J., 1997]

Les méthodes d'optimisation globales, ou stochastiques, constituent la troisième classe, qui inclut en particulier les algorithmes évolutionnistes (algorithmes génétiques). La synthèse [Yao X, 1999] montre que de tels algorithmes sont utilisés non seulement pour l'estimation des paramètres, mais aussi pour la détermination de l'architecture et l'adaptation des règles d'apprentissage. A noter, dans cette classe, les techniques hybrides qui combinent méthode du gradient et algorithmes évolutionnistes.

Les performances des algorithmes d'apprentissage pour les MLP sont sensibles à toute une gamme de facteurs :

- Le critère d'erreur, qui peut être simplement quadratique, robuste aux valeurs aberrantes ou régularisées,
- Le principe d'initialisation des poids, qui influence considérablement le nombre d'itérations et peut avoir un impact sur la généralisation,
- Le critère d'arrêt de l'apprentissage,
- Des paramètres spécifiques aux différentes méthodes ou enfin la maîtrise par l'utilisateur d'une méthode particulière.

La meilleure méthode est donc dépendante du problème. L'algorithme de Levenberg - Marquardt, même s'il ne fournit aucune garantie d'atteindre le minimum global, est cependant largement recommandé.

III-5.2.1 Règle de WIDROW-HOFF

La règle d'apprentissage de WIDROW-HOFF est une règle qui permet d'ajuster les poids d'un réseau de neurones pour diminuer à chaque étape l'erreur commise par celui-là (à condition que le facteur d'apprentissage soit bien choisi)

Un poids est modifié en utilisant la formule suivante:

$$w_{k+1} = w_k + \alpha \, \Delta_k \, x_k \qquad \textbf{(III. 7)}$$

Où:

- w_k est le poids à l'instant k,
- w_{k+1} le poids à l'instant k+1,
- α est le facteur d'apprentissage (pas d'adaptation),
- Δ_k caractérise la différence entre la sortie attendue et la sortie effective du neurone à l'instant k.
- x_k la valeur de l'entrée avec laquelle le poids w est associé à l'instant k.

Ainsi, si Δ_k et x_k sont positifs tous les deux, alors le poids doit être augmenté. La grandeur du changement dépend avant tout de la grandeur de Δ_k mais aussi de celle de x_k. Le coefficient α sert à diminuer les changements pour éviter qu'ils deviennent trop grands, ce qui peut entraîner des oscillations du poids.

Deux versions améliorées de cet apprentissage existent, la version *par lots* et la version *par inertie* (momentum en anglais) [HECHT-NIELSEN, 1990], dont l'une utilise plusieurs exemples pour calculer la moyenne des changements requis avant de modifier le poids et l'autre empêche que le changement du poids au moment k ne devienne beaucoup plus grand qu'au moment k-1.

III-5.2.2 Algorithme de rétropropagation du gradient (Annexe2)

L'algorithme de rétro-propagation du gradient est utilisé dans le perceptron multicouche. Un réseau de type rétropropagation se compose d'une couche d'entrée, une couche de sortie et zéro ou plusieurs couches cachées [RUMELHART et al., 1989] Les connections sont permises seulement d'une couche inférieure (plus proche de la couche d'entrée) vers une couche supérieure (plus proche de la couche de sortie). Il est aussi interdit d'avoir des connections entre des neurones de la même couche.

La couche de sortie se compose normalement des neurones linéaires qui calculent seulement une somme pondérée de toutes ses entrées.

Les couches cachées contiennent des neurones avec des fonctions d'activation non-linéaires, normalement la fonction sigmoïde.

Il a été prouvé [HECHT- NIELSEN, 1990] qu'il existe toujours un réseau de neurones de ce type avec trois couches seulement (couche d'entrée, couche de sortie et une couche cachée) qui peut approximer une fonction $f : [0,1]^n \rightarrow R^m$ avec n'importe quelle précision $\varepsilon > 0$ désirée. Un problème consiste à trouver combien de neurones cachés sont nécessaires pour obtenir cette précision. Un autre problème est de s'assurer a priori qu'il est possible d'apprendre cette fonction.

Initialement tous les poids peuvent avoir des valeurs aléatoires, qui sont normalement très petites avant de commencer l'apprentissage. La procédure d'apprentissage se décompose en deux étapes. Pour commencer, les valeurs d'entrées sont fournies à la couche d'entrée. Le réseau propage ensuite les valeurs jusqu'à la couche de sortie et donne ainsi la réponse du réseau. A la deuxième étape les bonnes sorties correspondantes sont présentées aux neurones de la couche de sortie, qui calculent l'écart, modifient leurs poids et rétropropagent l'erreur jusqu'à la couche d'entrée pour permettre aux neurones cachés de modifier leurs poids de la même façon. Le principe de modification des poids est normalement l'apprentissage de WIDROW-HOFF.

Le fait que l'apprentissage utilise un principe de descente de gradient sur la surface d'erreur pour modifier les poids, veut malheureusement dire qu'il est possible de tomber dans des minima locaux, auquel cas le réseau n'apprendra jamais l'approximation optimale. En effet, la convergence de l'algorithme n'est pas prouvée et de multiples variables sont à ajuster précisément en fonction du problème traité.

III-5.2.3 Algorithme de Levenberg – Marquardt (Annexe2)

Dans la méthode de rétropropagation du gradient, pour retrouver les poids qui minimisent l'erreur commise, on utilise uniquement l'information donnée par la dérivée du premier ordre ; Or, il est connu en optimisation classique que l'utilisation des informations données par les dérivées d'ordre supérieur peut accélérer le processus de recherche de la solution optimale. De nombreux travaux ont tenté d'étendre cette approche aux réseaux de neurones. On peut ainsi citer la méthode de Newton qui utilise les dérivées de second ordre (la matrice HESSIENNE)

Cependant, il est très difficile et très coûteux de calculer la matrice HESSIENNE dans le cas des réseaux multicouches. L'algorithme de Levenberg-Marquardt [HAGAN M.T., MENHAJ M., 1994] fait partie des méthodes dites Quasi-Newton où le Hessien n'est pas calculé de manière explicite mais approché en utilisant une fonction dépendant du gradient.

III-5.3 Validation

Parmi les techniques de validation, on a :

III-5.3.1 Fonction d'autocorrélation des résidus

On définit une fonction d'autocorrélation des résidus :

$$\hat{r}_{\varepsilon\varepsilon}(\tau) = \frac{\sum_{k=1}^{N-\tau}\left(\varepsilon\left(k,\hat{\theta}\right)-\bar{\varepsilon}\right)\left(\varepsilon\left(k-\tau,\hat{\theta}\right)-\bar{\varepsilon}\right)}{\sum_{k=1}^{N}\left(\varepsilon\left(k,\hat{\theta}\right)-\bar{\varepsilon}\right)^2} \qquad \textbf{(III. 11)}$$

III-5.3.2 Fonction d'intercorrélation

La fonction d'intercorrélation enter les résidus et les entrées précédentes :

$$\hat{r}_{u\varepsilon}(\tau) = \frac{\sum_{k=1}^{N-\tau}\left(u(k)-\bar{u}\right)\left(\varepsilon\left(k-\tau,\hat{\theta}\right)-\bar{\varepsilon}\right)}{\sqrt{\sum_{k=1}^{N}\left(u(k)-\bar{u}\right)^2}\sqrt{\sum_{k=1}^{N}\left(\varepsilon\left(k,\hat{\theta}\right)-\bar{\varepsilon}\right)^2}} \qquad \textbf{(III. 12)}$$

où :

La moyenne : $\bar{x} = 1/N \sum_{k=1}^{N} x(k)$ **(III. 13)**

ε : L'erreur de prédiction

u : entrée de système

Idéalement, si le modèle est valable, les résultats de ces tests de corrélations conduisent aux résultats suivants :

$$\hat{r}_{\varepsilon\varepsilon}(\tau) = \begin{cases} 1, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases} \quad \text{et, } \hat{r}_{u\varepsilon} = 0, \forall \tau \qquad \textbf{(III. 14)}$$

III-6 CONCEPTION D'UN RESEAU DE NEURONES

Pour construire un réseau de neurones, la première chose à faire n'est pas de choisir le type de réseau mais de bien choisir les échantillons de données d'apprentissage, test et validation. Ce n'est qu'ensuite que le choix du type du réseau interviendra.

Voici chronologiquement les quatre grandes étapes qui doivent guider la création d'un RNA.

III-6.1 Choix des échantillons

Cette étape est cruciale et va aider le concepteur à déterminer le type du réseau le plus approprié pour résoudre son problème. La façon dont se présente l'échantillon conditionne :

- Le type du RNA
- Le nombre de cellules d'entrée et de sortie
- La façon dont il faudra mener l'apprentissage, les tests et la validation.

III-6.2 Elaboration de la structure du réseau

La structure du réseau dépend étroitement du type des échantillons. Il faut tout d'abord choisir le type de réseau : Un perceptron standard, un réseau de HOPFIELD, un réseau de KOHONEN etc. Dans le cas du perceptron par exemple, il faudra aussi choisir le nombre de neurones dans la couche cachée.

III-6.3 Apprentissage

L'apprentissage est l'avant dernière phase d'élaboration d'un réseau de neurones, elle consiste à calculer les pondérations optimales des différentes liaisons, en utilisant un échantillon. La méthode la plus utilisée est la rétropropagation : On entre des valeurs dans la cellule d'entrée, et en fonction de l'erreur obtenue en sortie, on corrige les poids accordés aux pondérations. C'est un cycle qui est répété jusqu'à ce que la courbe d'erreur du réseau ne soit croissante.

III-6.4 Validation

Une fois le réseau calculé, il faut procéder à des tests pour vérifier qu'il réagit bien comme on le souhaite : C'est la validation. La méthode la plus simple étant de garder une partie de

l'échantillon réservée à l'apprentissage, pour la validation et faire ainsi une validation hors échantillon.

III-7 APPLICATIONS

Les Applications des Réseaux de Neurones représentent une liste non exhaustive

- Reconnaissance de formes, de caractère et de la parole : Détection de visages dans des fonds complexes, le scanner.
- Informatique, Bases de données : Fouille de Web (Web Mining), Fouille de données (Data mining)
- Traitement du signal : Estimation en temps réel des courants harmoniques pour la commande d'un filtre actif.
- Recherche Pharmaceutique, Diagnostic en médecine, Génétique
- Contrôle des processus industriels : Modélisation et la conduite des procédés
- Guidage de véhicules – modélisation. Commande sans capteur d'une machine asynchrone ; Diagnostic des systèmes industriels
- Applications en physique : Guidage avec précision de la visée balistique ; Physique des particules
- Modélisation mathématique : Apprentissage d'une fonction non linéaire
- Domaine spatial
- Prédiction des séries temporelles : Marché-Finance

III-8 COMPARAISON AVEC D'AUTRE METHODES

Les réseaux de neurones sont des outils de régression non linéaire asymptotiquement plus parcimonieux que les B-Splines ou les polynômes en effet, ils ont la capacité de représenter n'importe quelle fonction, linéaire ou pas simple ou complexe. Ils sont ce qu'on appelle en mathématique des *approximateurs universels*. Ils ont également la faculté d'apprentissage à partir d'exemples représentatifs. L'apprentissage est automatique.

Les techniques statistiques classiques imposent des restrictions sur le nombre d'entrée. Aussi, les régressions se font-elles sur des fonctions de dépendance simple, linéaire ou logarithmique. Pour l'utilisation de ces méthodes statistiques, beaucoup d'hypothèses sont imposées par les théories sous-jacentes : Distributions normales, variables non corrélées, etc. En plus ces approches

échouent avec les grosses bases de données, car en présence de milliers de variables, on trouvera forcément un niveau élevé de redondance parmi ces variables.

Au contraire, pour les réseaux de neurones, les hypothèses imposées sont moins contraignantes. Aucun prétraitement, simplification, réduction de variables quantitatives ne sont nécessaires. En outre, il a été démontré que la régression linéaire et la régression logistique sont des cas particuliers de réseaux de neurones, avec une seule couche et un neurone de sortie linéaire. En fait, les réseaux englobent et offrent plus de généralité que les techniques statistiques.

CONCLUSION

Après un aperçu historique sur les réseaux de neurones, et le rappel de quelques théorèmes relatifs à leurs capacités d'approximation, nous avons présenté un algorithme d'apprentissage très utilisé dans la communauté connexionniste : L'algorithme de rétro-propagation (*RP*). Plusieurs travaux lui ont été consacrés et un certain nombre de modifications ont été proposées pour accélérer le temps d'apprentissage et améliorer les performances du réseau obtenu. Nous en avons rappelé dans ce chapitre les principales variantes.

Dans la pratique, l'objectif d'une modélisation statistique n'est pas d'ajuster finement un modèle sur un ensemble d'apprentissage, mais d'obtenir un bon compromis entre performances d'apprentissage et performances de généralisation.

Lors de la construction d'un réseau de neurones, l'objectif donc est de déterminer la meilleure architecture possible et le meilleur jeu de poids au sens d'un critère donné (la généralisation par exemple)

Sous cet angle, la construction du réseau de neurones peut être vue comme un problème de recherche d'une solution dans l'espace des architectures et des poids.

REFERENCES

AL ANI T.
Introduction aux Systèmes Adaptatifs à Base d'Apprentissage ;
(2004) Partie 3 : Réseaux de neurones artificiels, pp.12-33.

BALMISSE G.
Les réseaux de neurones ;
(2002) KM CENTER.

BARRON A.R.
Universal approximation bounds for superpositions of a sigmoïdal function;
(1993) IEEE Transactions on Information Theory, 39(3), 930-945.

BATOUCHE S.
Prédiction des indices de rétention d'un certain groupe d'acrylates en chromatographie en phase gazeuse ;
(2002) Mémoire de Magister (chimie-physique) –Univ. Annaba-

BEAR M.F., CONNORS B.W., PARADISO M.A.
Neurosciences – A la découverte du cerveau ;
(1997) Pradel, Paris, 654 p., ISBN 2–907516–92–2.

BISHOP C.
Neural Networks for Pattern Recognition;
(1995) Oxford University Press.

CORDESCHI R.
Naissance et frontières des neurosciences, Cerveau, esprit et machines : Précis historique de l'intelligence artificielle.

CYBENKO G.
Approximation by superpositions of a sigmoidal function;
(1989) Mathematics of Control, Signals, and Systems, 2 (4), pp. 303-314.

DAMELINCOURT J.
Le neurone formel ;
(2000) Vie Artificielle. com.

DOUENCE V.
Circuits et systèmes de modélisation analogique de neurones biologiques ;
(2000) Thèse de docteur, pp.13–22.

DREYFUS G.
Les réseaux de neurones ;
(1998) Mécanique Industrielle et Matériaux, 51.

FOGELMAN-SOULIE F.
Méthodes connexionnistes pour l'apprentissage ;
(1988) Actes des Journées Nationales sur l'Intelligence Artificielle. Paris: Teknea, pp. 275-293.

HAGAN M.T., MENHAJ M.
(1994) IEEE Transactions on Neural Networks, 5.

HEBB D.O.
The Organization of Behaviour;
(1949) Wiley, New York.

HECHT-NIELSEN R.
Neurocomputing;
(1990) Addison-Wesley Publishing Company, 433 p.

HOPFIELD J.J.
Neural Networks and Physical Systems with Emergent Collective Computational Abilities;
(1982) Proceedings of the National Academy of Sciences, USA, pp. 2-554.

HORNIK K., STINCHCOMBE M., WHITE H.
Multilayer Feedforward Networks are Universal Approximators ;
(1989) Neural Networks Vol. 2, pp. 359-366.

HORNIK K., STINCHCOMBE M., WHITE H.
Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks;
(1990) Neural Networks, 3, p. 551-560.

HORNIK K.
Approximation capabilities of multilayer feedforward networks;
(1991) Neural Networks, 4, p. 251-257.

HORNIK K., STINCHCOMBE M., WHITE H. et AUER P.
Degree of approximation results for feedforward networks approximating unknown mappings and their derivatives ;
(1994) Neural Computation Vol.6, pp. 1262-1275.

KANDEL E.R., SCHWARTZ J.H., JESSELL T.M.
Principles of neural science;
(1991) Troisième édition, Prentice Hall International, Londres, 1135 p., ISBN 0–8385–8068-8.

KOHONEN T.
Self-organization and associative memory;
(1984) Berlin: Springer-Verlag.

LE MOIGNE J.L.
En hommage à H.A. SIMON (1916-2001), texte d'une note de lecture rédigée en 1991, consacrée au livre de mémoires que H.A. SIMON avait publié en 1990, sous le titre symbolique : " Models of my life " : Quelques remarques historiques sur la formation des sciences de la cognition.

MINSKY M. et PAPERT S.
Perceptrons;
(1969) The MIT Press, Cambridge.

PARISI R., DICLAUDIO E., ORLANDI G. et RAO B.D.
A generalized learning paradigm exploiting the structure of feedforward neural networks;
(1996) IEEE Transactions on Neural Networks, 7 (6), 1450-1460.

RIVALS I., PERSONNAZ L., DREYFUS G.
Modélisation, classification et commande par réseaux de neurones : principes fondamentaux, méthodologie de conception et illustrations industrielles ;
(1995) Les réseaux de neurones pour la modélisation et la conduite des procédés, J.-P. Corriou coordonnateur, Lavoisier, Technique et Documentation, Paris.

ROSENBLATT R.
Principles of Neurodynamics*;*
(1962) Spartan Books, New York.

RUMELHART D.E., G. HINTON et R. WILLIAMS
Learning Representations by Back Propagating Errors;
(1986) Nature, 323, pp. 533-536.

RUMELHART D. E., Mc CLELLAND J. L. et al.
Parallel Distributed Processing;
(1988) Massachusetts: MIT Press 1, 547 p.

SHEPERD A.J.
Second-Order Methods for Neural Networks;
(1997) Springer, London.

SAÏGHI S.
Circuits et systèmes de modélisation analogique de réseaux de neurones biologiques : Application au développement d'outils pour les neurosciences computationnelles ;
(2004) Thèse de docteur, pp. 18–23.

YAO X.
Evolving Artificial Neural Networks;
(1999) Proceedings of the IEEE, 87 (9), pp. 1423-1447.

OBJECTIF

Nous voulons simuler une loi de comportement effective d'un polyéthylène haute densité en traction uniaxiale, par un réseau de neurones type feedforward ; Nous formons ce réseau par des valeurs, contraintes vraies - déformations vraies (Annexe3) corrigées par la méthode de BRIDGMAN (chapitre II), prises pour quatre vitesses de déformation : $\dot{\varepsilon}_1 = 1.\ 10^{-3}\ s^{-1}$, $\dot{\varepsilon}_2 = 5.\ 10^{-4}\ s^{-1}$, $\dot{\varepsilon}_3 = 5.\ 10^{-5}\ s^{-1}$, $\dot{\varepsilon}_4 = 1.\ 10^{-5}\ s^{-1}$. Nous essayons de construire un réseau capable de prédire cette loi de comportement pour une vitesse différente de celles utilisées pendant l'apprentissage. Nous notons ensuite sa réponse et nous la comparons avec l'expérimentation.

Nous scindons notre ensemble de données en deux sous-ensembles : Avec le premier nous effectuons l'apprentissage, le second pour la validation de la réponse du réseau (validation hors échantillon). Nous augmentons la taille de la matrice d'apprentissage si nécessaire, pour avoir une meilleure prédiction.

VI-1 CONSTRUCTION DU RESEAU

Au cours de ces dernières années, l'une des évolutions les plus remarquables des réseaux de neurones formels (RN) a été, pour les chercheurs, l'abandon de la métaphore biologique au profit de fondements théoriques solides : La propriété fondamentale des RN est l'approximation universelle parcimonieuse.

Nous définissons :

L'erreur moyenne de prédiction (Mean Square Error)

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2 \qquad \textbf{(VI.1)}$$

Où :

N : Nombre de données (échantillons),

$\hat{y}$: Sortie estimée (simulée),

y : Sortie réelle (désirée).

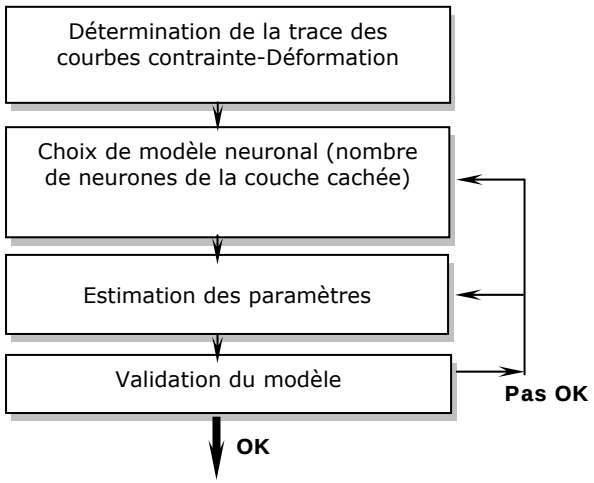
L'écart type de l'erreur de prédiction, noté RMSE (Root Mean Square Error) :

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \qquad \textbf{(VI.2)}$$

Lors de l'établissement des RNA plusieurs paramètres entrent en jeu comme les fonctions d'activation, les algorithmes d'apprentissage et d'entraînement, le nombre de couches dans le réseau, etc.

Le choix a été effectué suivant la littérature afin d'obtenir un réseau à même de donner un résultat optimal.

Seul le choix du nombre de neurones dans la couche cachée ne peut être fixe [HECHT-NIELSEN, 1990], nous devons donc le faire varier jusqu'à obtenir l'optimum dans la plage correspondante aux nombres de données entrées : L'écart type RMSE calculé selon l'équation (VI.2) doit être minimum (figureVI.1).



FigureVI.1. Diagramme de la procédure

VI-1.1 Choix de l'architecture

TableauVI.1. Paramètres choisis

Paramètres	Choix
Modèle utilisé	RP
Entrée	Vitesses et déformations
Fonction d'entraînement	Levenberg-Marquardt
Fonction d'apprentissage	Descente du gradient
Nombre de couches	2

Fonction d'activation au niveau de la 1ère couche	sigmoïde
Fonction d'activation au niveau de la 2ème couche	linéaire
Durée d'entraînement	10000 cycles
Performance	10^{-10}

C'est un perceptron monocouche (chapitreIII) type feedforward [HORNIK et al., 1989], utilisé pour la prédiction [COINTE A. et ROUGER F., 2004], avec un apprentissage supervisé (figureVI.2) et un algorithme d'apprentissage selon la méthode de Levenberg - Marquardt [LEVENBERG K., 1944 ; Marquardt D.W., 1963 ; RUMELHART et al., 1986; FOGELMAN-SOULIE, 1988; HECHT-NIELSEN, 1990 ; HAGAN M.T., MENHAJ M., 1994] qui consiste à :

1- Initialiser aléatoirement les poids du réseau

2- Présenter le premier exemple et calculer la sortie

3- Comparer la sortie avec la sortie réelle et calculer l'erreur

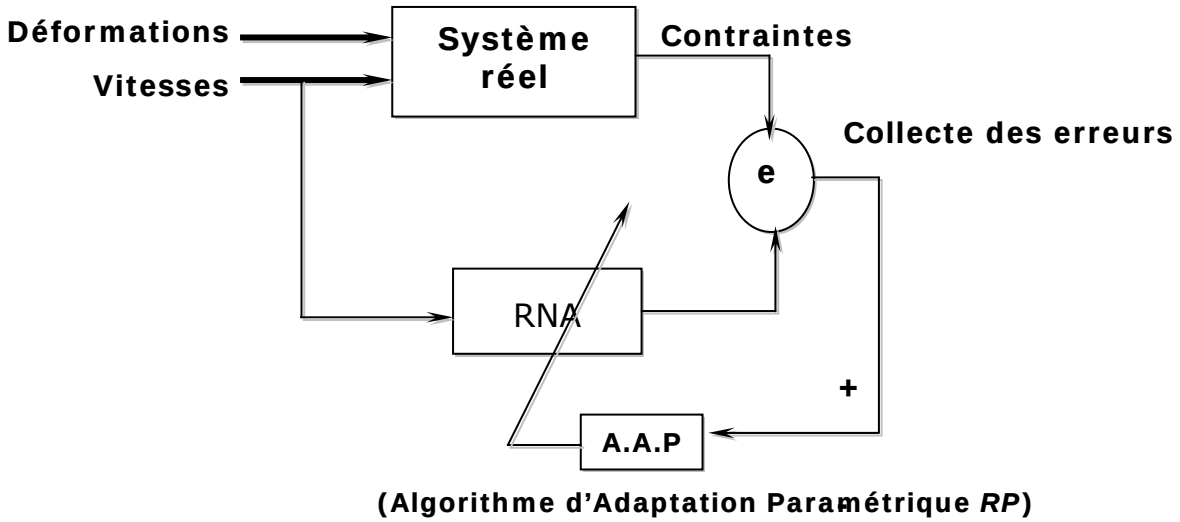
4- Rétropropager l'erreur et adapter les poids :

- Si l'erreur $_{nouvelle}$ > l'erreur $_{ancienne}$

  $\Rightarrow W_{nouveau} = W_{ancien} + \Delta W$, règle d'apprentissage (Annexe2)

  Où :

  W : vecteur des poids synaptiques (chapitreIII, Annexe2),
- Si l'erreur $_{nouvelle}$ < l'erreur $_{ancienne}$ $\Rightarrow$ Pas de modification de poids
- Test si la performance désirée est atteinte $\Rightarrow$ Arrêt
- Sinon arrêt avec le nombre d'itérations (Cycles, epochs.)

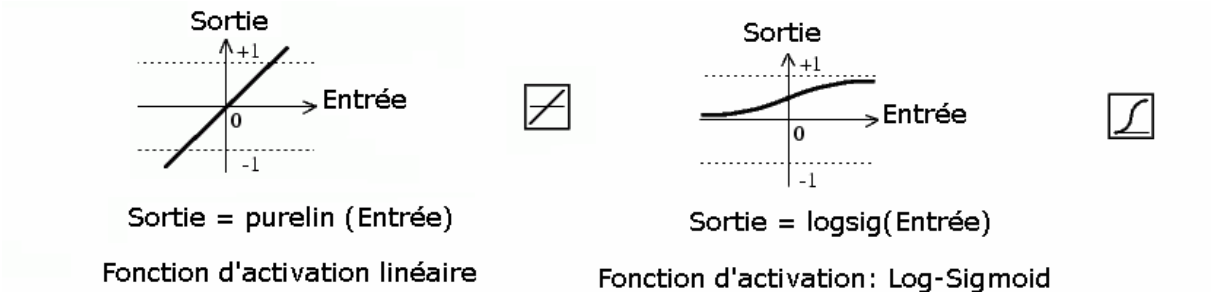


FigureVI.2. Schéma bloc de l'algorithme de prédiction

VI-1.2 Choix de la fonction d'activation

La fonction d'activation des neurones de la couche cachée est une sigmoïde (Chapitre III), celle de la sortie est linéaire (figureVI.3) ; Ce qui donne au réseau la propriété d'approximateur universelle [CYBENKO G., 1989]

La combinaison linéaire des sigmoïdes permet d'approximer n'importe quelle fonction d'intérêt avec une précision arbitraire, à condition de disposer de suffisamment de neurones sur la couche cachée [PARIZEAU M., 2004].



FigureVI.3. Type de fonctions d'activation utilisées

VI-1.3 Choix de nombre de neurones dans la couche cachée

VI-1.3.1 Premier paquet

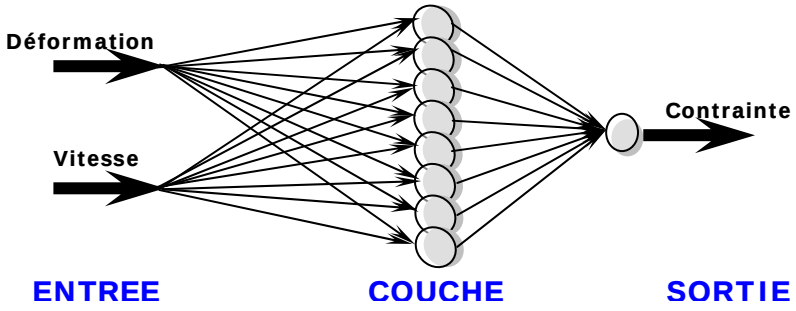
A cause de l'ordre des grandeurs non homogènes des valeurs : Contraintes (de l'ordre de 10MPa), déformations (de l'ordre de 1%) et des vitesses de déformation (de l'ordre de 10^{-4} s^{-1}), nous normalisons nos données.

Ce premier paquet est constitué des trois cent (300) premières valeurs prises de chaque courbe contrainte-déformation, pour effectuer l'apprentissage. Le reste sert pour la validation. C'est à dire moins d'un tiers des données pour la validation et plus de deux tiers pour l'apprentissage. Nous avons réalisé plusieurs essais, à raison de dix milles (10000) itérations pour chacun, afin de déterminer le nombre de neurones dans la couche cachée.

Le tableau suivant montre que les plus faibles erreurs d'apprentissage et de validation sont obtenue pour une architecture avec huit neurones (8) dans la couche cachée (figureVI.4).

TableauVI.2. Choix de nombre de neurones dans la couche cachée.

Nombre de neurones dans la couche cachée	RMSE due à l'apprentissage (10^{-4})	RMSE due à la validation (10^{-4})
5	49	822
6	9	139
7	23	1196
8	5	132
9	5	3255
10	7	1279
12	24	1003
14	10	12234



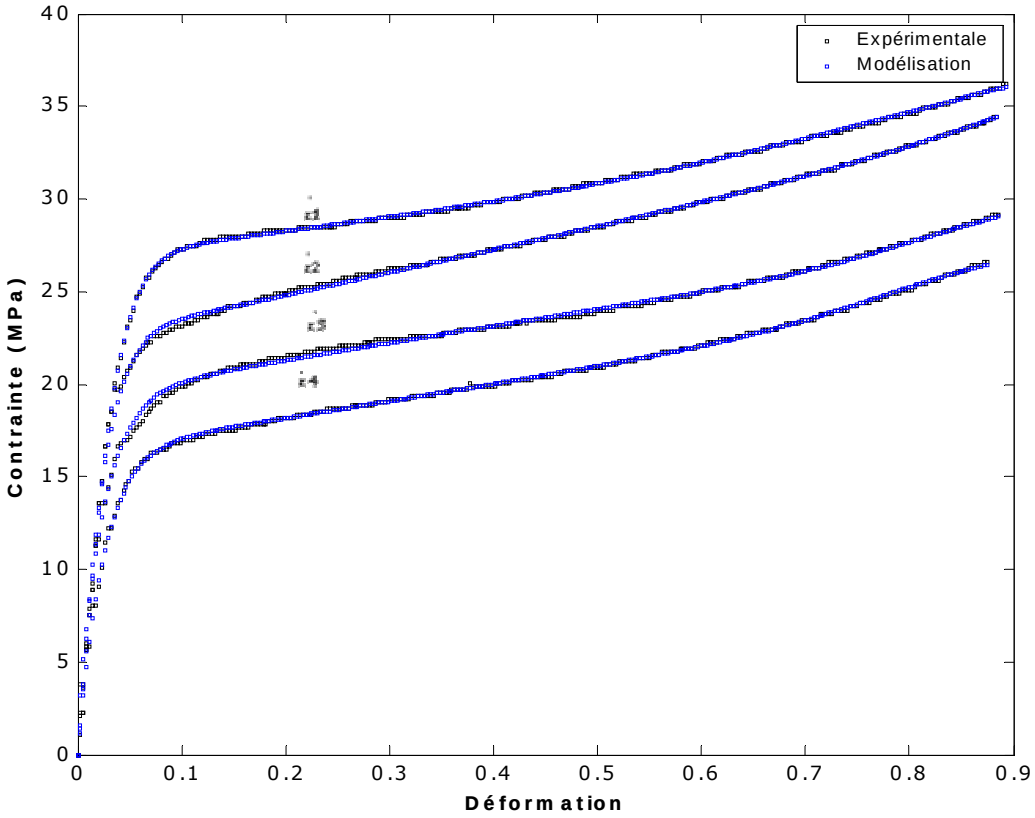
FigureVI.4. Architecture du réseau choisi

Les courbes obtenues après apprentissage sont identiques à celles expérimentales (figureVI.5), à l'exception de la deuxième et la troisième vitesse de déformation où le début du domaine plastique est mal appris par le réseau : Le passage du linéaire vers le non linéaire fait que les valeurs obtenues après simulation sont écartées de 5‰, ce qui est indiqué par la variation de l'erreur au cours de l'apprentissage (figureVI.6).

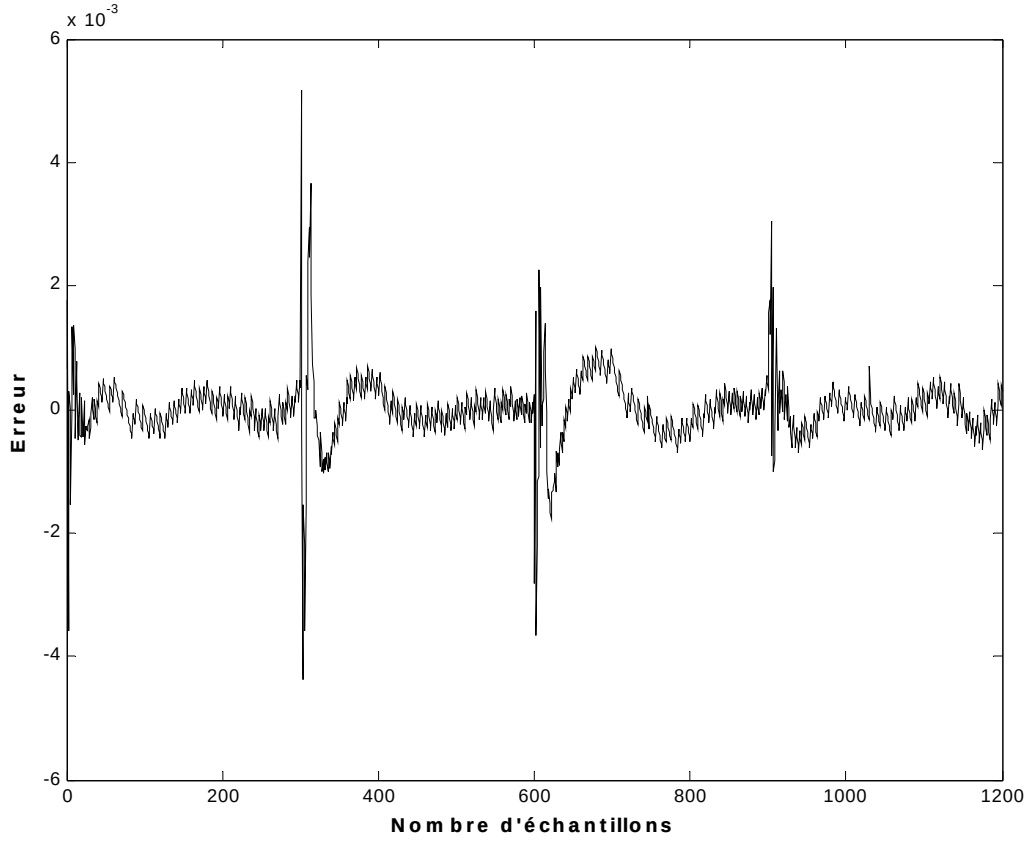
Le réseau valide mal pour la deuxième vitesse et très mal pour les vitesses restantes (figureVI.7). L'erreur due à la validation (figureVI.8) est plus marquée pour la première et la troisième vitesse de déformation : C'est parce qu'on passe, dans le domaine de validation, du comportement plastique au durcissement du matériau, c'est un autre type de comportement que le réseau n'a pas appris pendant la phase d'apprentissage alors il n'a pas pu le prédire correctement.

En ce qui concerne le test de généralisation (figureVI.9), pour une vitesse prise dans la gamme des vitesses d'apprentissage $\dot{\varepsilon}_T = 1.\ 10^{-4}\ s^{-1}$, la courbe obtenue a la même allure générale que celle expérimentale, sauf dans l'intervalle des données de validation, on constate l'augmentation

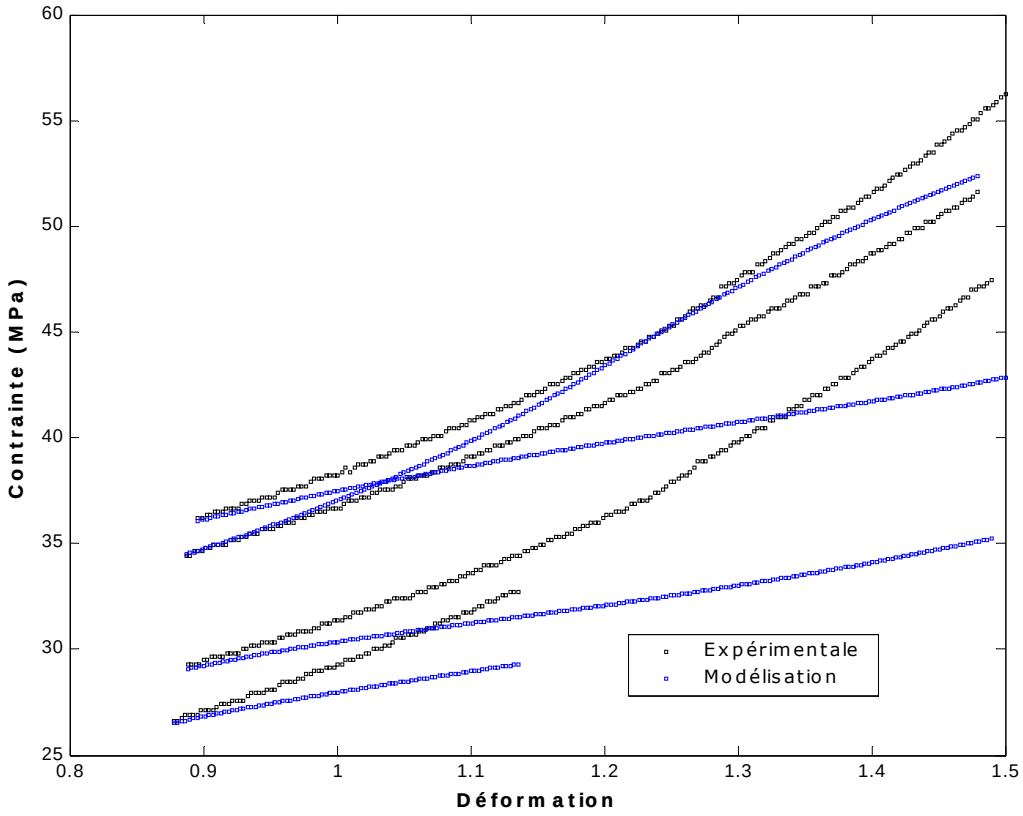
de l'erreur (figureVI.10) dans ce domaine où le réseau généralise mal le durcissement du matériau.



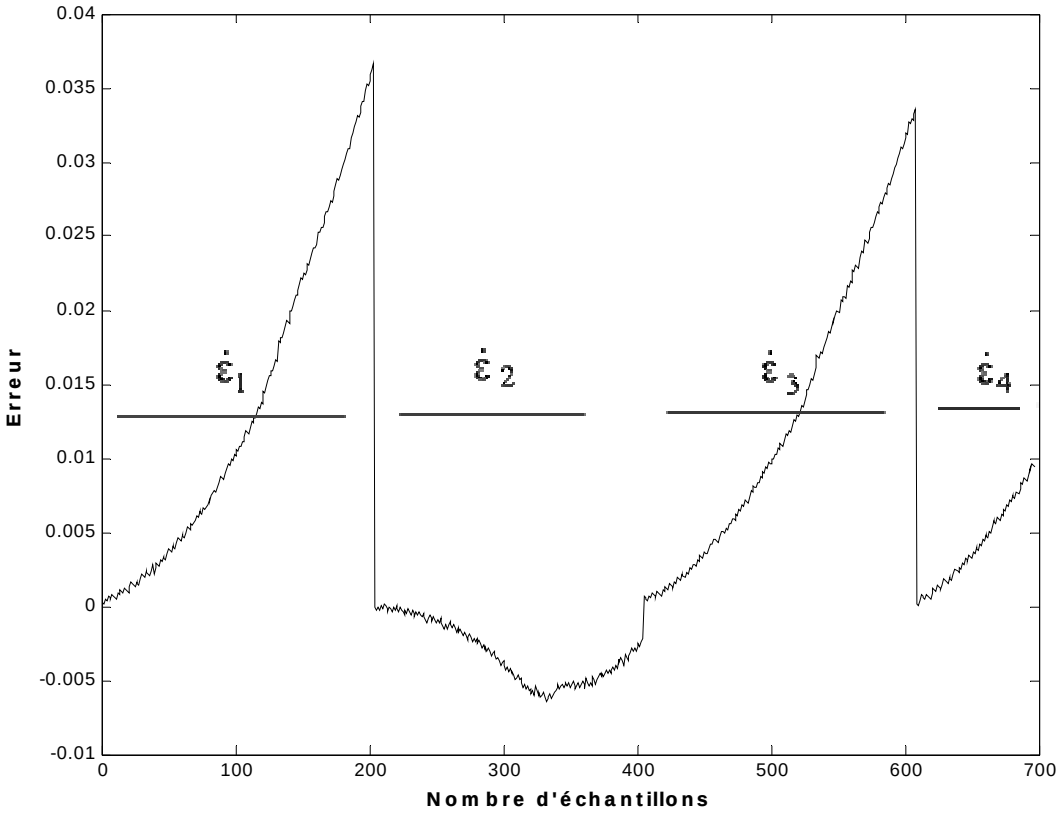
FigureVI.5. Comparaison entre les valeurs expérimentales et celles obtenues après apprentissage



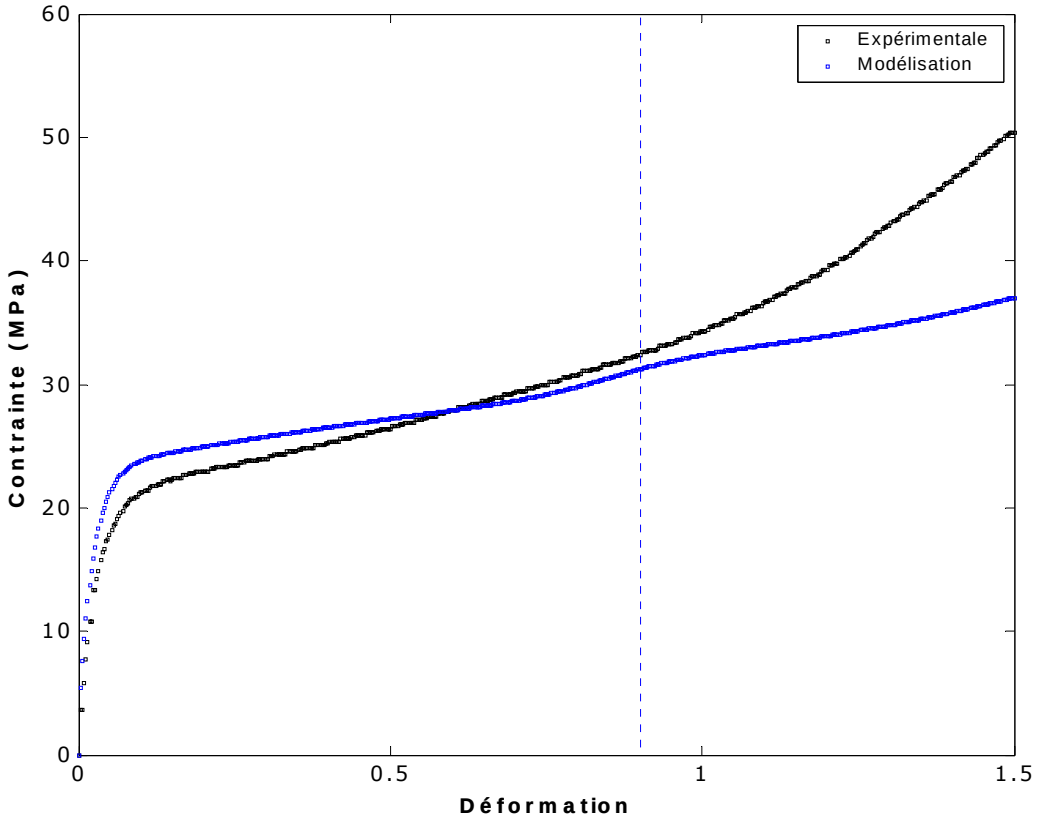
FigureVI.6. Erreur due à l'apprentissage



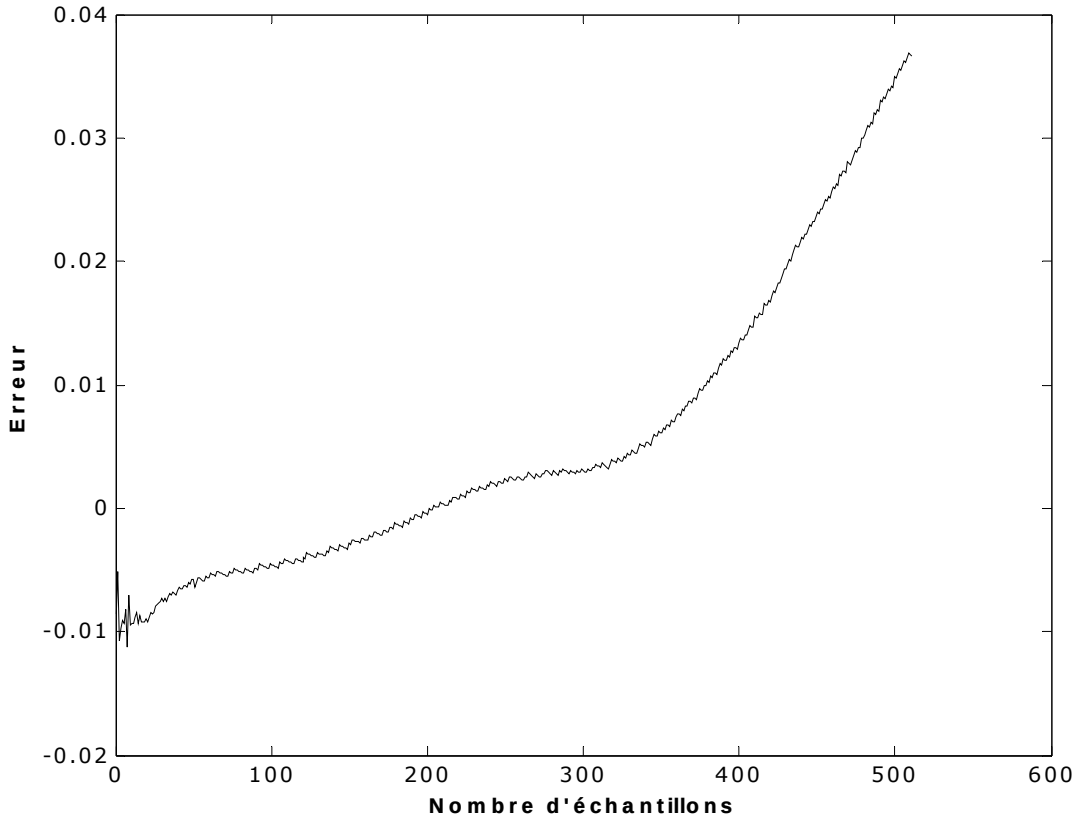
FigureVI.7. Test de validation



FigureVI.8. Erreur due à la validation



FigureVI.9. Test de généralisation pour une vitesse de déformation $\dot{\varepsilon}_T = 1.\ 10^{-4}\ s^{-1}$



FigureVI.10. Erreur commise sur l'ensemble de test

VI-1.3.2 Deuxième paquet :

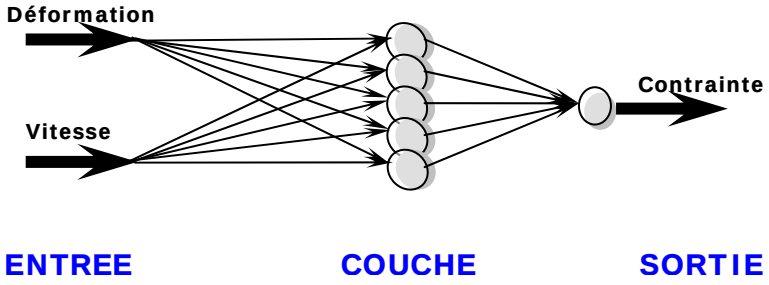
Nous augmentons l'ensemble des données d'apprentissage : Nous prenons les trois cent quarante (340) premières valeurs des contraintes et déformations de chaque courbe ici présente. C'est à dire les deux tiers de nombre des données, le reste nous l'utilisons pour la validation.

Nous effectuons la même procédure qu'avec le premier paquet, normalisation des données, apprentissage, validation et traçage des courbes.

Nous optons pour l'architecture avec cinq neurones (5) dans la couche cachée (figureVI.11), vu que c'est l'architecture pour laquelle nous avons obtenu l'écart type minimal pour l'apprentissage ainsi que la validation. Le tableauVI.3 montre bien ça :

TableauVI.3. Choix de nombre de neurones dans la couche cachée.

Nombre de neurones dans la couche cachée	RMSE due à l'apprentissage (10^{-4})	RMSE due à la validation (10^{-4})
5	7	56
6	24	125
7	23	202
8	24	3297
9	14	3006
10	10	130
12	3	87
14	11	3402



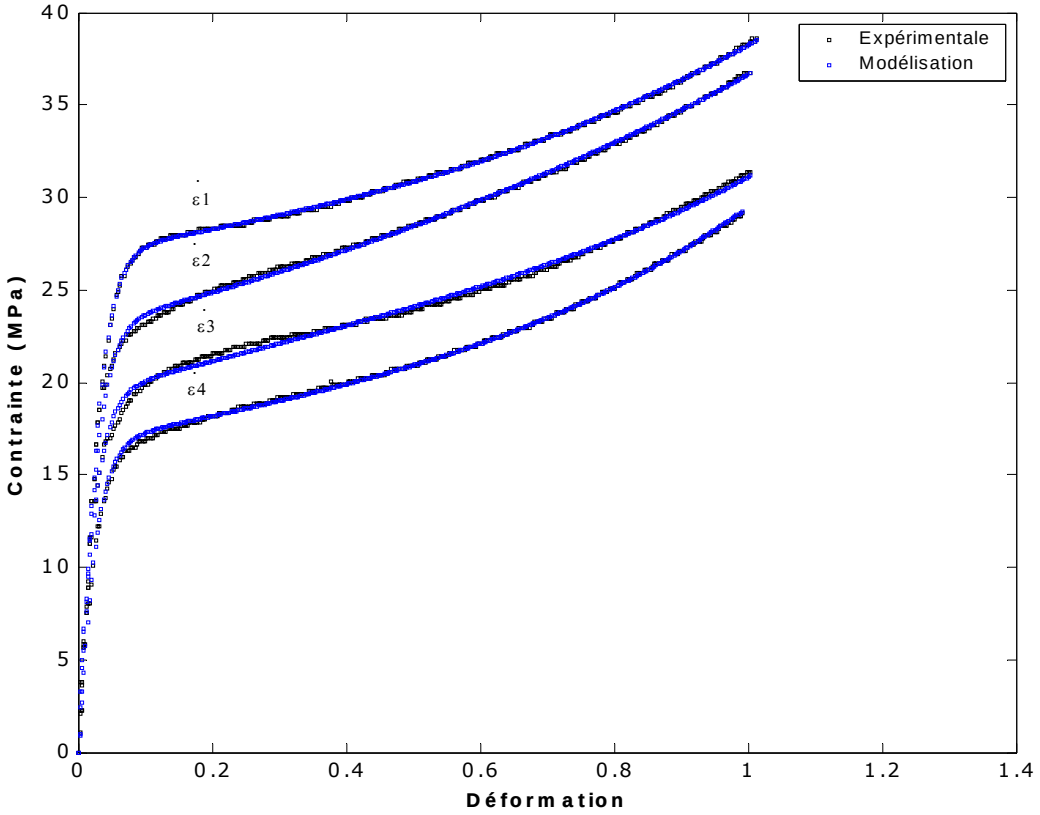
FigureVI.11. Architecture du réseau choisi

Comme précédemment, les courbes obtenues sont très proches de l'expérimentales (figureVI.12), surtout pour la première vitesse. Pour la deuxième et la quatrième vitesse, c'est la limite élastique qui est mal prédite, pour la troisième vitesse, le passage vers le domaine plastique se fait avec un écart type de 6‰.

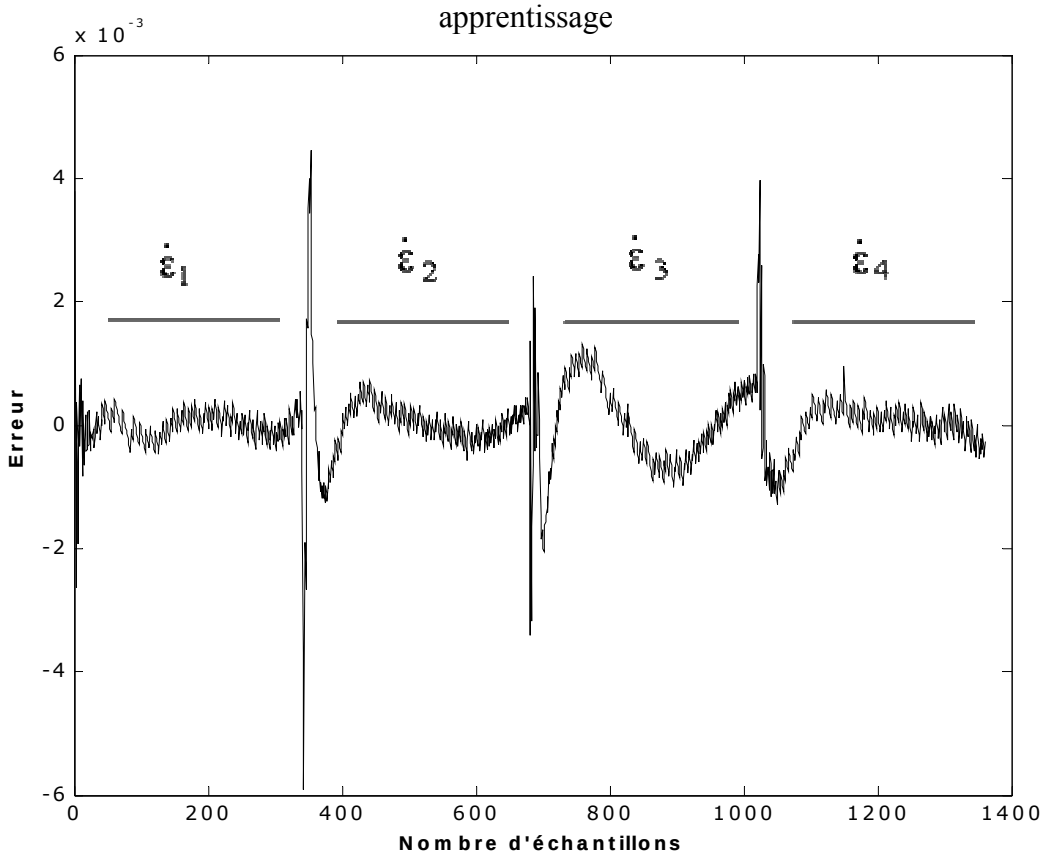
L'évolution de l'erreur (figureVI.13) confirme ces observations, elle est très importante au début de la deuxième vitesse, et au milieu de la troisième vitesse.

Avec l'augmentation de l'ensemble d'apprentissage, nous nous attendant à une meilleure prédiction dans le domaine de validation, c'est bien fait pour la quatrième vitesse, assez bien pour la deuxième et la troisième vitesse mais pas pour la première vitesse de déformation (figureVI.14), la figureVI.15 montre l'importance de l'erreur due à la validation pour cette dernière.

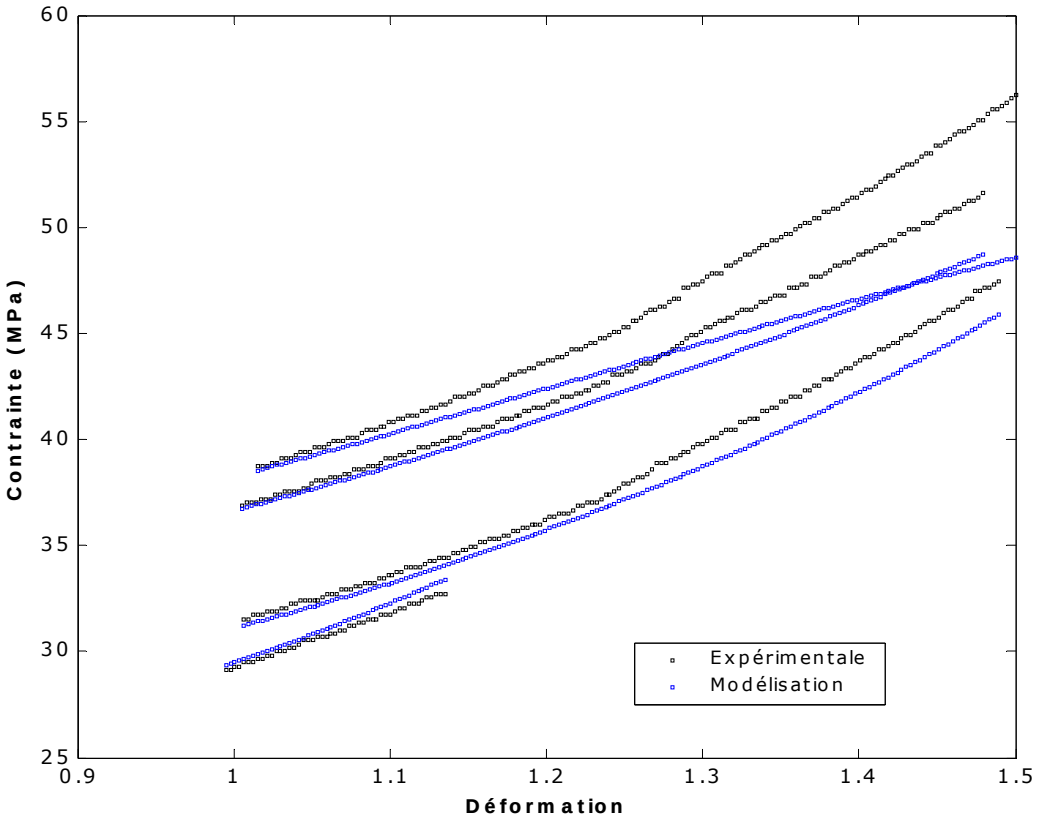
Pour le test de généralisation, nous avons remarqué la similarité de l'allure des deux courbes : Expérimentale et celle obtenue après simulation (figureVI.15). Cette dernière est un peu décalée vers le haut au début du domaine plastique et vers le bas pendent le durcissement (partie de la validation), ce n'est que le résultat de l'apprentissage : La vitesse de généralisation est prise entre la deuxième et la troisième vitesse, pour lesquelles nous avons remarqué l'importance de l'erreur due à l'apprentissage au début du domaine plastique, et leurs décalages lors du durcissement du matériau.



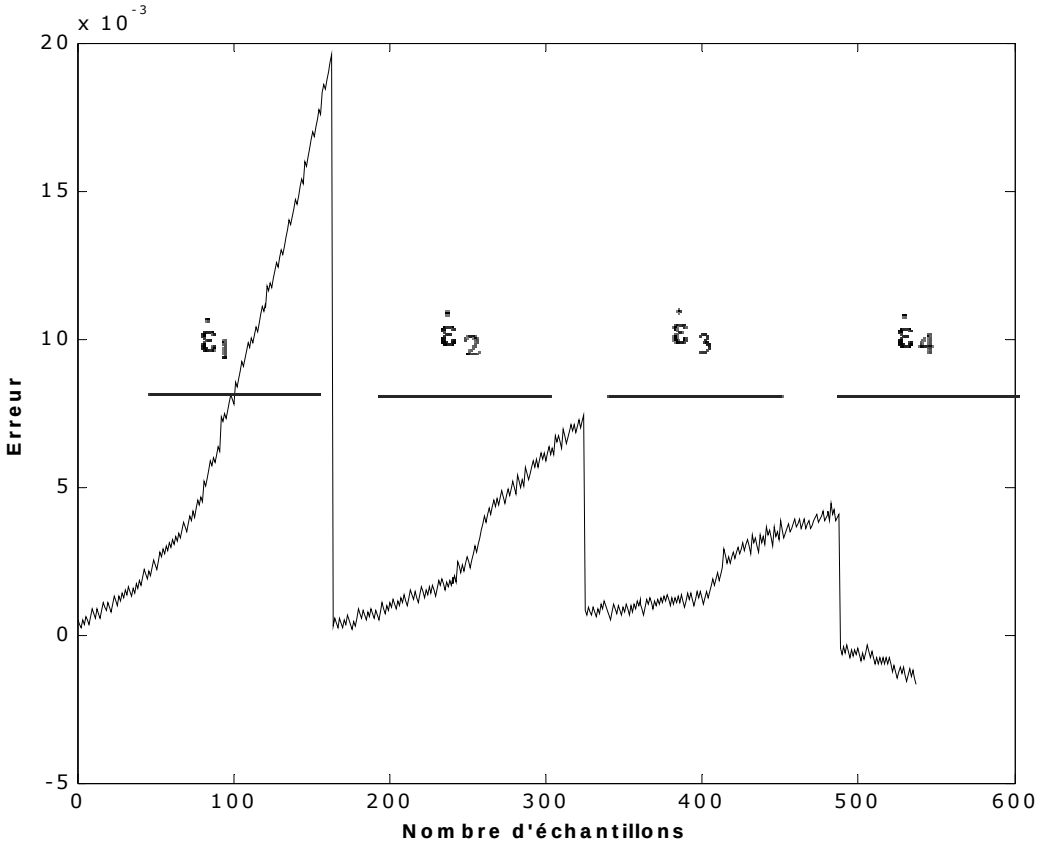
FigureVI.12. Comparaison entre les valeurs expérimentales et celles obtenues après apprentissage



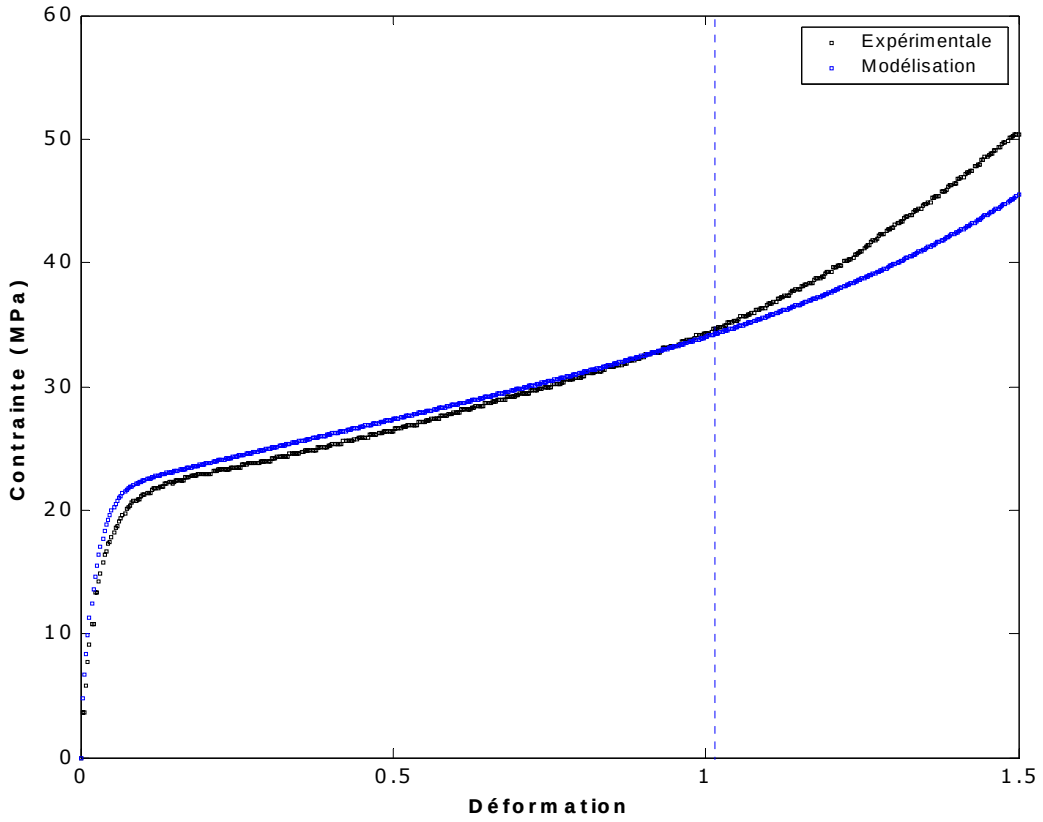
FigureVI.13. Erreur due à l'apprentissage



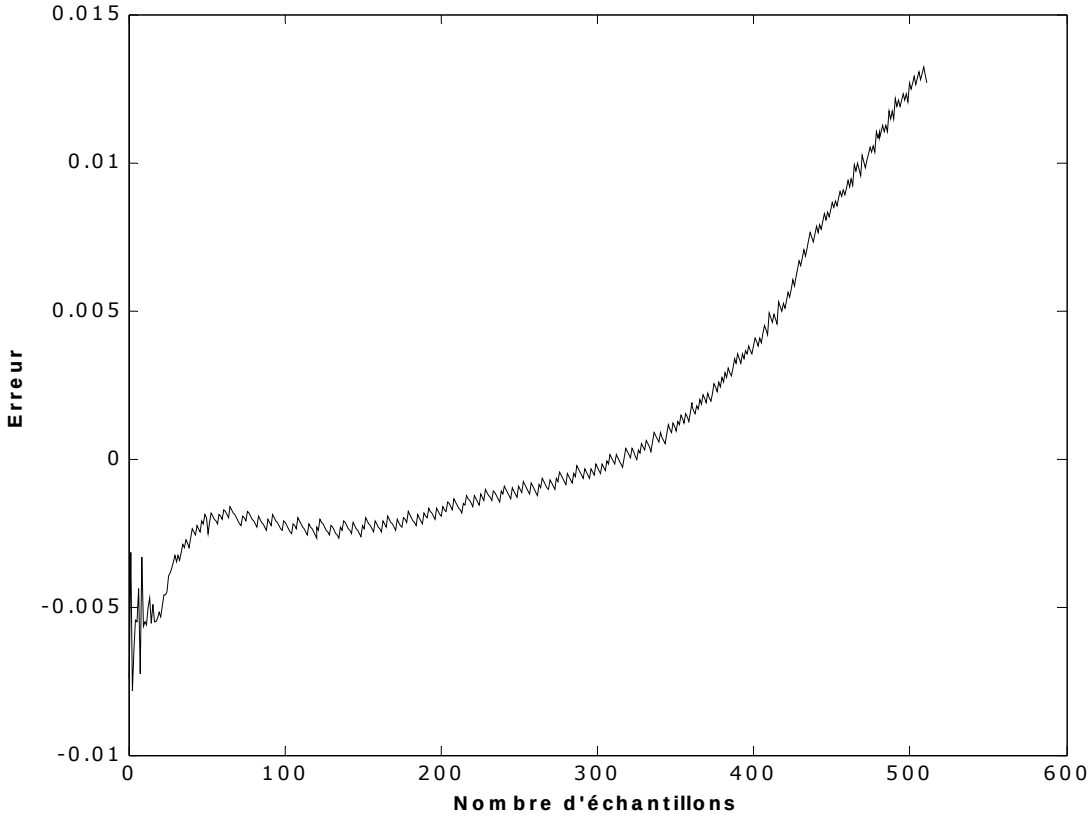
FigureVI.14. Test de validation



FigureVI.15. Erreur due à la validation



FigureVI.16. Test de généralisation pour une vitesse de déformation $\dot{\varepsilon}_T = 1.\ 10^{-4}\ s^{-1}$



FigureVI.17. Erreur commise sur l'ensemble de test

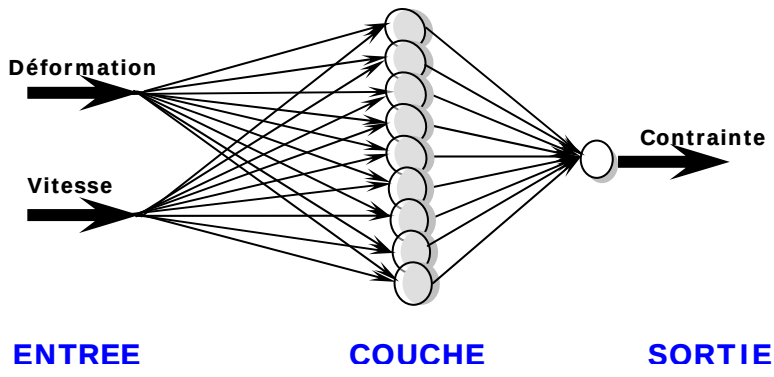
VI-1.3.3 Troisième paquet :

Nous augmentons l'ensemble des données d'apprentissage : Nous prenons les trois cents quatre-vingt neuf (390) premières valeurs des contraintes et déformations de chaque courbe. C'est à dire plus que les deux tiers du nombre de données pour l'apprentissage, le reste nous l'utilisons pour la validation. La quatrième courbe (de la cinquième vitesse de déformation) entre entièrement dans l'ensemble d'apprentissage.

Nous optons pour une architecture avec neuf neurones dans la couche cachée (figureVI.18), vu que c'est l'architecture pour laquelle nous avons obtenu l'écart type minimal pour l'apprentissage ainsi que la validation, ceci est exposé dans le tableauVI.4 :

TableauVI.4 : Choix du nombre de neurones dans la couche cachée.

Nombre de neurones dans la couche cachée	RMSE due à l'apprentissage (10^{-4})	RMSE due à la validation (10^{-4})
5	44	215
6	11	52
7	21	87
8	34	25591
9	5	90
10	23	811
12	13	2115
14	3	150

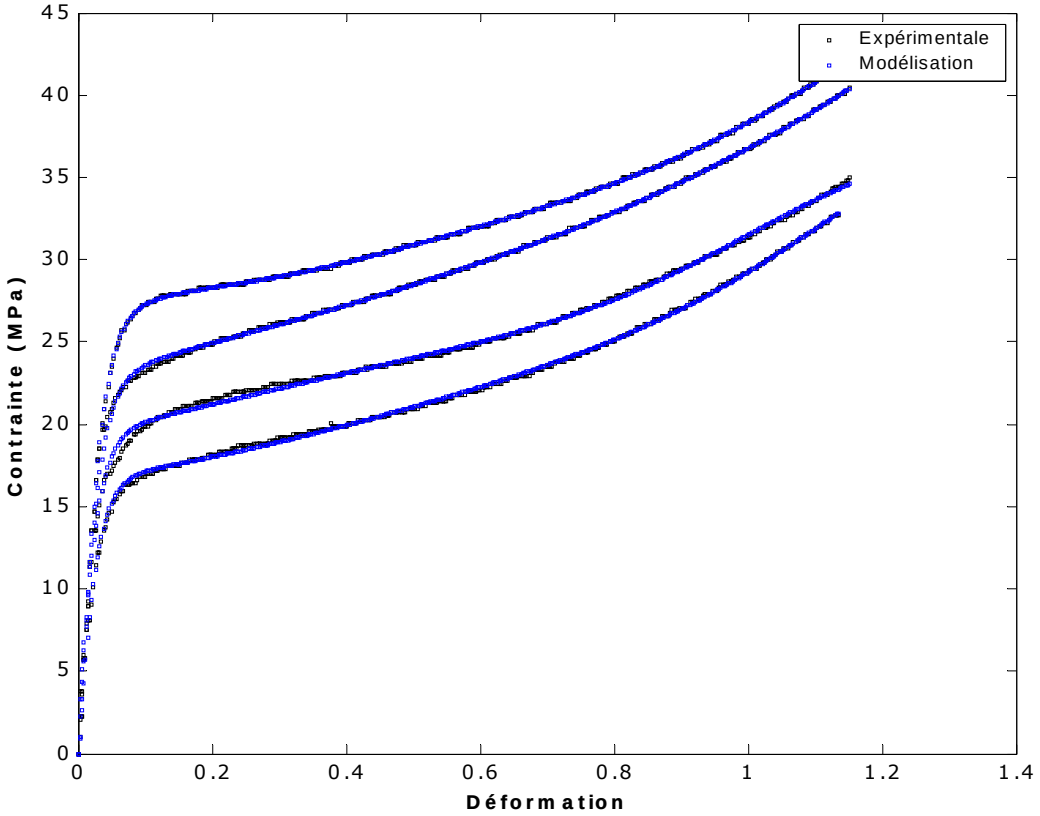


FigureVI.18. Architecture du réseau choisi

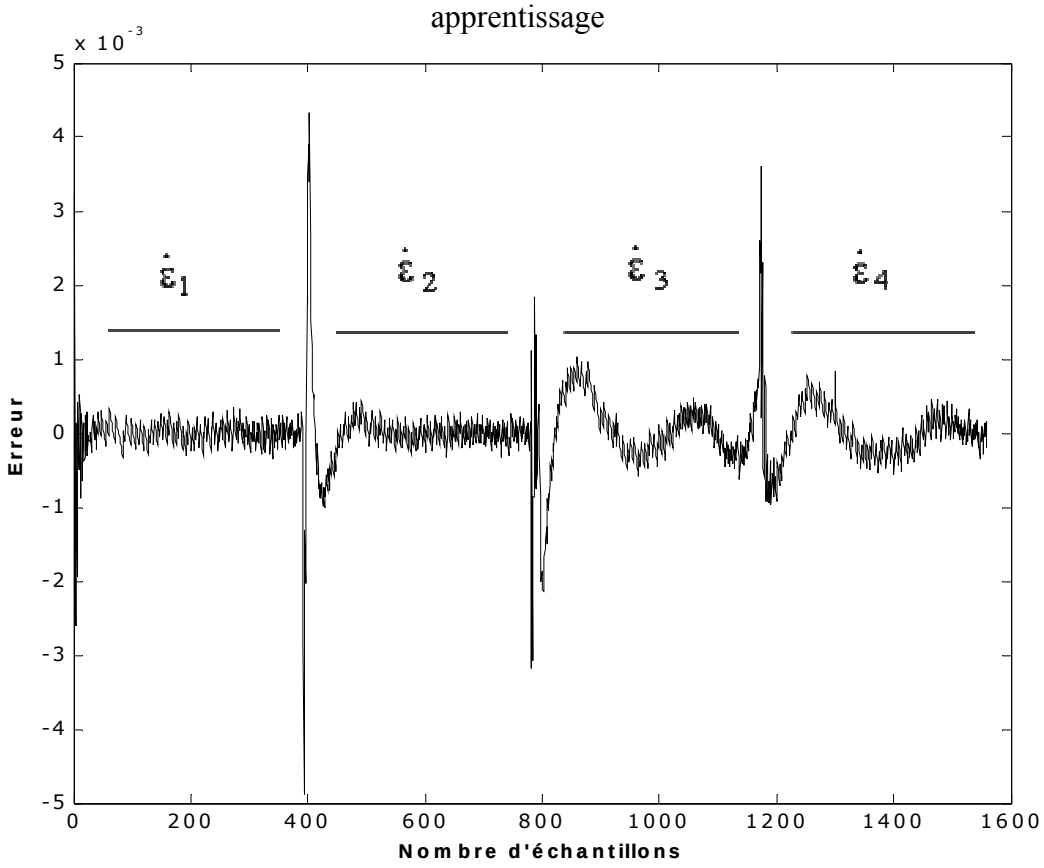
Les courbes obtenues après simulation (figureVI.19), passe par les données expérimentales, sauf pour la troisième vitesse : La courbe est un peu décalée vers le haut, toujours le passage vers le non linéaire qui cause un écart de 4‰. L'évolution de l'erreur au cours de l'apprentissage (figureVI.20) montre son importance au début de la deuxième courbe et le long de la troisième et quatrième courbe.

Le réseau valide bien pour la première et la deuxième vitesse, mal pour la troisième (figureVI.21). L'évolution de l'erreur de validation montre l'importance de cette dernière pour la troisième vitesse (figureVI.22).

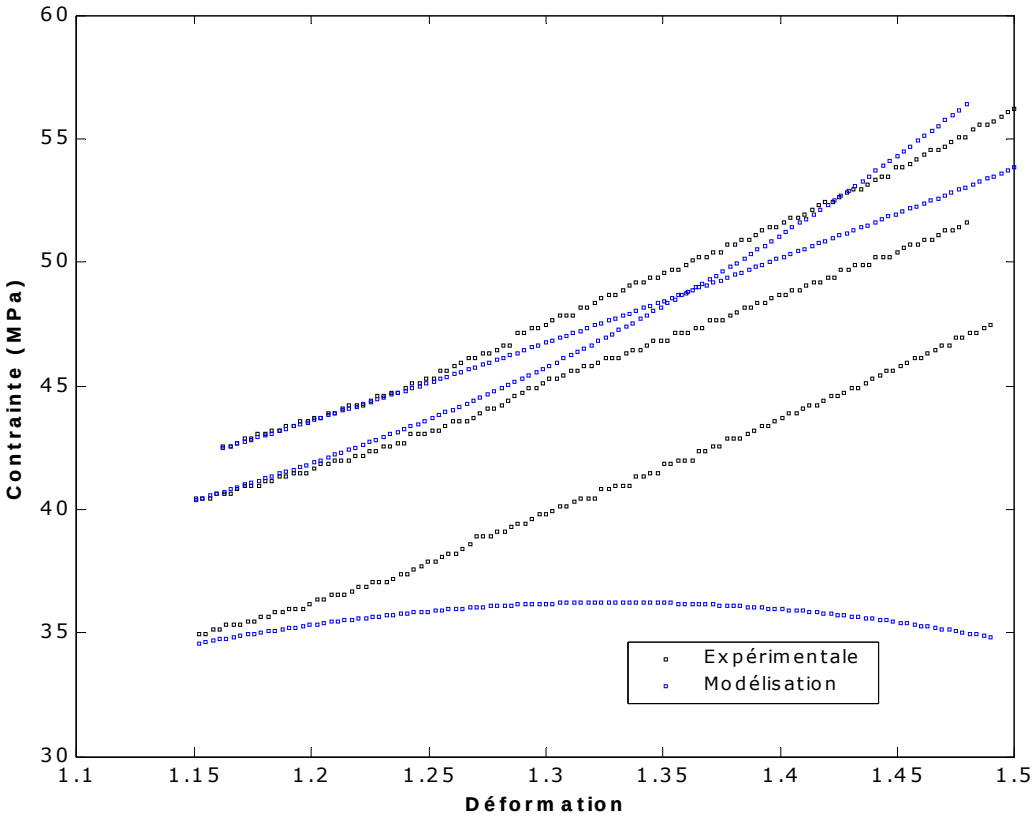
En ce qui concerne la généralisation (figureVI.23), le réseau n'arrive pas à donné des valeurs proches de l'expérimentale ou de l'allure générale de la courbe : Comme nous venons de le constater précédemment, la généralisation n'est que le résultat d'apprentissage, l'erreur est bien importante, surtout à partir du domaine des déformations plastiques (figureVI.24).



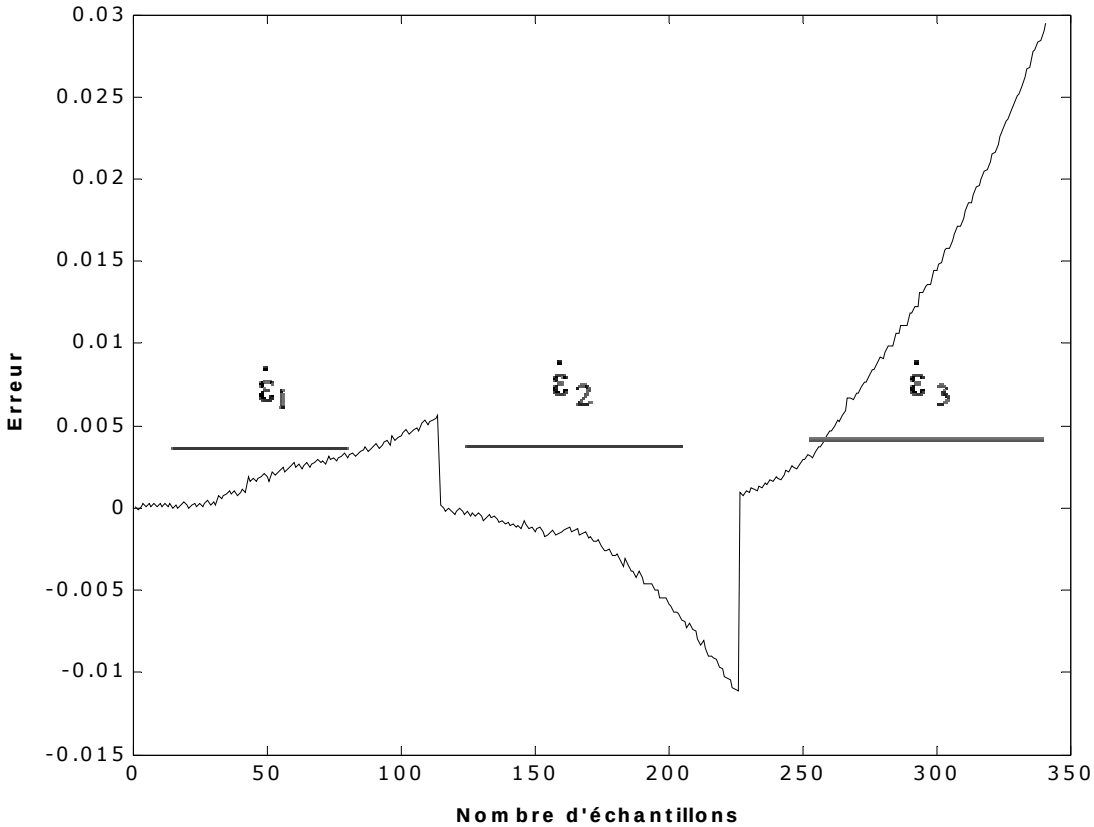
FigureVI.19. Comparaison entre les valeurs expérimentales et celles obtenues après apprentissage



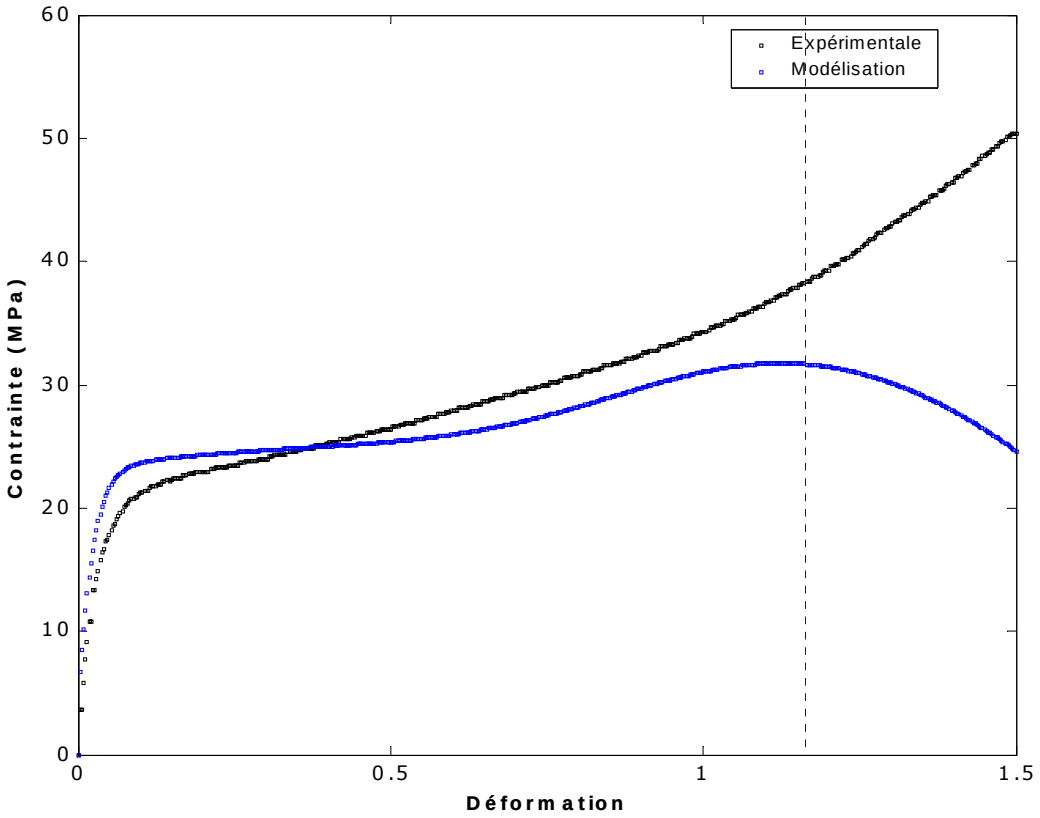
FigureVI.20. Erreur due à l'apprentissage



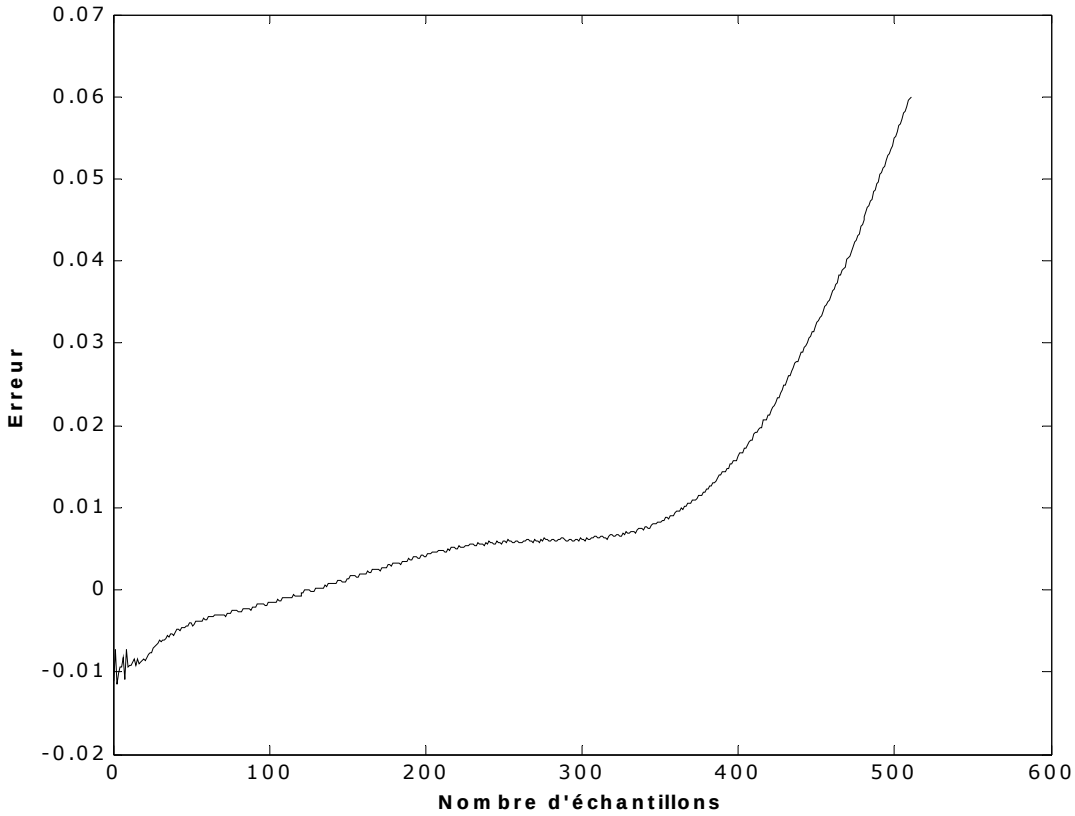
FigureVI.21. Test de validation



FigureVI.22. Erreur due à la validation



FigureVI.23. Test de généralisation pour une vitesse de déformation $\dot{\varepsilon}_T = 1.\ 10^{-4}\ s^{-1}$



FigureVI.24. Erreur commise sur l'ensemble de test

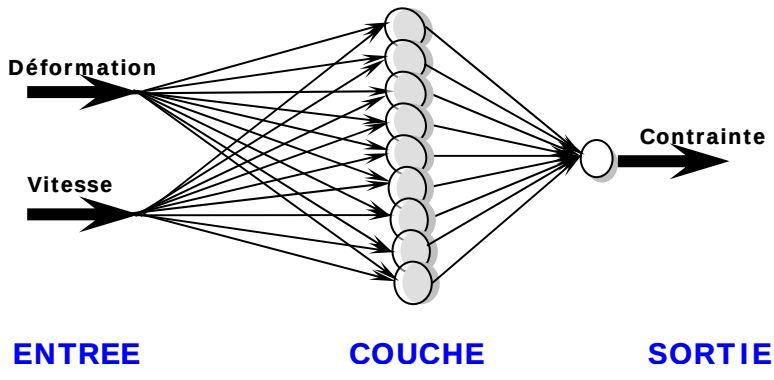
VI-1.3.4 Quatrième paquet :

Nous allons refaire la même chose que précédemment, mais en augmentant encore la taille de l'ensemble d'apprentissage, nous prenons les quatre cents (400) premières valeurs des contraintes et déformations de chaque courbe, c'est à dire plus de deux tiers du nombre de données, le reste nous l'utilisons pour la validation. La quatrième courbe, cette fois-ci encore, entre entièrement dans l'ensemble d'apprentissage.

Nous optons pour une architecture avec neuf neurones dans la couche cachée (figureVI.25), le tableauVI.4 justifie ce choix :

TableauVI.5 : Choix du nombre de neurones dans la couche cachée.

Nombre de neurones dans la couche cachée	RMSE due à l'apprentissage (10^{-4})	RMSE due à la validation (10^{-4})
5	45	125
6	24	140
7	19	159
8	27	233
9	4	27
10	6	113
12	3	68
14	5	290

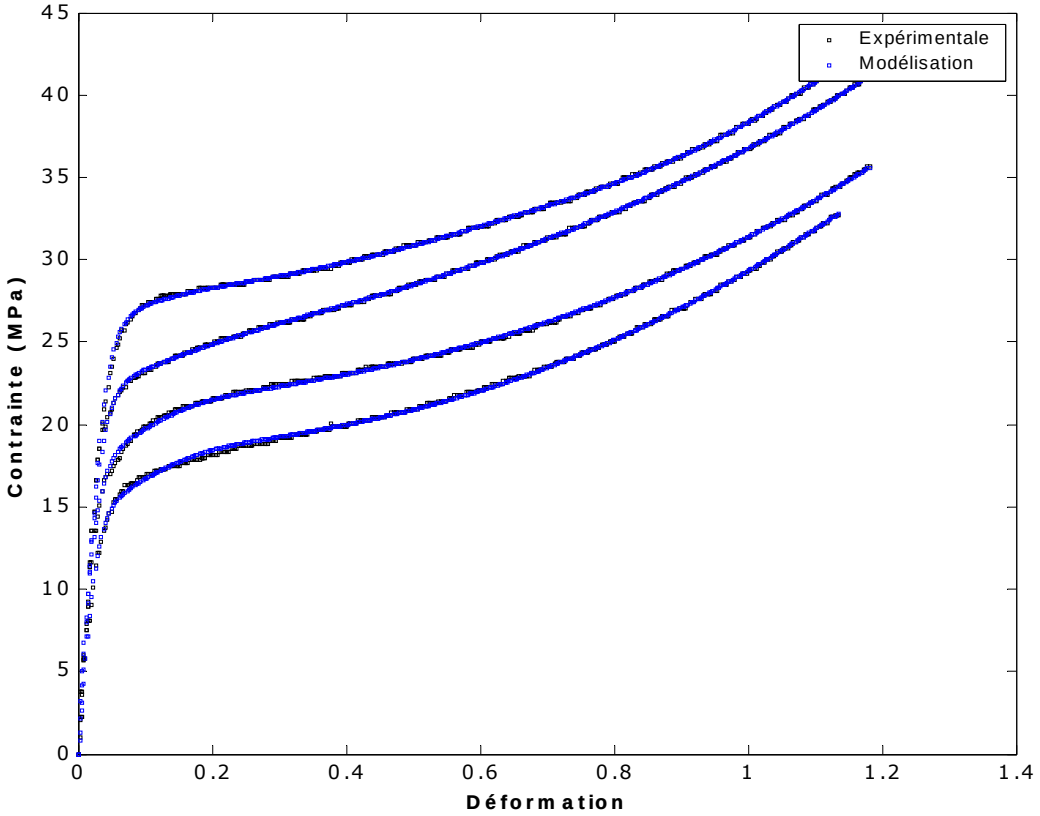


FigureVI.25. Architecture du réseau choisi

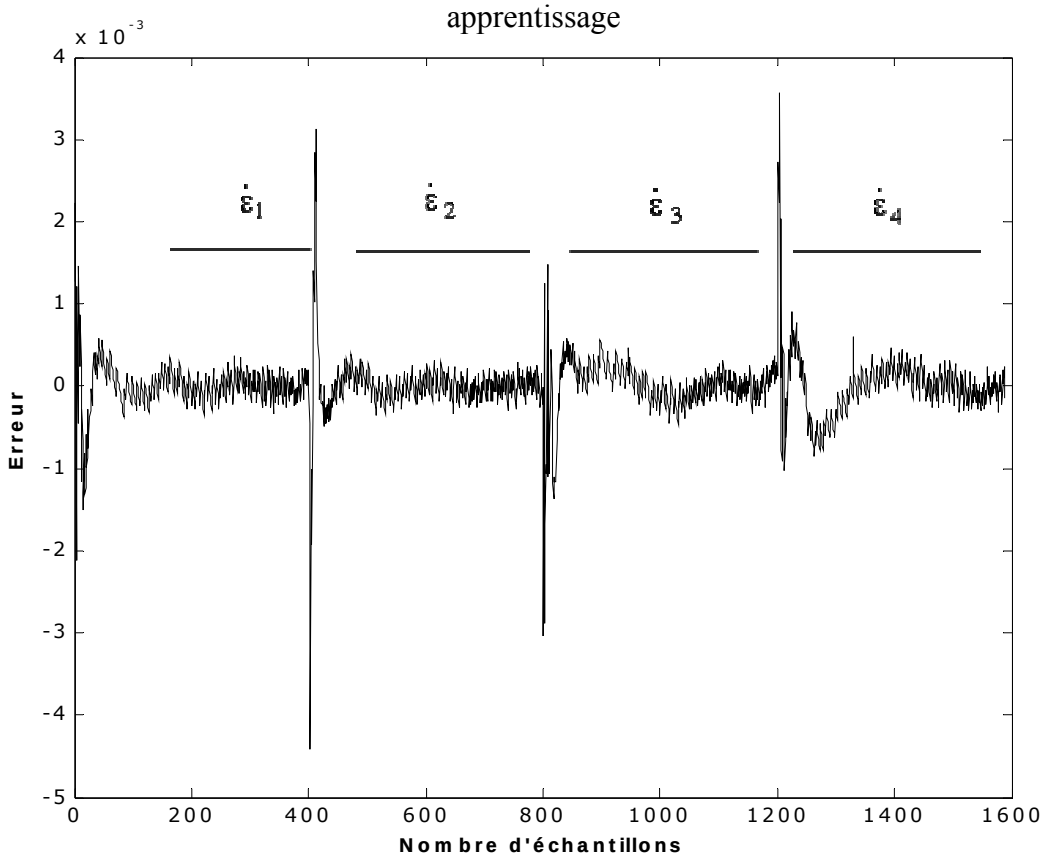
Les courbes obtenues après apprentissage passent par tous les points expérimentaux (figureVI.26), pour la quatrième vitesse de déformation, il y a un léger décalage au début du

domaine plastique. L'erreur est régulière pour les trois premières vitesses (figureVI.27) sauf pour la quatrième ce qui affirme nos observations. La validation est meilleure que la précédente (figureVI.28), les courbes sont proches de l'expérimentale avec un faible écart de l'ordre de 3‰, l'erreur de validation est importante pour la première et la troisième vitesse (figurVI.29).

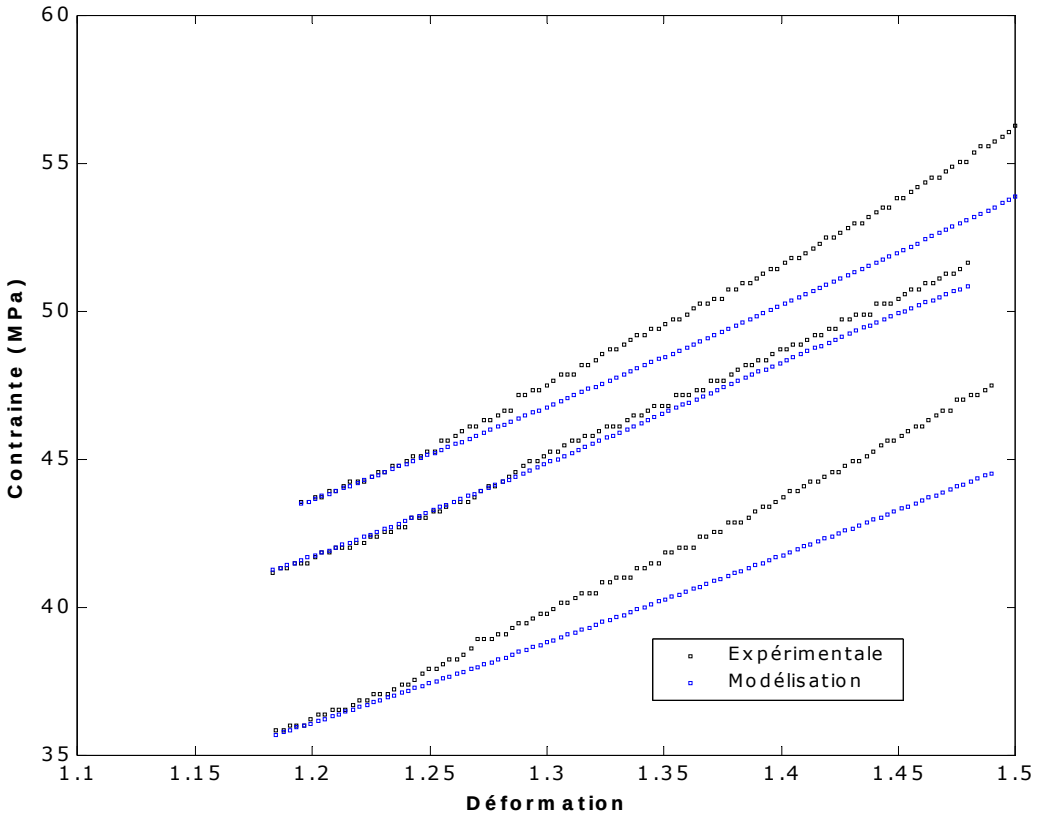
La figureVI.30 illustre la courbe simulée lors de la généralisation, elle est très proche de l'expérimentale. L'allure générale inclut le comportement mécanique enseigné au réseau de neurones : Un domaine élastique suivi par un long stade plastique caractérisé par un durcissement progressif. C'est cette progression dans le durcissement qui est mal prédite par le réseau (figureVI.31), aussi le passage vers le non linéaire fait que la courbe s'écarte avec 4‰ de l'expérience.



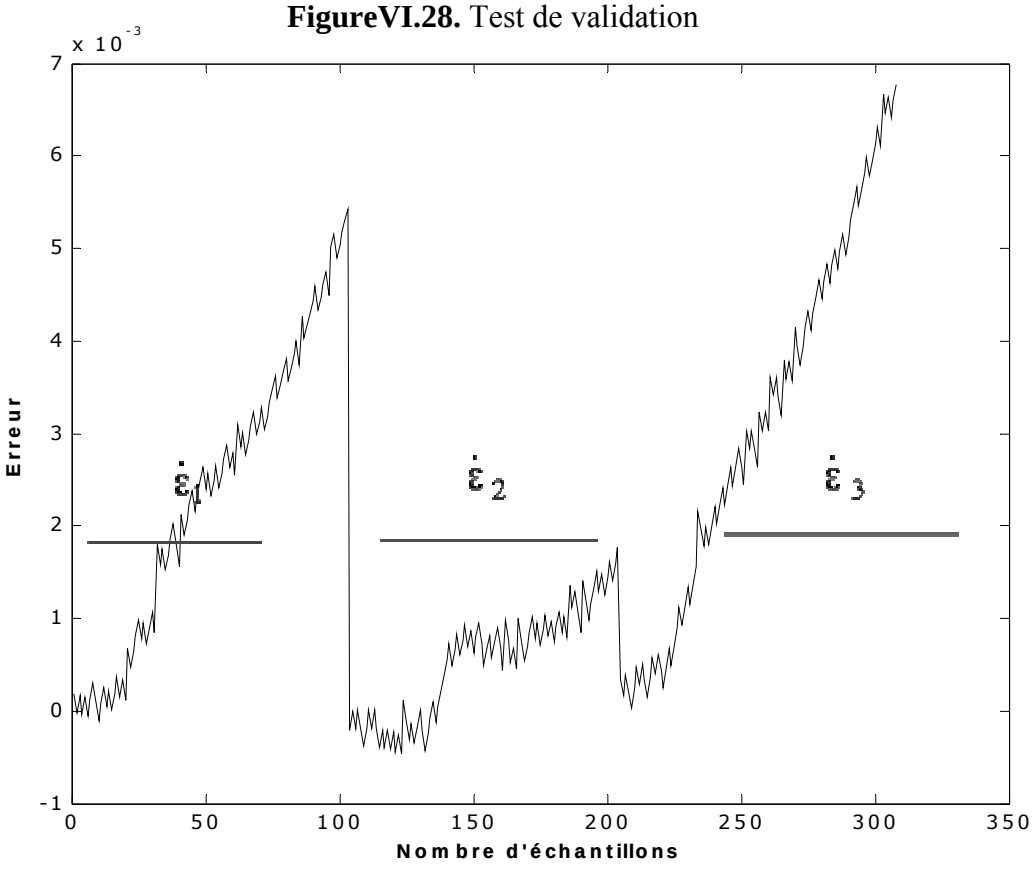
FigureVI.26. Comparaison entre les valeurs expérimentales et celles obtenues après apprentissage



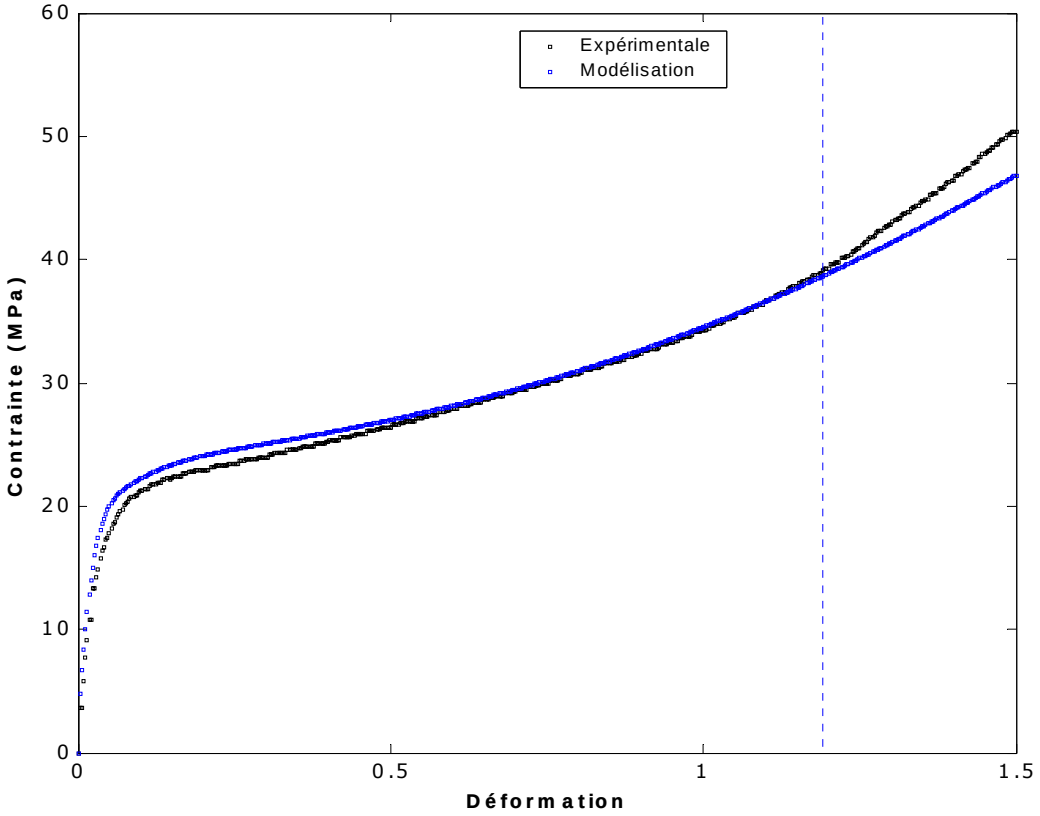
FigureVI.27. Erreur due à l'apprentissage



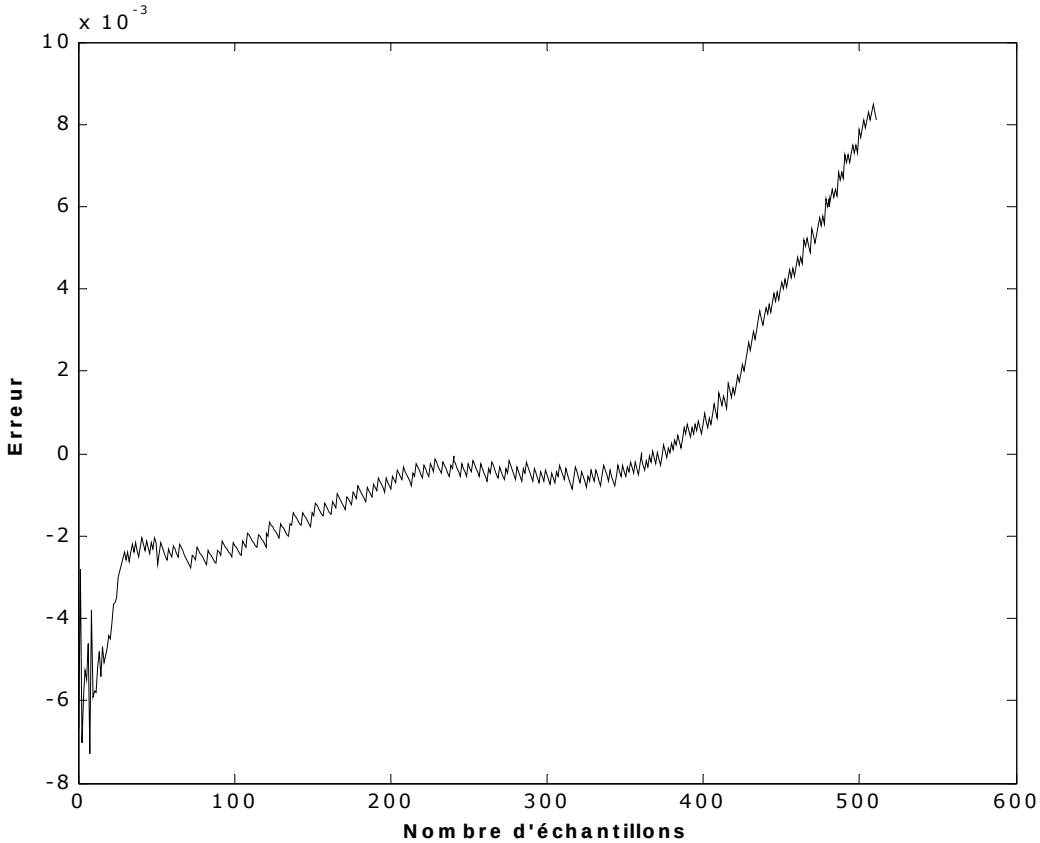
FigureVI.28. Test de validation



FigureVI.29. Erreur due à la validation



FigureVI.30. Test de généralisation pour une vitesse de déformation $\dot{\varepsilon}_T = 1.\ 10^{-4}\ s^{-1}$



FigureVI.31. Erreur commise sur l'ensemble de test

Nous venons de constater l'importance du choix de jeu de donnée pour la réalisation de l'apprentissage. Notre loi de comportement est complexe [HISS R. et al. 1999], elle varie de l'élastique, le plastique au durcissement ce qui rend le choix des échantillons qui la présente fidèlement difficile, ça nous a poussé à utiliser une autre méthode.

VI-2 LA MÉTHODE *EARLY STOPPING*:

On s'intéresse d'habitude, lors de l'apprentissage d'un réseau neurone, à obtenir un réseau avec une performance optimale [PRECHELT L., 1998].

Une autre méthode pour améliorer la généralisation s'appelle *Early stopping* ; Dans cette technique les données disponibles sont divisées en trois sous-ensembles :

Le premier est celui d'apprentissage, qui est utilisé pour calculer le gradient et mettre à jour les poids et les biais de réseau.

Le deuxième est celui de validation. L'erreur sur l'ensemble de validation est surveillée pendant le processus d'apprentissage. Elle diminuera normalement pendant la phase initiale d'apprentissage, de même que l'erreur d'apprentissage. Cependant, quand le réseau commence à l'*overfit.* (Sur-apprentissage) des données, l'erreur sur l'ensemble de validation commencera typiquement à monter. Quand elle augmente pour un numéro indiqué des itérations, l'apprentissage est arrêté, les paramètres (poids et biais) sont retournés aux minimums.

L'erreur sur le sous-ensemble de test n'est pas utilisée pendant l'apprentissage, mais elle est employée pour comparer différents modèles. Il est également utile de tracer l'erreur sur l'ensemble de test pendant le processus d'apprentissage. Si l'erreur sur l'ensemble du test atteint un minimum à un numéro sensiblement différent d'itération que l'erreur sur celui de validation, ceci peut indiquer une division faible de l'ensemble général des données.

La méthode Early stopping peut être utilisée avec n'importe quelle fonction de transfert qui a été décrite précédemment (chapitreIII). On doit simplement passer les données de validation à la fonction de transfert.

Soit le vecteur qui rassemble les données d'entrée P, les sous ensembles sont construites comme suit :

Pour le sous ensemble de validation on prend chaque quatre valeurs une valeur à partir de la deuxième valeur de P : 2:4:q (q est la taille de P)

Pour le sous ensemble de test on prend chaque quatre valeurs une valeur à partir de la quatrième valeur de P : 4:4:q (q est la taille de P)

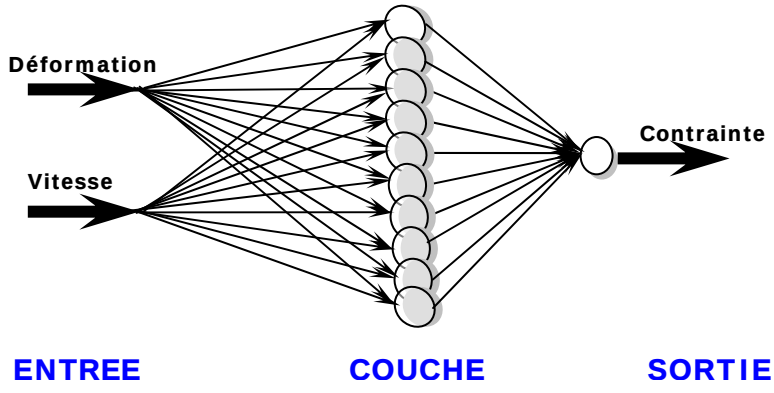
Le reste est pour le sous-ensemble d'apprentissage : [1:4:q3:4:q]

P = [x_1 x_2 x_3 x_4 x_5 x_6 x_7 x_8 x_9 x_{10} x_{11} x_{12} x_{13} x_{14} x_{15} x_{16} …]

De cette façon, l'apprentissage se fait le long des courbes et passe par les trois domaines de comportement élasto-plastique de notre matériau. Nous avons choisi un feedforward avec dix neurones (figureVI.32) dans la couche cachée, le tableauVI.6 justifie ce choix :

TableauVI.6 : Choix du nombre de neurones dans la couche cachée.

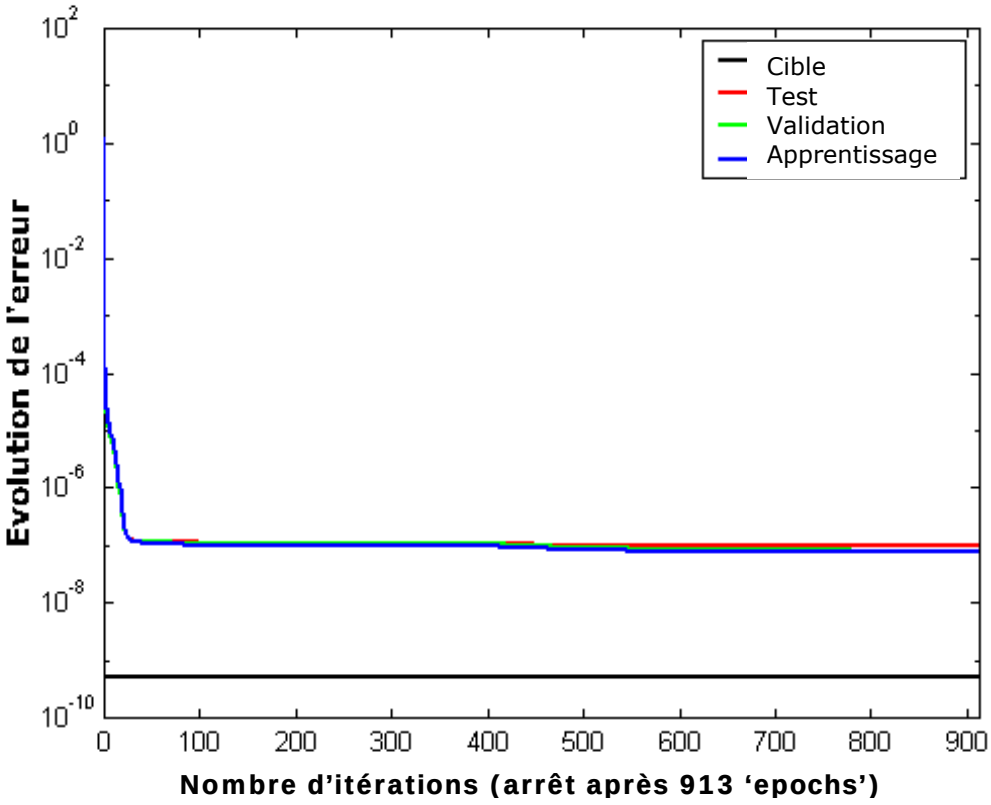
Nombre de neurones dans la couche cachée	**RMSE (10^{-4})** (Apprentissage)	**RMSE (10^{-4})** (Validation)	**RMSE (10^{-4})** (Test)
5	28	27	27
6	23	21	22
7	28	25	27
8	11	10	10
9	9	8	8
10	3	3	3
12	17	17	17
14	7	6	7



FigureVI.32. Architecture du réseau choisi

L'application de cette méthode dans notre cas a donné les résultas suivants :

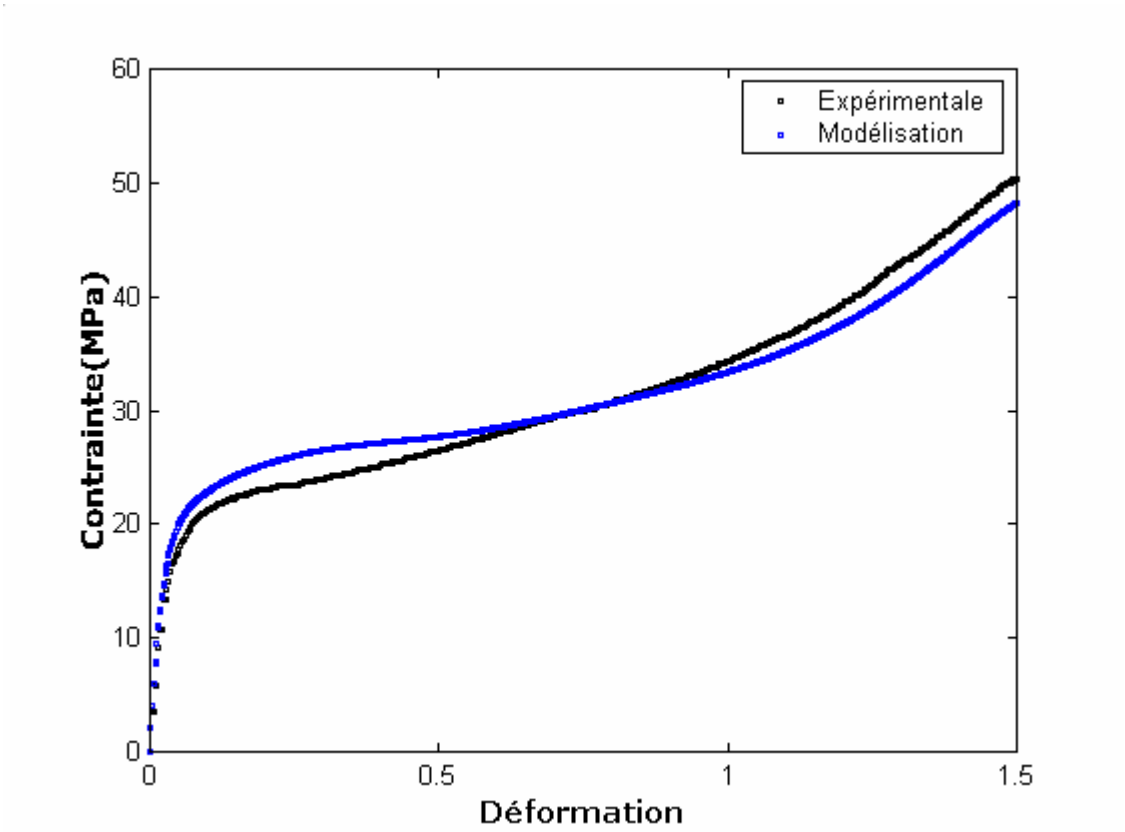
Après neuf cent treize (913) itérations, l'apprentissage s'arrête parce que l'erreur de validation a augmenté, c'est un outil de vérification très utile : Nous remarquons que l'erreur de validation et l'erreur commise lors du test ont les même caractéristiques, c'est raisonnable.



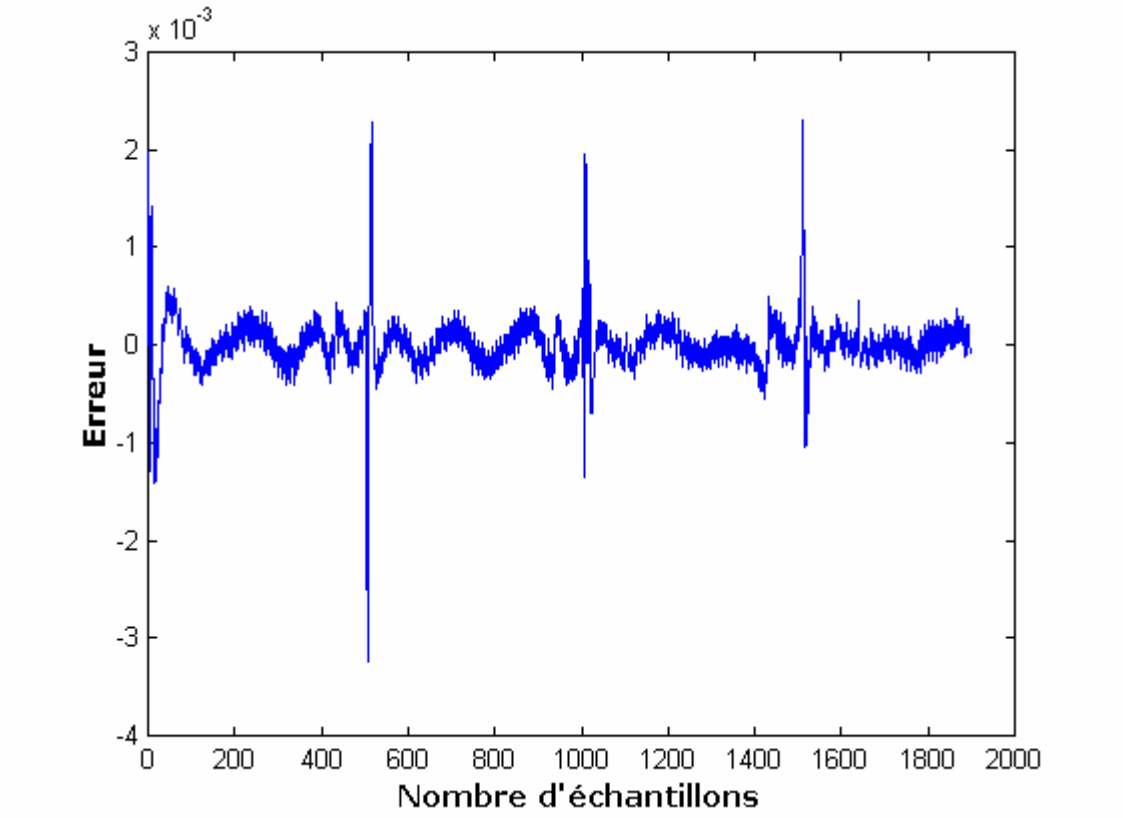
FigureVI.33. Evolution de l'erreur.

La courbe de simulation obtenue présente une allure générale proche de l'expérience : Une limite élastique diffuse, un long stade plastique suivi par un durcissement structural du matériau. Les valeurs de la simulation sont un peu écartées de l'expérimentation (figureVI.34). Nous expliquons cet écart par un manque dans la base des données, les réseaux de neurones sont connus par leur robustesse. Pour présenter un comportement élastoplastique avec seulement quatre courbes correspondantes à quatre vitesses de déformation il faut s'attendre à un tel écart.

L'erreur est plutôt régulière (figureVI.35), elle est importante lors du passage du comportement de l'élastique vers le plastique.



FigureVI.34. Comparaison entre les valeurs expérimentales et celles obtenues après simulation pour une vitesse de déformation $\dot{\varepsilon}_T = 1.\ 10^{-4}\ s^{-1}$



FigureVI.35. Erreur commise sur l'ensemble d'apprentissage

CONCLUSION

Les modèles neuronaux les plus utilisés, le perceptron multicouches : MLP, sont présentés dans le cadre de la régression des fonctions non linéaires et de la prédiction. Le perceptron à une couche cachée, à fonction d'activation linéaire en sortie, fournit une structure simple et flexible pour la modélisation non linéaire.

Nous avons voulu démontrer dans ce travail les réels avantages des RNA en les utilisant pour la détermination d'une loi de comportement effective (fonction non connue). Nous avons utilisé un perceptron à une couche cachée, connu par sa propriété d'approximation universelle, pour plus de précision un feedforward avec un apprentissage supervisé. Le réseau, quoiqu'il donne des écarts considérables, approche bien notre loi de comportement.

Généralement, le réseau fait un bon apprentissage et validation pour la première et la quatrième vitesse de déformation et un mauvais apprentissage et validation pour la troisième vitesse de déformation pour les quatre paquets.

Mais, c'est le quatrième paquet qui approche le plus notre loi de comportement malgré les écarts de la courbe au début et vers la fin du domaine plastique : Ce n'est que le résultat d'apprentissage et c'est dû à l'écart entre les courbes qui ont alimenté notre réseau.

L'utilisation de la méthode early stopping a amélioré le résultat : Une allure générale de la courbe de contrainte-déformation plus proche de la courbe réelle, le domaine élastique à pente décroissante, la limite élastique diffuse, le long domaine plastique, le durcissement structural du matériau qui caractérise la fin de ce domaine.

L'avantage réel de l'utilisation des RNA pour la prédiction est que c'est facile à réaliser, une fois le bon choix effectué. Il ne reste que la vérification de la reproductibilité du modèle établi par un test de généralisation, les RNA ne demandent aucun prétraitement ni simplification des valeurs d'entrée : Flexibilité et parcimonie vérifiées.

Mais, sans une banque de données bien homogène et dense les avantages des RNA s'évanouissent et montrent leur limite comme l'atteste les résultats des premier et troisième paquet.

Ces différentes remarques prouvent que le succès d'une méthode dépend d'un ensemble de paramètres qui vont de la collecte des données expérimentales plus raffinées sur le

comportement du matériau au choix du type de RNA et de son utilisation, et qu'il est, par conséquent, extrêmement difficile de tirer des conclusions définitives sur une approche.

L'introduction de différentes variables concernant le matériau (module de Young, taux de cristallinité, contrainte maximale, limite élastique, coefficient de poisson, etc.) ou les conditions expérimentales (température, temps, etc.) amélioreront le résultat de prédiction et le rendront optimal.

Avec les résultats ainsi obtenus, nous espérons aboutir au but de ce travail :

* Construire un réseau de neurone ayant les capacités de prédire la loi de comportement effective du HDPE en traction pour des vitesses de déformation prises entre 10^{-5} et 10^{-3} s^{-1}. Bien que les cinq réseaux construits présentent certaines limitations (au début et vers la fin du comportement plastique), ils ont bien appris le comportement : Un domaine de déformation élastique, une limite élastique diffuse qui ne dépasse pas 25 MPa, un long domaine de déformation plastique et, pour une vitesse de déformation $\dot{\varepsilon}_T = 1.10^{-4}$ s^{-1} prise entre $\dot{\varepsilon}_2 = 5.10^{-4}$ s^{-1} et $\dot{\varepsilon}_3 = 5.10^{-5}$ s^{-1}, la courbe obtenue se trouve entre les deux courbes correspondantes aux deux vitesses.

* Apprendre à programmer avec le MATLAB : Dans le MATLAB, programmer un RNA devient facile du moment où l'application est claire (choix de l'échantillon d'apprentissage, type du RNA, nombre de neurones dans la couche cachée,..), les fonctions nécessaires lors de la programmation sont prédéfinies, il faut seulement savoir leurs noms et leurs applications ainsi que les conditions relatives à ces applications.

* S'initier aux méthodes connexionnistes d'approximation et de prédiction : Profiter de la propriété d'apprentissage des RNA pour l'évaluation rapide de la fonction de comportement effective du HDPE en traction, cette fonction qui peut être approcher pendent le comportement élastique vue que c'est un comportement où les déformations sont réversibles, régi par une loi linéaire qui est la loi de HOOKE, mais difficile à approcher mathématiquement lors du comportement plastique, domaine où les déformations sont irréversibles et la loi est non linéaire.

REFERENCES

COINTE A., ROUGER F.
Prédiction de résistances d'assemblages par organes de type tige en utilisant l'approche des réseaux de neurones ;
(2004) Séminaire LRBB-CTHA, poster : Thème prioritaire n°2, Endommagement-Rupture.

CYBENKO G.
Approximation by superpositions of a sigmoidal function;
(1989) Mathematics of Control, Signals, and Systems, 2 (4), pp. 303-314.

FOGELMAN-SOULIE F.
Méthodes connexionnistes pour l'apprentissage ;
(1988) Actes des Journées Nationales sur l'Intelligence Artificielle. Paris : Teknea, p. 275-293.

HAGAN M.T., MENHAJ M.
(1994) IEEE Transactions on Neural Networks, 5.

HECHT-NIELSEN R.
Neurocomputing;
(1990) Addison-Wesley Publishing Company, 433 p.

HISS R., HOBEIKA S., LYNN C. et STROBL G.
Network stretching, slip processes, and fragmentation of crystallites during uniaxial drawing of polyethylene and related copolymers. A comparative study;
(1999) Macromolecules, 32, pp. 4390-4403.

LEVENBERG K.
A méthode for the solution of certain nonlinear problems in least squares;
(1944) Quarterly of applied mathematics, 2 (2), pp. 164-168.

MARQUARDT D.W.
An algorithm for least squares estimation of nonlinear parameters.
(1963) J. Soc. Indust. Appl. Math., 11, pp. 431-441.

PARIZEAU M.
Réseau de neurones ;
(2004) Université LAVAL, pp.49-50.

PRECHELT L.
Early stopping - But when?
(1998) In Orr and Mueller eds. (1998), pp. 55-69.

ORR G.B. et MUELLER K.-R.
Neural Networks: Tricks of the Trade;
(1998) Berlin : Springer, ISBN 3-540-65311-2.

RUMELHART D.E., HINTON G. et WILLIAMS R
Learning Representations by Back Propagating Errors;
(1986) Nature, 323, pp. 533-536.

Conclusion générale

Le comportement du HDPE est divisé en deux domaines : Elastique au début de l'étirage, suivi par un long domaine plastique caractérisé par un durcissement progressif : C'est un comportement dit élasto-plastique. Il est difficile à approcher mathématiquement, l'implication du temps et de la température pose problème. Nous avons trouvé dans les réseaux de neurones statiques le très bon outil pour réaliser cette étude de prédiction.

Les réseaux à une couche cachée type *feedforward* se révèlent de bons approximateurs universels parcimonieux et flexibles, avec lesquels nous avons approché la loi de comportement, à partir des données entrées-sorties.

Nous avons constaté la capacité des réseaux de neurones à estimer une fonction non connue avec un écart type de l'ordre de 1%, le passage du domaine de comportement linéaire (élastique) au non linéaire (plastique) pose un problème qui apparaît négligeable pendant l'apprentissage et la validation mais donne une erreur globale considérable pendant la généralisation : La généralisation reste à ce stade mal maîtrisée pour des valeurs de vitesse de déformation hors intervalle d'apprentissage.

Généralement, pour les quatre architectures (premier, deuxième, troisième et quatrième paquet), le réseau fait un bon apprentissage pour la première et la quatrième vitesse de déformation et un mauvais apprentissage pour la troisième vitesse de déformation. Le test de validation est utilisé pour éviter le sur-apprentissage, ou *overfitting*, du réseau où ce dernier n'arrive plus à généraliser pour de nouvelles situations : Il apprend par cœur.

Mais, c'est le quatrième paquet qui approche le plus la loi de comportement malgré les divergences de la courbe au début et vers la fin du domaine plastique : C'est dû à l'écart entre les courbes qui ont alimenté le réseau.

Nous nous sommes initiés à un nouveau domaine d'études : Apprendre à programmer un type de réseaux de neurones, feedforward, ce qui nous a permis de constater la flexibilité des RNA et la facilité de leur programmation avec le MATLAB : Il suffit de disposer et de bien choisir l'ensemble qui présente le plus fidèlement le système à prédire pour effectuer l'apprentissage. Eviter le sur-apprentissage qui nuira aux résultats de prédiction pour des nouvelles situations présentées au réseau, par une validation hors échantillon. La reproductibilité des résultats

rendra l'utilisation des RNA très intéressante, pour optimiser les résultats obtenus, nous envisageons l'introduction d'autres variables concernant le matériau (E, σ_y, taux de cristallinité, etc.) ou le procédé expérimental (température de l'expérience, temps d'étirage, etc.). C'est une perspective qui reste à mettre en oeuvre.

Une autre étude sera intéressante à mettre en pratique : une interprétation semi empirique des paramètres du réseau, donc passage d'une boite noire vers une modélisation semi physique.

Annexe1

CALCUL DU FACTEUR DE TRIAXIALITE D'APRES BRIDGMAN

INTRODUCTION

Le développement d'une instabilité plastique dans un matériau polymère a fait l'objet de plusieurs études [MILLS, 1978 ; HUTCHINSON et NEALE, 1983 ; HAUDIN et al., 1984 ; ESTRIN et al., 1991] L'analyse la plus simple consiste à considérer le champ de contraintes uniaxial. Cependant les composantes non-axiales du tenseur des contraintes cessent d'être négligeables dès que la courbure relative R/Rc du profil de l'éprouvette devient importante au niveau de l'instabilité. R et Rc sont le rayon courant et le rayon de courbure de l'éprouvette, comme le montre la figure1 il apparaît, au centre de l'éprouvette une striction.

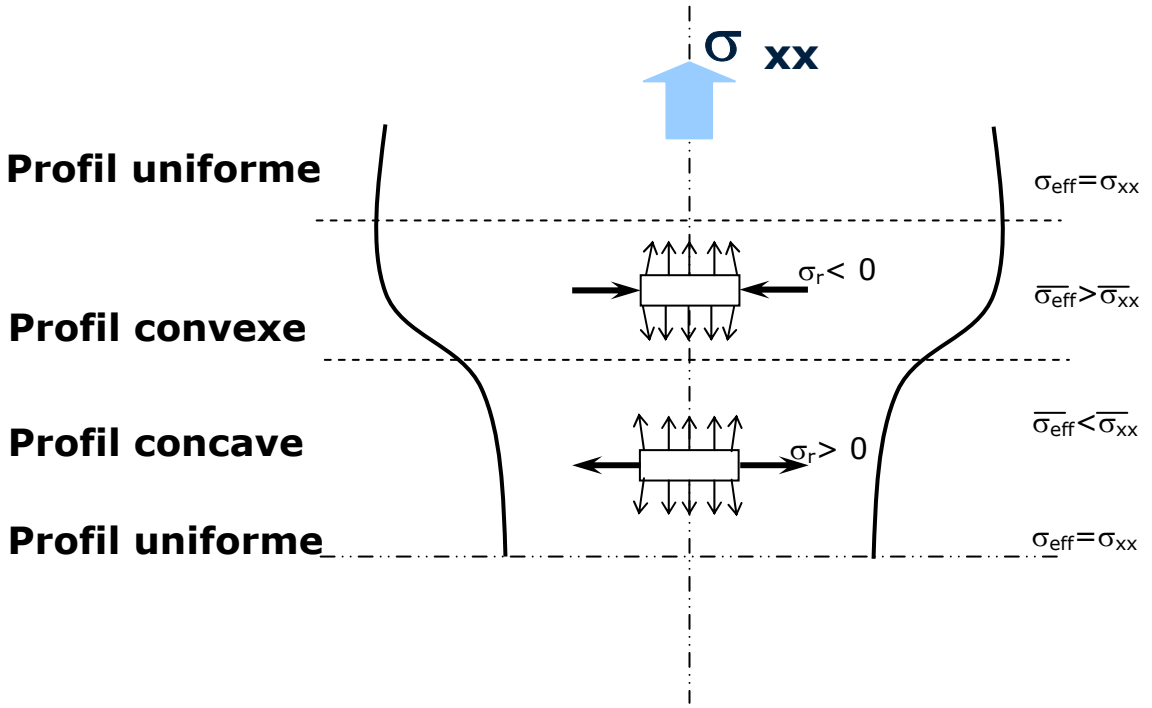


Figure1.1. Influence du profil sur la contrainte effective dans le cas de la traction.

Le phénomène de triaxialité affecte donc l'écoulement plastique car la vitesse de déformation plastique est sensible au déviateur de contraintes. Pour comparer le régime plastique en contraintes triaxiales à celui dans un essai purement uniaxial, on introduit la contrainte effective σ_{eff} qui est la contrainte de traction uniaxiale qui produirait la même vitesse de déformation effective que le champ de contraintes réel triaxial [DIETER, 1961]

La courbure est au centre de la striction et les contraintes transversales σ_{rr} et $\sigma_{\theta\theta}$ sont positives. Par conséquent, σ_{eff} y est inférieur à σ_{xx}, ce qui entraîne une certaine stabilisation de la striction. Au contraire, au niveau des épaules de la striction, la courbure est convexe avec σ_{rr} et $\sigma_{\theta\theta}$ négatives. La contrainte effective est donc supérieure à σ_{xx}. On a alors une accélération locale de la déformation, ce qui favorise la propagation de la striction vers les extrémités de l'éprouvette.

L'effet de triaxialité se traduit par l'apparition de contraintes transversales qui modifient la contrainte effective qui, à son tour, contrôle la déformation généralisée du matériau. Il semble donc bien que cet effet soit en mesure de rendre compte du phénomène de l'instabilité plastique et de combler les insuffisances de l'analyse uniaxiale qui n'explique que de manière incomplète ce processus. Pour cela, on introduit un paramètre scalaire qui exprime l'effet des contraintes transversales : $F_T = \sigma_{eff}/\sigma_{xx}$, appelé **Facteur** de **Triaxialité**. En toute rigueur, ce facteur est fonction de la géométrie du profil et des propriétés du matériau. Cependant, plusieurs auteurs ont montré que l'influence de la géométrie est prépondérante quand on considère le cas d'une striction symétrique et d'un rapport R/Rc assez faible [Bridgman, 1964 ; A. Helal, 1982 ; A. Souahi, 1986]

Dans la littérature, on trouve plusieurs descriptions des contraintes transversales dans une striction axisymétrique, mais toutes reposent sur des hypothèses simplificatrices. SIEBEL (1925) ainsi que DAVIDENKOV et SPIRIDONOVA (1946) se sont fondés sur le critère de plasticité de TRESCA. Ils trouvent FT = 1/(1+R/3Rc). L'analyse la plus couramment utilisée et celle de BRIDGMAN (1644) qui semble la plus réaliste. Dans une étude détaillée qui reste la plus citée un demi-siècle plus tard, BRIDGMAN a dérivé une autre expression du facteur de triaxialité en se fondant sur une loi d'écoulement déduite de critère de VON MISES.

FACTEUR DE TRIAXIALITE DE BRIDGMAN

Dans une analyse classique, BRIDGMAN (1944 et 1964) a considéré théoriquement le problème des effets de triaxialité dans une éprouvette axisymétrique présentant une courbure de profil, lors d'un essai de traction. Ce calcule suppose les hypothèses suivantes :

- Hypothèse 1 : Le matériau a un comportement rigide-plastique isotrope sans écrouissage.
- Hypothèse 2 : Le comportement plastique est régi par la loi de LEVY-MISES.
- Hypothèse 3 : Le matériau est insensible à la vitesse de déformation.
- Hypothèse 4 : La courbure relative du profil est modérée.
- Hypothèse 5 : La géométrie de l'éprouvette est axisymétrique.

BRIDGMAN s'est intéressé à la distribution des contraintes locales dans le plan médian de l'éprouvette, c'est à dire dans la section de plus faible surface, compte tenu de ces hypothèses. L'analyse montre que, dans ce plan, les axes cylindriques x, θ et r représentent les directions principales des contraintes et que les composantes σ_{rr} et $\sigma_{\theta\theta}$ sont égales et positives. Elles sont maximales près de l'axe et décroissent jusqu'à une valeur nulle au voisinage de la surface libre.

D'après les lois de LEVY-MISES, la contrainte effective (ou généralisée) s'écrit en chaque point de la section par la relation :

$$\sigma_{eff} = \frac{1}{\sqrt{2}}\left[(\sigma_{xx}-\sigma_{rr})^2+(\sigma_{xx}-\sigma_{\theta\theta})^2+(\sigma_{rr}-\sigma_{\theta\theta})^2\right]^{1/2} \qquad \textbf{(1.1)}$$

BRIDGMAN écrit que la contrainte effective est constante dans toute la section, d'après l'hypothèse 1, et que la valeur moyenne de la contrainte axiale supporte la charge $F = \overline{\sigma}_{xx}\,\pi R^2$. Par conséquent, il trouve :

$$F_T = \frac{1}{\left[1+\frac{2Rc}{R}\right]\log\left[1+\frac{R}{2Rc}\right]} \qquad \textbf{(1.2)}$$

On remarque que cette expression ne dépend pas des paramètres mécaniques du matériau mais uniquement du facteur géométrique R/Rc. Elle est donc applicable de manière très commode et a été fréquemment utilisée dans la littérature pour corriger la contrainte de traction des métaux après l'apparition de la striction [NEEDLEMAN, 1972 ; SEMIATIN et al., 1985]

CALCUL DANS LE CAS GENERAL

Le calcul de BRIDGMAN n'est valide, de façon stricte que dans le centre d'une striction diffuse, et pour un matériau parfaitement plastique et isotrope. Cependant, il a donné d'excellents résultats dans de nombreux travaux précédents où aucune restriction n'a semblé prohibitive [ALY HELAL, 1982 ; G'SELL et al., 1985 ; SOUAHI, 1986], malgré les hypothèses très simples qui le sous- tendent. Dans ces travaux, la validité du facteur de triaxialité calculé théoriquement a été vérifiée de manière très satisfaisante dans toute la gamme des courbures choisies. A cet effet, on reprend en détail le calcul de BRIDGMAN en l'appliquant pour la striction. On avait représenté schématiquement l'allure d'une instabilité lors d'un étirage (figure1.2)

Comme le montre cette figure, on suppose que le profil de l'éprouvette, dans la section médiane, est un arc de cercle ainsi que les lignes d'écoulement des contraintes principales. Pour des considérations de symétrie autour de l'axe de l'éprouvette, on avait opté pour un système de coordonnées cylindriques r, θ et x. De plus, les grandeurs sont exprimées en valeurs algébriques car, certaines variables changent de signe lorsqu'on passe d'une courbure concave à une courbure convexe. Par exemple, le rayon de courbure Rc est positif pour un profil concave et négatif pour un profil convexe.

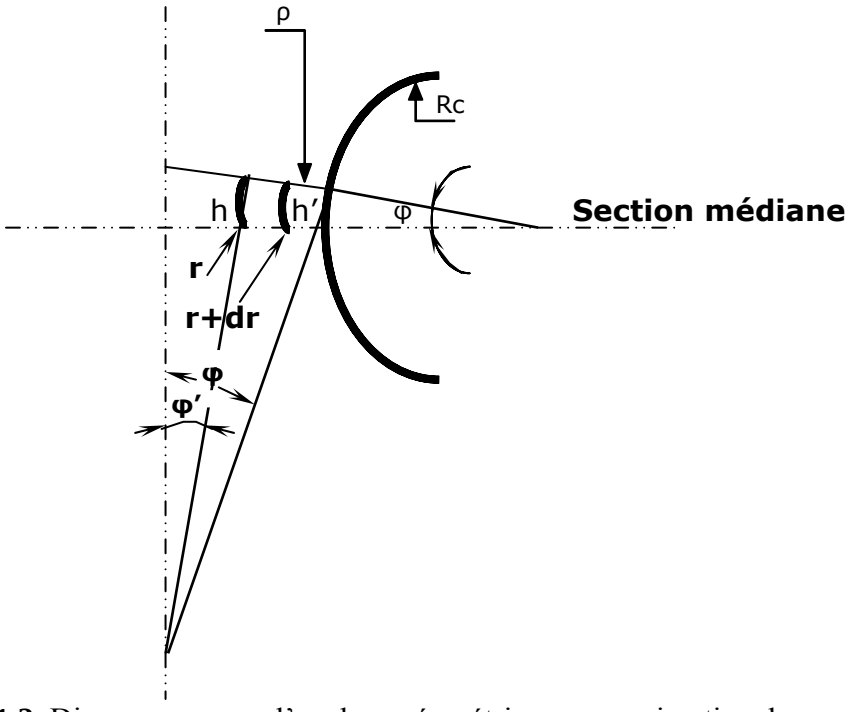


Figure1.2. Diagramme pour l'analyse géométrique approximative des conditions sur les contraintes dans le voisinage d'une instabilité plastique : striction.

Dans cette méthode, on va d'abord faire des calculs géométriques simples. Considérons un élément infinitésimal de surface dont nous avons représenté la moitié, par souci de clarté, dans la figure1.2. Cet élément dont le centre est situé dans la section médiane de l'éprouvette, est délimité par les lignes d'écoulement des contraintes principales. Il y a d'abord deux cercles, concentriques avec le profil de l'instabilité plastique, situés à une distance r et r+dr de l'axe et, ensuite, deux autres cercles qui sont situés de part et d'autre de la section médiane. Ces derniers sont centrés sur l'axe et sont normaux à la fois au profil et à l'axe de l'éprouvette. Ils ont pour rayon δ et sous-tendent un angle φ. Par conséquent, on a $\delta = R/\varphi$. Si l'on considère un point distant de r par rapport à l'axe, il sous-tend un angle φ' avec l'axe de telle façon que : $\varphi' = \varphi (r/R)$. On peut alors déduire les demi-hauteurs h et h' :

$$\begin{cases} h = Rc\,\varphi + \delta\,(\cos\varphi' - \cos\varphi) \\ h' = Rc\,\varphi + \delta\,(\cos(\varphi' + d\varphi') - \cos\varphi) \end{cases} \qquad \textbf{(1.3)}$$

Où :

$d\varphi'$ est l'incrémentation infinitésimale de l'angle φ' lorsque l'on passe de r à r+dr.

On va maintenant faire un bilan des contraintes. Lorsqu'on écrit les équations d'équilibre des contraintes, nous remarquons qu'elles se simplifient puisque toutes les dérivées par rapport à θ s'annulent, à cause de la symétrie axiale. Pour la même raison, les composantes $\sigma_{r\theta}$ et $\sigma_{x\theta}$ sont nulles. Des trois équations d'équilibre ne subsistent donc que ces deux relations :

$$\begin{cases} \dfrac{\partial \sigma_{rr}}{\partial r} + \dfrac{\partial \sigma_{rx}}{\partial x} + \dfrac{\sigma_{rr} - \sigma_{\theta\theta}}{r} = 0 \\ \dfrac{\partial \sigma_{rx}}{\partial r} + \dfrac{\partial \sigma_{xx}}{\partial x} + \dfrac{\sigma_{rx}}{r} = 0 \end{cases} \qquad \textbf{(1.4)}$$

A la surface de l'éprouvette qui est libre de contraintes, les conditions aux limites s'écrivent :

$$\begin{cases} \sigma_{rr}\cos\alpha - \sigma_{rx}\sin\alpha = 0 \\ \sigma_{rx}\cos\alpha - \sigma_{xx}\sin\alpha = 0 \end{cases} \qquad \textbf{(1.5)}$$

Où :

α représente l'angle entre la génératrice du profil et l'axe de l'éprouvette (figure1.2 et 3) On simplifie ces deux relations en écrivant σ_{rr} et σ_{rx} en fonction de σ_{xx} ; On obtient :

$$\begin{cases} \sigma_{rr} = \sigma_{xx}\, tg^2\alpha \\ \sigma_{rx} = \sigma_{xx}\, tg\ \alpha \end{cases} \qquad \textbf{(1.6)}$$

A ces deux équations, il faut ajouter la condition imposée par la force résultante F :

$$F = 2\pi \int_0^R r\, \sigma_{xx}\, dr = \text{constante} \qquad \textbf{(1.7)}$$

De plus, dans la section médiane, on a quatre conditions. Les deux premières sont issues de considérations de symétrie lorsqu'on passe de +x à –x. La troisième équation stipule que la surface est libre. Quant à la quatrième relation, elle se déduit de la première équation d'équilibre énoncée plus haut, dans le but d'éviter des infinités sur l'axe x. Ces conditions s'écrivent :

$$\begin{cases} \dfrac{\partial \sigma_{xx}}{\partial x} = 0 \qquad \text{quelque soit r} \\ \sigma_{rx} = 0 \qquad \text{quelque soit r} \\ \sigma_{rr}(r=R) = 0 \\ \sigma_{rr}(r=0) = \sigma_{\theta\theta}(r=0) \end{cases} \qquad \textbf{(1.8)}$$

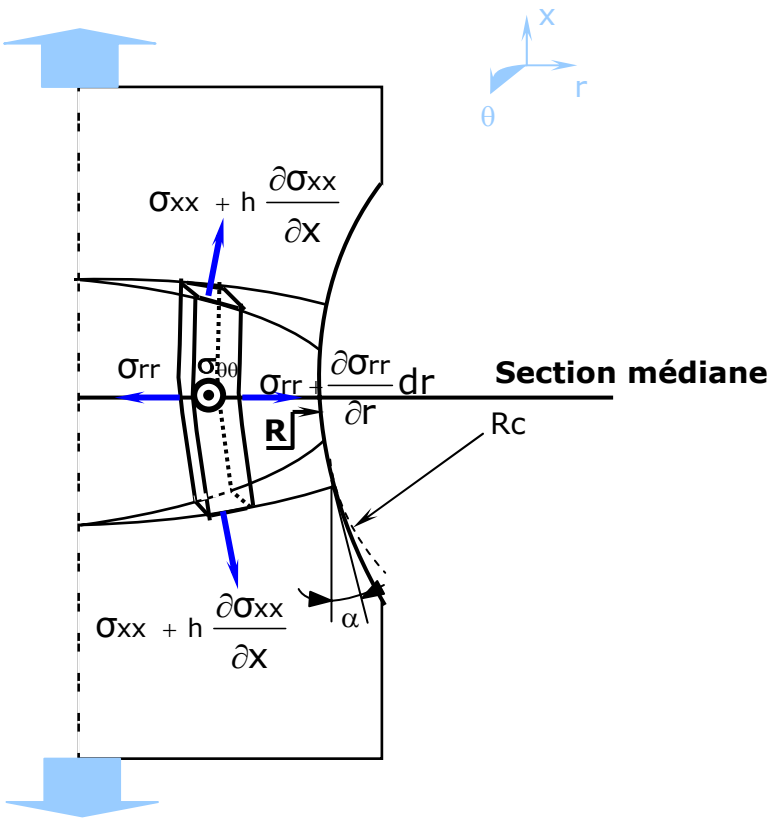


Figure1.3. Equilibre des contraintes dans le plan médian d'une éprouvette soumise à un étirage.

La grandeur, pour laquelle r=R, est estimée à la surface de l'échantillon, alors qu'elle est calculée au centre de l'axe, lorsque r=0. On remarque que la composante orthogonale σ_{rx} est nulle, dans la section médiane, au niveau du profil de l'éprouvette. On va donc éliminer σ_{rx} des équations d'équilibre, énoncées plus haut. Il ne resterait plus que les trois composantes orthogonales σ_{rr}, $\sigma_{\theta\theta}$ et σ_{xx} reliées entre elles par une seule équation :

$$\frac{\partial(r\sigma_{rr})}{\partial r} = \sigma_{\theta\theta} + \int_0^R r\frac{\partial^2\sigma_{xx}}{\partial x^2} \qquad \textbf{(1.9)}$$

Cette condition ne suffit pas pour déterminer une solution. On doit donc considérer les lois de comportement plastique du matériau qui obéit au critère de VON MISES (hypothèse 2) :

$$(\sigma_{xx} - \sigma_{rr})^2 + (\sigma_{xx} - \sigma_{\theta\theta})^2 + (\sigma_{rr} - \sigma_{\theta\theta})^2 = C \qquad \textbf{(1.10)}$$

Où :

C est une fonction qui dépend du matériau. S'il est idéalement plastique, C est une constante. Cette supposition donne habituellement une première approximation. Cependant, les matériaux polymères pré&sentent, en général, un durcissement aux fortes déformations : C dépend donc des déformations présentes. Cette dépendance est très complexe et n'a pu être déterminée complètement, quel que soit le matériau considéré. On est donc obligé de faire une seconde approximation. Dans cette nouvelle approche, on reconnaisse la dépendance de C par rapport aux déformations mais on supposera que celles-ci varient de manière très progressive dans une section donnée, au niveau de la striction pour une traction. Par conséquent, C est supposée indépendante de r dans toute la section.

De plus, on peut aussi écrire les conditions sur les déformations qui doivent être satisfaites dans le domaine plastique. Lorsque l'on suppose une isotropie du durcissement du matériau (hypothèse 1), elles deviennent :

$$\left\{\begin{array}{l} e_r = \beta\,(\sigma_{rr} - 1/2\,(\sigma_{\theta\theta} + \sigma_{xx})) \\ e_\theta = \beta\,(\sigma_{\theta\theta} - 1/2\,(\sigma_{xx} + \sigma_{rr})) \\ e_x = \beta\,(\sigma_{xx} - 1/2\,(\sigma_{rr} + \sigma_{\theta\theta})) \end{array}\right\} \qquad \textbf{(1.11)}$$

Où :

β représente une fonction qui dépend des déformations et que l'on détermine expérimentalement. Ces déformations sont sujettes à une condition supplémentaire puisque e_x est supposée constante le long de la section comme on l'a énoncé précédemment. De plus, on doit tenir compte de la condition de conservation volumique. Pour un déplacement radial ρ, elle s'écrit sous ces conditions : $\rho = -e_x\, r/2 + c/r$. c est une constante. Pour éviter une infinité dans les axes, c est égale à 0. Il vient alors : $\rho = -e_x\, r/2$. On a finalement :

$$e_r = e_\theta = -1/2\ e_x \qquad \textbf{(1.12)}$$

Si l'on pose $\sigma_{rr} = \sigma_{\theta\theta}$, on peut satisfaire la condition sur e_r et e_θ. Encore, le critère de plasticité de VON MISES est vérifié lorsqu'on écrit : $\sigma_{xx} = \sigma_{xx}\ (r=R) + \sigma_{rr}$. Ce système de contraintes se décrit simplement. C'est une tension longitudinale, toujours uniforme, dans la section médiane avec une tension hydrostatique superposée ($\sigma_{rr} = \sigma_{xx} - \sigma_{xx}\ (r=R) = \sigma_{\theta\theta}$). Cette dernière diminue lorsqu'on s'approche de la surfacre où elle s'annule ($\sigma_{rr}(r=R) = 0$) et à son maximum dans l'axe. Pour compléter la détermination de la solution, on a une relation entre $\partial^2\sigma_{xx}/\partial x^2$ et σ_{rr}, déjà explicitée précédemment. A cause des conditions qu'on a posées auparavant, on a deux équations qui concernent les deux grandeurs $\partial^2\sigma_{xx}/\partial x^2$ et σ_{rr}. Il n'y a pas de méthode relativement simple qui permettent de résoudre ces équations de manière rigoureuse. Pour cela, on se contente d'une méthode de résolution approximative qui n'est pas valide qu'au voisinage immédiat de l'instabilité plastique. Cependant, ceci est acceptable puisque on s'intéresse au comportement intrinsèque du matériau (au voisinage de la striction).

Considérons maintenant un élément de matière ayant pour épaisseur dθ et pour surface l'élément décrit précédemment (figure3). Après avoir fait le bilan des forces, on écrit les équations d 'équilibres :

$$\begin{aligned}&\left(\sigma_{xx} + h\frac{\partial\sigma_{xx}}{\partial x}\right)\sin\varphi\left(r + \frac{dr}{2}\right)dr = \\ &\sigma_{rr}\ hr - \left(\sigma_{rr} + \frac{\partial\sigma_{rr}}{\partial r}\ dr\right)(r+dr)h' + \sigma_{\theta\theta}\, h\frac{\sin\theta}{\theta}\ dr\end{aligned} \qquad \textbf{(1.13)}$$

Lorsqu'on ne tient compte que des termes de plus bas ordre, on trouve après simplification :

$$\sigma_{xx}\frac{r^2}{R} = \sigma_{rr}\ \left(\frac{3r^2}{2R} - \frac{R}{2} - Rc\right) - \left(r\frac{d\sigma_{rr}}{dr} - \sigma_{\theta\theta}\right)\left(Rc + \frac{R^2 - r^2}{2R}\right) \qquad \textbf{(1.14)}$$

Comme on l'a vu précédemment, on a : $\sigma_{\theta\theta}=\sigma_{rr}$ et $\sigma_{xx}=\sigma_{rr}+\sigma_{xx}(r=R)$ où $\sigma_{xx}(r=R)$ est la contrainte longitudinale sur la surface externe pendant que les deux autres composantes $\sigma_{\theta\theta}$ et σ_{rr} sont nulles. σ_{xx} (r=R) est la contrainte d'écoulement lors d'un essai purement uniaxial. On la remplace par σ_{eff} et obtient alors une équation comprenant une seule composante σ_{rr} :

$$\frac{d\sigma_{rr}}{dr}\left(Rc+\frac{R^2-r^2}{2R}\right)+\frac{r}{R}\,\sigma_{eff}=0 \qquad \textbf{(1.15)}$$

Afin d'intégrer cette équation, on sépare les variables en tenant compte des conditions aux limites déjà énumérées précédemment.

On a par conséquent :

$$\begin{cases}\sigma_{rr}=\sigma_{eff}\,\log\dfrac{R^2+2RRc-r^2}{2RRc}\\ \sigma_{xx}=\sigma_{eff}\left(1+\log\dfrac{R^2+2RRc-r^2}{2RRc}\right)\end{cases} \qquad \textbf{(1.16)}$$

On déduit alors l'expression de la force F en intégrant σ_{xx} sur toute la section :

$$F=2\pi\int_0^R r\;\sigma_{xx}\,dr=\pi\,\sigma_{eff}\left(R^2+2RRc\right)\log(1+R/Rc) \qquad \textbf{(1.17)}$$

En sachant que $\sigma_{xx}=F/\pi R^2$, cette expréssion s'écrit aussi :

$$\sigma_{xx}=\frac{F}{\pi R^2}=\sigma_{eff}\left(1+\frac{2Rc}{R}\right)\log\left(1+\frac{R}{2Rc}\right) \qquad \textbf{(1.18)}$$

De cette dernière relation, on a le facteur de triaxialité F_T. Il est supérieur à 1 pour un profil convexe et inférieur à 1 pour un profil concave.

$$F_T=\frac{1}{\left(1+\dfrac{2Rc}{R}\right)\log\left(1+\dfrac{R}{2Rc}\right)} \qquad \textbf{(1.19)}$$

CONCLUSION

On a donc parvenu à trouver une solution qui satisfait les équations d'équilibre, les conditions aux limites et la loi de plasticité de VON MISES. Mathématiquement, ces conditions déterminent une solution unique. On a abouti à cette solution en supposant d'une part que les contraintes sont constantes dans toute la section et, d'autre part, que les contraintes sur la surface sont nulles.

REFERENCES

ALY HELAL N.A.
Etude et simulation du développement de la striction dans le polyéthylène à haute densité ;
(1982) Thèse de docteur-ingénieur, I.N.P.L., Nancy.

BRIDGMAN P.W.
The stress distribution at the neck of a tension specimen;
(1944) Proc. Am. Soc. Metals, 32, 553.

BRIDGMAN P.W.
Studies in large plastique flow and fracture;
(1964) Harvard University Press, Cambridge, Massachusetts.

DAVIDENKOV N.N. et SPIRIDONOVA N.I.
Mechanical methods of testing- Analysis of the state of stress in the neck of a tension test specimen;
(1946) Proc. Am. Soc. Test. Mat., 46, 1147.

DIETER G.E.
Mechanical metallurgy;
(1961) Mc Graw Hill, New York.

ESTRIN Y., LING C.P. et McCORMICK P.G.
Localization of plastic flow: spatial vs temporal instabilities;
(1991) Acta Metall. Mater., 39 (11), 2943.

G'SELL C., MARQUEZ-LUCERO A., SOUAHI A. et TONG Y.S.
Determination of stress triaxiality during the necking of high-density polyethylene ;
(1985) Instabilités Plastiques, Presses des Ponts et Chaussées, Paris, 165-174.

HAUDIN J.M., COTTO D. et MONTHEILLET F.
Application of SWIFT criterion to the theoretical study of biaxial stretching instability in polymeric films;
(1984) C. R. acad. Sc., Paris, 298 (2), 511.

HUTCHINSON J. W. et NEALE K. W.
Neck propagation ;
(1983) J. Mech. Phys. Solids, 31, 405.

MILLS N. J.
Theory of neck development in tensile tests on thermoplastics;
(1978) Bri. Polym. J., 10 (3),1.

NEEDLEMAN A.
A numerical study of necking in circular cylindrical bars;
(1972) J. Mech. Phys. Solids, 20, 111.

SEMIATIN S.L., GHOSH A. K. et JONAS J. J.
A simplified numerical analysis of the sheet tensile test;
(1985) Metall. Trans. A.

SIEBEL E.
(1925) Werkstoffaus, 71.

SOUAHI A.
Effet de la triaxialité des contraintes pendant la striction diffuse en traction dans le polyéthylène à haute densité ;
(1986) Thèse de Magister, Université d'Annaba.

Annexe2

ALGORITHME DE RETRO-PROPAGATION DU GRADIENT (*RP*)

Le principe de l'algorithme est, comme dans le cas du perceptron linéaire, de minimiser une fonction de coût ou d'erreur. Il s'agit en suite de calculer la contribution à cette erreur de chacun des poids synaptiques. C'est cette étape qui est difficile.

En effet, chacun des poids influe sur le neurone correspondant, mais la modification pour ce neurone va influer sur tous les neurones des couches suivantes. Ce problème est parfois désigné sous le nom de *Credit Assignment Problem.*

Cependant, la popularisation de la *RP* et son développement restent liés aux travaux du groupe PDP dirigé par Rumelhart.

Pour introduire l'algorithme de rétropropagation du gradient, on va l'appliquer sur un perceptron multicouches (*MLP*) avec une fonction d'activation sigmoïde :

$$f(x) = \frac{e^x}{e^x + 1} \qquad \textbf{(2.1)}$$

La fonction σ (x) est indéfiniment dérivable et dont la dérivée est simple à calculer

$$f(x) = \frac{e^x}{\left(e^x + 1\right)^2} = f(x)\,(1 - f(x)) \qquad \textbf{(2.2)}$$

1 INTRODUCTION DE L'ALGORITHME

On considère un réseau à trois couches illustré par la figure2.1

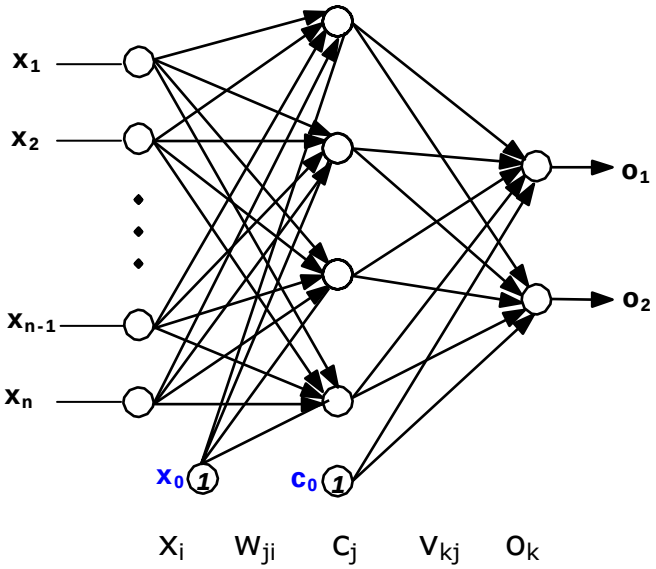


Figure2.1. Définition des notations pour un MLP

Les conventions de notation sont les suivantes :

- o_k : Activation de la $k^{ème}$ unité de sortie, k = 1, …, M;
- t_k : Activation désirée de la $k^{ème}$ unité de sortie ;
- c_j : Activation de la $j^{ème}$ unité cachée, j=0, 1, …, n_h ; c_0=1 : Entrée du biais pour la couche de sortie ;
- x_i : $i^{ème}$ entrée externe du réseau ; i=0, 1, …, n; x_0=1 : Entrée du biais pour la couche cachée;
- w_{ji} : Poids d'une connexion entre la $i^{ème}$ entrée et la $j^{ème}$ unité cachée;
- v_{kj} : Poids d'une connexion entre la $j^{ème}$ unité cachée et la $k^{ème}$ unité de sortie.

Les indices i, j et k font référence aux unités d'entrée, aux unités cachées et aux unités de sortie, respectivement. L'exposant p correspond au numéro de l'exemple présenté à l'entrée du réseau : p = 1,..., n_A, où n_A est le nombre d'exemples d'apprentissage. Le $p^{ème}$ exemple est noté $X^p = \left[X_0^p \ ,\ldots, X_i^p \ ,\ldots, X_n^p \right]$ et la $i^{ème}$ composante X_i^p désigne la $i^{ème}$ entrée lorsque le $p^{ème}$ exemple est présenté au réseau. Pour un exemple p, la $j^{ème}$ unité cachée a l'entrée résultante I_j^p :

$$I_j^p = \sum_{i=0}^{n} w_{ji} x_i^p \qquad \textbf{(2.3)}$$

Et une activation c_j^p :

$$c_j^p = f\left(I_j^p\right) = f\left(\sum_{i=0}^{n} w_{ji}\, x_i^p\right) \qquad \textbf{(2.4)}$$

Où :

f est la fonction d'activation. La k^{ème} unité de sortie reçoit une entrée résultante I_k^p définie par :

$$I_k^p = \sum_{i=0}^{n} v_{kj} c_j^p \qquad \textbf{(2.5)}$$

Et génère en sortie l'activation o_k^p :

$$o_k^p = f\left(I_k^p\right) \qquad \textbf{(2.6)}$$

Comme la fonction d'activation des neurones de sorties pour un PMC est linéaire, l'équation (2.6) devient :

$$o_k^p = I_k^p \qquad \textbf{(2.7)}$$

La fonction de coût usuelle est l'erreur quadratique moyenne définie par :

$$E(w) = \frac{1}{2} \sum_{k,p} \left(t_k^p - o_k^p\right)^2 \qquad \textbf{(2.8)}$$

Où :

w est le vecteur contenant tous les poids du réseau. La fonction E (**w**) est continue et différentiable par rapport à chaque poids. Pour déterminer les poids qui la minimisent, on peut donc utiliser l'algorithme de descente du gradient. Pour faciliter la notation, E(**w**) sera notée *E* dans ce qui suit [Moussaoui et al., 1999]

Pour les poids des connexions des unités cachées vers les unités de sortie, le terme d'adaptation des poids au cours de l'apprentissage est défini par :

$$\Delta v_{kj} = -\eta \frac{\partial E}{\partial v_{kj}} = \eta \sum_p \delta_k^p c_j^p \qquad \textbf{(2.9)}$$

Avec :

$$\delta_k^p = \left(t_k^p - o_k^p\right) \qquad \textbf{(2.10)}$$

Dans le cas de sorties linéaires.

Pour les poids des connexions entre la couche d'entrée et la couche cachée, le terme d'adaptation des poids est :

$$\begin{aligned}\Delta v_{kj} &= -\eta \frac{\partial E}{\partial w_{ji}} = -\eta \sum_p \frac{\partial E}{\partial c_j^p} \frac{\partial c_j^p}{\partial w_{ji}} \\ &= \eta \sum_{k,p} \left(t_k^p - o_k^p\right) v_{kj} f'(I_j^p) x_i^p = \eta \sum_p \delta_j^p x_i^p\end{aligned} \qquad \textbf{(2.11)}$$

Avec :

$$\delta_j^p = f'\left(I_j^p\right) \sum_k v_{kj} \delta_k^p \qquad \textbf{(2.12)}$$

On peut constater que les équations (2.9) et (2.11) ont la même forme, et ne diffèrent que par la définition de la quantité δ. Ces formules se généralisent facilement aux cas des réseaux possédant un nombre quelconque de couches cachées.

D'après l'équation (2.12), le calcul de δ_j pour une unité cachée j nécessite les δ_k des unités de sortie, qui sont fonctions des erreurs en sortie du réseau (t_k - o_k). Ainsi, pour corriger les poids des connexions entre la couche d'entrée et la couche cachée, on a besoin de rétropropager l'erreur depuis les sorties vers les entrées, d'où le nom de l'algorithme d'apprentissage : *rétro-propagation de l'erreur*.

D'après l'équation (2.12), le calcul de δ_j pour une unité cachée j utilise la dérivée de l'activation de cette même unité et la somme des δ_k pour toutes les unités de sortie. Le calcul de δ_j se fait donc de manière locale (indépendamment de toutes les autres unités cachées). Ceci permet d'envisager une *parallélisation* de l'algorithme de rétro-propagation *(RP).*

Après avoir calculé le terme d'adaptation des poids, la mise à jour se fait selon la formule suivante :

$$w_{ji}(t+1) = w_{ji}(t) + \Delta w_{ji} \quad \textbf{(2.13)}$$

ET

$$v_{kj}(t+1) = v_{kj}(t) + \Delta v_{kj} \quad \textbf{(2.14)}$$

Où :
t est l'indice de l'itération. Le mode d'adaptation des poids tel qu'il est présenté par les équations (2.9) et (2.11) s'appelle mode *batch*. La mise à jour des poids se fait après avoir passé en revue tous les exemples d'apprentissage. Ce mode d'apprentissage est encore appelé déterministe hors ligne *off-line* ou *by epoch.*

Une autre approche consiste à modifier les poids après chaque présentation d'un exemple d'apprentissage. C'est l'apprentissage en mode en ligne *on-line* ou *by pattern*.

Il y a des avantages et des inconvénients pour chaque mode d'adaptation des poids, et le choix de l'un ou de l'autre dépend du problème à traiter.

Les algorithmes *off-line* sont faciles à analyser pour ce qui concerne les propriétés de convergence. Ils peuvent utiliser un taux d'apprentissage optimum à chaque itération et peuvent conduire à des solutions assez précises (avec de faibles variantes). En revanche, ils ont l'inconvénient d'induire un temps de calcul du terme d'adaptation de poids dépendant de la taille de l'ensemble d'apprentissage.

Les méthodes *on-line* peuvent être utilisées lorsque les exemples ne sont pas tous disponibles au début de l'apprentissage, et quand on désire réaliser une adaptation continue à partir d'une suite de couples *entrée-sortie* issus d'une relation qu'on cherche à identifier.

2 VARIANTES DE L'ALGORITHME DE *RP*

Depuis son introduction, l'algorithme de *RP* a été largement étudié et plusieurs modifications y ont été apportées. L'algorithme de base décrit ci-dessus converge très lentement pour les réseaux multicouches.

Les variations apportées à l'algorithme de *RP* ont pour objectifs l'accélération de la convergence du processus d'apprentissage et l'amélioration de la capacité de généralisation. Nous présentons ci-dessous quelques variantes parmi les plus importantes.

2.1 Choix de la fonction coût

Le choix de la fonction d'erreur utilisée pour l'apprentissage des réseaux de neurones multicouches a une certaine influence sur la rapidité d'apprentissage et sur la qualité de généralisation du réseau. Cette question a été étudiée par plusieurs chercheurs [Gayner et Downs, 1994]. Le critère d'erreur le plus utilisé est la fonction d'erreur quadratique moyenne (cf. équation (2.8)).

Cette fonction a tendance à amplifier les erreurs les plus importantes. Par conséquent, au cours de l'apprentissage, la mise à jour des poids est largement déterminée par la correction des grandes erreurs, ce qui est recherché en général [Hassibi et Stork, 1993].

Cependant le choix de fonction quadratique n'est pas la seule possibilité. On peut remplacer $\left(t_k^p - o_k^p\right)^2$ par toute autre fonction $\varepsilon\left(t_k^p, o_k^p\right)$ différentiable et minimale lorsque ses deux arguments sont égaux. Le développement précédent montre que seule l'expression de l'équation (2.9) dépend de la fonction coût. Le reste de l'algorithme de *RP* reste inchangé.

2.2 Introduction d'un terme de moment

Le paramètre η (appelé taux d'apprentissage) joue un rôle important. S'il est trop faible, la convergence est lente, s'il est trop grand, l'algorithme oscille entre des points différents à cause de l'existence de vallées et de plateaux à la surface de la fonction coût.

Pour stabiliser la recherche des poids optimisant la fonction coût, une méthode consiste à ajouter un terme dit de *moment* à l'expression d'adaptation des poids [Plaut et al., 1990]. L'idée est de donner une certaine *inertie* pour chaque poids, de sorte que sa mise à jour ne se fasse pas de manière brutale.

Ceci permet alors d'utiliser un taux d'apprentissage relativement important sans pour autant augmenter les oscillations de la trajectoire sur la surface d'erreur.

La nouvelle formule d'adaptation des poids est définie par :

$$\Delta w(t+1) = -\eta \frac{\partial E}{\partial w} + \alpha \Delta w(t) \qquad \textbf{(2.15)}$$

Où :

α est le terme de moment dont la valeur est souvent prise proche de 1 (~0.9). Cette méthode peut être utilisée en modes *off-line* ou *on-line*.

2.3 Taux d'apprentissage adaptatifs

Les paramètres η et α de l'algorithme *RP* ne sont pas faciles à fixer a priori pour un problème donné. En outre, des valeurs données peuvent être bien adaptées au début de l'apprentissage, sans l'être nécessairement au milieu ou à la fin du processus. Pour résoudre ce problème, certains auteurs ont proposé d'ajuster automatiquement ces paramètres au cours de l'apprentissage.

Plusieurs heuristiques d'adaptation dynamique du taux d'apprentissage ou du terme de moment ont été proposées [Jacobs, 1988], [Silva et Almeida, 1990], [Tollenaere, 1990]. On se contente ici de rappeler le principe des méthodes adaptatives et de citer l'algorithme de Silva et Almeida comme exemple.

Le principe des méthodes adaptatives est d'étudier à chaque itération l'effet de la mise à jour des poids des connexions sur la fonction coût. Si cette fonction augmente, alors le taux d'apprentissage η_{pq} associé au poids w_{pq} doit être réduit. En revanche, si la fonction coût décroît régulièrement suite aux modifications des poids, alors on peut augmenter η_{pq} pour accélérer la recherche de l'optimum. Différents critères peuvent être retenus, comme par exemple la diminution de la fonction coût pendant plusieurs itérations successives. En partant de ce principe, Silva et Almeida ont proposé la règle suivante :

$$\eta_{pq}(t) = \begin{cases} \gamma \eta_{pq}(t-1), & \text{si } \dfrac{\partial E}{\partial w_{pq}(t)} \dfrac{\partial E}{\partial w_{pq}(t-1)} > 0 \\ \zeta \eta_{pq}(t-1) & \text{sinon} \end{cases} \qquad \textbf{(2.16)}$$

Où :

w_{pq} est un poids quelconque du réseau, γ et ζ sont des paramètres respectivement supérieur et inférieur à 1 (par exemple : γ = 1.2 et ζ = 0.7). Les auteurs montrent expérimentalement que les

valeurs précises de ces deux paramètres ne sont pas décisives. Ils proposent également d'utiliser un terme de moment identique et non adaptatif pour tous les poids du réseau. La procédure d'adaptation des poids est alors la suivante :

$$\begin{cases} w_{pq}(t+1) = w_{pq}(t) - \eta_{pq}(t)\ z_{pq}(t) \\ \text{et} \\ z_{pq}(t) = \dfrac{\partial E}{\partial w_{pq}(t)} + \alpha\ z_{pq}(t-1) \end{cases} \quad \textbf{(2.17)}$$

2.4 Autres procédures d'optimisation

L'algorithme de *RP* dans sa forme de base utilise la technique de descente du gradient. Celle-ci est parmi les plus simples, mais elle n'est pas très efficace car elle utilise peu d'information sur la surface de l'erreur. Dans la littérature, on trouve une grande quantité de techniques plus sophistiquées [Fletcher, 1987]. Une synthèse des principales méthodes est présentée dans [Hertz et al., 1991].

En effectuant le développement au second ordre de la fonction coût E (**w**) autour du point courant $\mathbf{w_0}$ et en négligeant les termes d'ordre supérieur, on peut écrire :

$$E(w) = E(w_0) + (w - w_0)^T \nabla E(w_0) + \frac{1}{2}(w - w_0)^T H(w - w_0) \quad \textbf{(2.18)}$$

Avec **H** la matrice Hessienne calculée au point $\mathbf{w_0}$. En dérivant l'équation (2.18), on obtient :

$$\nabla E(w) = \nabla E(w_0) + H(w - w_0) \quad \textbf{(2.19)}$$

En annulant la dérivée, l'équation (2.19) donne une estimation de la localisation du minimum :

$$w = w_0 - H^{-1} \nabla E(w_0) \quad \textbf{(2.20)}$$

En posant $\mathbf{w_0}$ = **w** (t) et **w** = **w** (t+1), on obtient une procédure itérative d'estimation du vecteur des poids minimisant la fonction coût :

$$w(t+1) = w(t) - H^{-1} \nabla E\left(w(t)\right) \qquad \textbf{(2.21)}$$

Cette méthode est appelée *méthode de Newton*. Elle permet une convergence rapide chaque fois que les poids se trouvent au voisinage de la solution. Son inconvénient est la nécessité de calculer $\mathbf{H}^{-1}$.

Deux solutions ont été proposées face à ce problème. L'une consiste à négliger les termes non diagonaux de **H** (*méthode pseudo-Newton*) [Le Cun et al., 1989], l'autre à estimer itérativement la matrice $\mathbf{H}^{-1}$ (*méthode quasi-Newton*)

Les méthodes de minimisation présentées jusqu'ici se basent sur l'utilisation de dérivées de la fonction coût. Une autre approche consiste à utiliser des procédures de *minimisation* stochastique [Brunelli R., 1994] ou les *algorithmes génétiques* [Goldberg, 1989]. Ces algorithmes effectuent une recherche globale et présentent par conséquent moins de risques d'aboutir à des minima locaux assez superficiels. Cependant, ils nécessitent des calculs plus lourds.

3 CHOIX DES SEQUENCES D'APPRENTISSAGE

Ce paragraphe donne quelques indications sur le choix des séquences d'apprentissage.

3.1 Séquence des entrées de commande

3.1.1 Contraintes sur les entrées de commande

Elles portent sur l'amplitude et le type de signaux de commande que le processus est susceptible de recevoir pendant son fonctionnement. Les amplitudes maximales sont en général faciles à déterminer, car leur ordre de grandeur correspond aux valeurs de saturation des actionneurs, qui peuvent être estimées physiquement (puissance maximale que peut délivrer un moteur, pression maximale d'un circuit de freinage ou d'un mécanisme hydraulique, etc.).

En ce qui concerne le type de signaux à utiliser, un principe général est que les signaux utilisés pour l'identification doivent être de même nature que ceux qui seront calculés par l'organe de commande pendant l'utilisation du processus. L'idéal serait d'utiliser pour l'identification le correcteur même qui sera synthétisé à l'aide du prédicteur identifié.

Si le bruit est négligeable, une bonne démarche consiste à effectuer les expériences *en asservissement*, avec un correcteur simple (par exemple, un PID). Cette démarche, ou

identification *en boucle fermée*, permet d'explorer le domaine de fonctionnement désiré, en imposant une séquence de consigne correspondant au cahier des charges.

Une autre démarche, couramment utilisée, consiste à explorer au mieux le domaine de fonctionnement, par exemple avec des créneaux de commande (riches en fréquences), d'amplitudes et de durées diverses. Notons que cette dernière solution n'est pas praticable s'il existe des contraintes sur des sorties ou des variables internes du processus, et en particulier si le processus est instable.

3.1.2 Fréquence d'échantillonnage :

Les réseaux de neurones étant des modèles non linéaires, il n'est pas possible de passer simplement d'un modèle discret, valable à une fréquence d'échantillonnage donnée, à un autre : Il est donc nécessaire d'effectuer l'identification à la fréquence qui sera utilisée pour la commande du processus.

Si des contraintes diverses, par exemple le temps de calcul de la commande, nécessitent le choix d'une fréquence d'asservissement plus basse que la fréquence utilisée pour l'identification, il faut procéder à une nouvelle identification (un nouvel apprentissage).

3.2 Le surajustement

Après cette revue des outils les plus fréquemment utilisés dans le domaine des réseaux de neurones pour sélectionner le meilleur modèle tout en évitant les solutions surajustées, les questions suivantes se posent :

- Quelle est l'origine exacte du surajustement ?
- Quelle démarche faudrait-il mettre en œuvre pour régler ce problème à la base ?

Le surajustement caractérise une fonction dont la complexité (c'est-à-dire le nombre et la nature des degrés de libertés) est telle qu'elle est capable de s'ajuster exactement aux exemples d'apprentissage, même si ceux-ci sont entachés de bruit.

Ce phénomène est donc, à l'origine, un phénomène local : Dans certains domaines des entrées, la fonction utilise localement certains de ses degrés de liberté de manière à passer précisément par certains exemples.

Cette définition du surajustement suppose que tous les exemples ont la même importance et que l'on recherche effectivement une solution dont la réponse est satisfaisante.

Généralement, cette hypothèse est formalisée dans la fonction de coût choisie. Ainsi, pour éviter le surajustement, il faudrait limiter l'influence de chaque exemple sur l'estimation des poids du modèle. Les valeurs de ceux-ci doivent être déterminées par l'ensemble de la base d'apprentissage, et non par un exemple particulier.

3.3 Problème des minima locaux

Les minima trouvés par les algorithmes d'apprentissages sont souvent des minima locaux. Le minimum trouvé dépend du point de départ de la recherche, en d'autres termes de l'initialisation des poids. En pratique, il faut effectuer plusieurs minimisations avec des initialisations différentes, pour trouver plusieurs minima et retenir le *meilleur*.

Il est néanmoins impossible et généralement inutile, de s'assurer que le minimum choisi est le minimum global [C. Bishop, 1995].

CONCLUSION

L'algorithme d'apprentissage le plus utilisé dans la communauté connexionniste est l'algorithme de rétro-propagation (*RP*) Plusieurs travaux lui ont été consacrés et un certain nombre de modifications ont été proposées pour accélérer le temps d'apprentissage et améliorer les performances du réseau obtenu.

Dans la pratique, l'objectif d'une modélisation statistique n'est pas d'ajuster finement un modèle sur un ensemble d'apprentissage, mais d'obtenir un bon compromis entre performances d'apprentissage et performances de généralisation.

Lors de la construction d'un réseau de neurones, l'objectif donc est de déterminer la meilleure architecture possible et le meilleur jeu de poids au sens d'un critère donné (la généralisation par exemple)

Sous cet angle, la construction du réseau de neurones peut être vue comme un problème de recherche d'une solution dans l'espace des architectures et des poids.

Le prix à payer pour cette approche est le temps de calcul. En général, on se contente dans la pratique d'utiliser des méthodes qui ne garantissent pas l'optimalité de la solution, mais qui ont l'avantage d'être plus rapides.

REFERENCES

BISHOP C.
Neural Networks for Pattern Recognition;
(1995) Oxford University Press.

Brunelli R.
Estimation of Pose and Illuminant Direction for Face Process;
(1994) IEEE Transactions, on Pattern Analysis and Machine Intelligence.

Fletcher R.
Practical methods of optimisation;
(1987) John Wiley and Sons Ltd, second edition, New-York, USA

Gayner R.J. et Downs T.
On the properties of error functions that affect the speed of backpropagation learning;
(1994) Marinaro M. and Morasso P. G., editors, ICANN'94, Proceeding of the International Conference on Artificial Neural Networks; Sorrento, Italy, pp.557-560.

Goldberg D.E.
Genetic algorithms in search, optimization and machine learning;
(1989) Addison Wesley, New-York, NY

Hassibi et D. G. Stork
Second order derivatives for network pruning: Optimal brain surgeon;
(1993) Advances in neural information processing systems, 5, San Mateo, CA, Morgan Kaufman Publishers Inc. pp. 164-171.

Hertz J., Krogh A., Palmer R.G.
Introduction To The Theory of Neural Computation;
(1991) Addison-Wesley.

Moussaoui A.K., Bouazza S.E. et Abbassi H.A.
Feedforward Control of Hot Strip thickness Process Using Neural Networks;
(1999) Proceedings du Séminaire National sur l'Automatique et les Signaux, SNAS'99, U.B.M. Annaba, Algérie, pp.237-241.

Plaut D.C., Nowlan S.J. et Hinton G.E.
Experiments on learning by back-propagation;
(1990) Technical Report CMU-CS-86-126, Carnegie Mellon University - Pittsburgh, PA 15213.

Annexe3

SOURCE EXPERIMENTALE DE LA BASE DES DONNEES

Ce sont les courbes contraintes vraies -Déformations vraies pour plusieurs vitesses de déformation tirées d'un travail expérimental [A.SOUAHI, 1986]

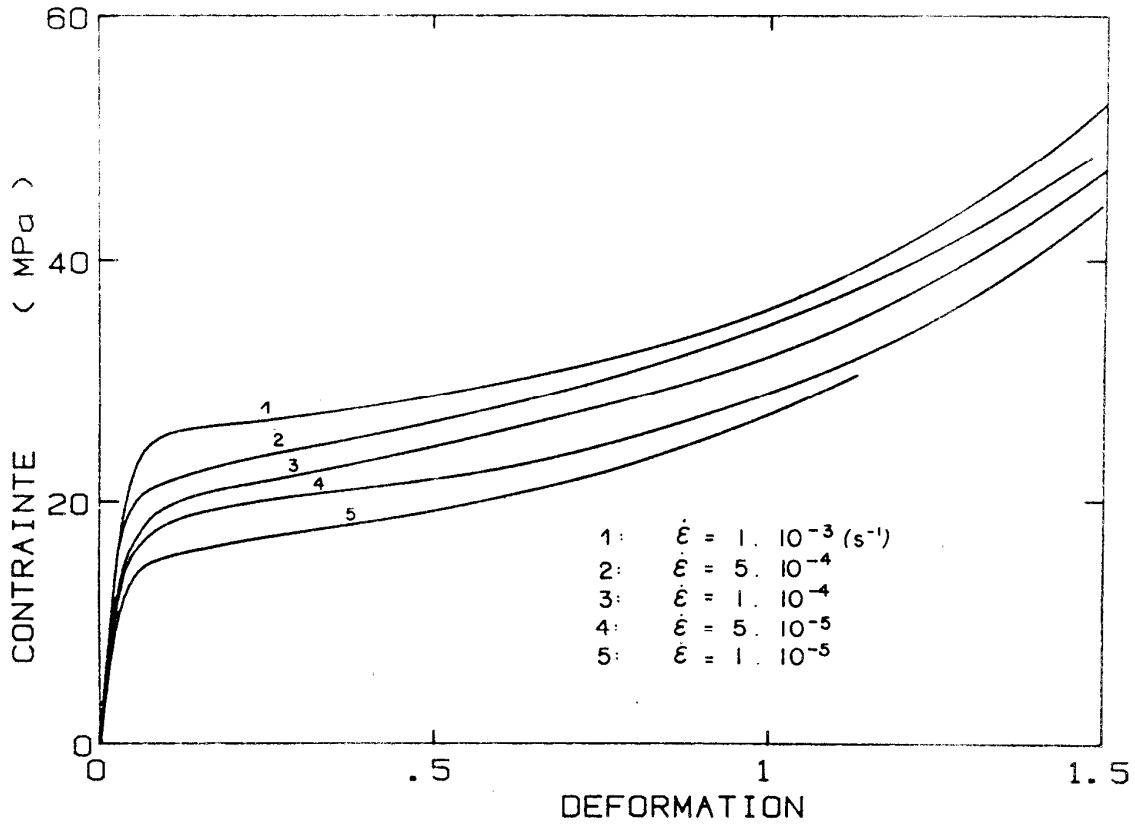


Figure3.1. Variation de la contrainte en fonction de la déformation pour des essais effectués avec des différentes vitesses de déformation [A. SOUAHI, 1986]

Nous avons déterminé la trace de ces courbes par un programme élaboré par M. BOUGHAZI (département d'électronique), exécuté en MATLAB.

REFERENCE

SOUAHI A.
Effet de la triaxialité des contraintes pendant la striction diffuse en traction dans le polyéthylène à haute densité ;
(1986) Thèse de Magister, pp. 86.