

La modélisation du battement cardiaque est une étape essentielle pour l'identification automatique des ondes caractéristiques. Elle a pour objectif de trouver une *représentation mathématique*, aussi simple et compacte que possible, de la forme de chaque onde constitutive du battement cardiaque. En effet, la représentation la plus « naturelle » des ondes consisterait à décrire le signal par son amplitude à chaque instant ; cette représentation serait donc un vecteur dans un espace dont la dimension serait égale à quelques centaines. Dans un tel espace, les étapes de traitement nécessaires à la reconnaissance des pathologies se heurteraient à des problèmes insurmontables.

L'approche proposée ici consiste à décomposer un battement en une somme de fonctions paramétrées qui permettent de localiser et de caractériser les ondes : la dimension de l'espace de représentation est alors égale au nombre de paramètres qui interviennent dans la décomposition. On peut fonder ce type de modélisation sur des algorithmes aussi divers que la décomposition en ondelettes, la régression polynomiale, l'approximation par réseaux de neurones, l'approximation par fonctions radiales de base (RBF),..., ce qui conduit chaque fois à une représentation analytique du battement considéré.

Après avoir défini précisément la principale propriété que nous attendons de la modélisation d'un battement, nous présentons brièvement quelques résultats obtenus par les méthodes classiques citées ci-dessus. Nous verrons ensuite en quoi ces méthodes ne constituent pas une modélisation satisfaisante selon nos critères, et nous présenterons enfin une décomposition entièrement originale, applicable à une grande variété de problèmes, et particulièrement bien adaptée à la modélisation des battements cardiaques.

I Objectif de la modélisation

I.1 Présentation

L'identification des ondes caractéristiques du battement est réalisable en deux étapes : la segmentation et l'étiquetage.

- La segmentation correspond au « découpage » du battement en zones susceptibles de contenir chacune une onde cardiaque ; il s'agit donc, à ce niveau, de repérer les formes qui ressemblent aux ondes cardiaques.
- L'étiquetage correspond à l'attribution d'un label médical (P, Q, R, S ou T) à chacune des zones définies lors de la segmentation. Les zones qui contiennent une onde cardiaque bien identifiée reçoivent le label médical correspondant, tandis que celles dont l'onde associée ne correspond pas à une onde cardiaque se voient attribuer l'étiquette X.

L'affectation d'un label médical à chaque forme de l'ECG est *indispensable* pour « *communiquer* » ensuite avec les cardiologues. En effet, les pathologies cardiaques sont classiquement exprimées sous forme d'anomalies des distances entre les ondes caractéristiques, ou en termes de problèmes dans la forme de ces ondes (cf. chapitres 1 et 2) : le fait de localiser précisément ces ondes permettra de fournir aux cardiologues des informations sur leurs formes et sur leurs distances mutuelles, et inversement de définir des seuils de « normalité » pour chacune des ondes.

La méthode d'étiquetage sera décrite en détail dans le chapitre suivant ; le présent chapitre est consacré à la segmentation.

Puisque l'étiquetage repose entièrement sur les résultats de la segmentation, celle-ci doit être réalisée avec beaucoup d'attention : plus la segmentation est pertinente, plus l'étiquetage est simple et robuste.

I.2 Objectif

La méthodologie est la suivante : chaque battement est décomposé en une somme pondérée de fonctions paramétrées appelées régresseurs. La modélisation d'un battement donné pose donc trois problèmes :

- le choix de la *famille de fonctions* à l'intérieur de laquelle seront choisis les régresseurs,
- la détermination du *nombre* de régresseurs nécessaires à la modélisation,
- l'estimation des valeurs des *paramètres* de ces régresseurs et de leur *pondération* dans la somme.

Idéalement, afin de faciliter l'étiquetage des ondes constitutives du battement, il serait souhaitable que chaque onde du battement soit modélisée de manière satisfaisante par *un régresseur et un seul* (ce qui élimine notamment le problème de l'estimation des pondérations). C'est pourquoi nous nous sommes attaché à trouver une famille de fonctions, que nous appellerons « fonctions bosses », pour lesquelles cette propriété est vérifiée pour la majorité des battements.

II Modélisation classique

Avant d'aborder cette décomposition « sur mesure » du battement cardiaque, nous allons présenter différentes modélisations qui cherchent les régresseurs dans des familles de fonctions conventionnelles.

La modélisation la plus couramment utilisée aujourd'hui dans les logiciels complets d'analyse de l'ECG est l'approximation du signal par une ligne brisée [Pavlidis, 1974], [Ray, 1992] et [Naken, 1993]. Les paramètres de ces lignes sont ensuite étudiés pour le repérage des QRS notamment [Koski, 1996]. Cette représentation est loin de notre objectif, nous ne nous étendrons donc pas plus sur ce type d'analyse.

II.1 Décomposition en ondelettes

La première décomposition proposée ici est une transformée en ondelettes. Ces méthodes sont abondamment décrites dans la littérature, dont on peut trouver des synthèses dans [Mallat, 2000] [Torrésani, 1995]. Il existe un grand nombre de types d'ondelettes telles que celles de Haar, de Morlet, de Daubechies, les « symlets » [Poularikas, 1996], etc. Le choix du type d'ondelettes dépend essentiellement des propriétés recherchées, par exemple l'orthogonalité de la base de décomposition, ou la continuité de la transformation.

Pour s'approcher de l'objectif fixé précédemment, nous proposons ici une décomposition en fonctions de type Coiflet (Mexican Hat) [Poularikas, 1996], fonctions qui ont une forme qui évoque celle des ondes cardiaques (Figure 1) ; de plus, ces fonctions constituent une base orthogonale, ce qui assure l'unicité de la décomposition.

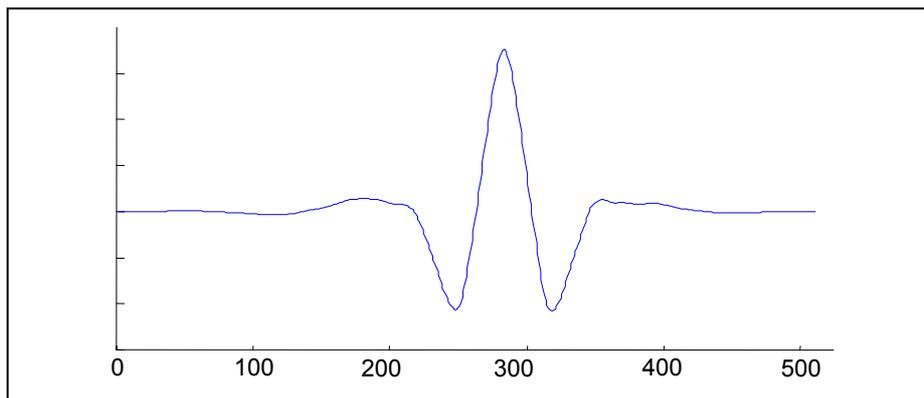


Figure 1: Représentation de l'ondelette mère Coiflet. La forme de cette fonction est proche de celle des ondes cardiaques à modéliser.

II.1.1 Principe de la décomposition

Nous présentons ici les grandes lignes de l'algorithme qui permet de passer d'un battement échantillonné, donc représenté par le vecteur constitué des amplitudes des points d'échantillonnages, à un battement représenté comme une somme pondérée de fonctions.

II.1.1.a Signal ECG

Le signal à décomposer est donc un battement cardiaque isolé ; la transformée en ondelettes discrète impose une contrainte : le nombre de points d'échantillonnage des signaux doit être une puissance entière de 2 ; or la durée des battements dépend du rythme cardiaque. Pour cette décomposition, le battement sera donc représenté par un vecteur de dimension égale à la puissance de 2 la plus proche, par valeur supérieure, du nombre de points d'échantillonnage du battement, en complétant par des zéros placés avant et après celui-ci.

Par exemple, considérons le signal S d'un battement à modéliser (Figure 2). Échantillonné à 500Hz, il est composé de 342 points. Le signal S_0 utilisé pour la décomposition est le vecteur composé du signal S précédé de 85 zéros et suivi de 84 zéros, ce qui porte la dimension de ce vecteur à 512, soit 2^9 .

S_0 est donc aussi un vecteur \mathbf{S}_0 de l'espace à 512 dimensions dont la i -ème coordonnée dans la base canonique est la valeur du signal au point i . Dans tout ce qui suit, les vecteurs de cet espace sont notés en caractères gras.

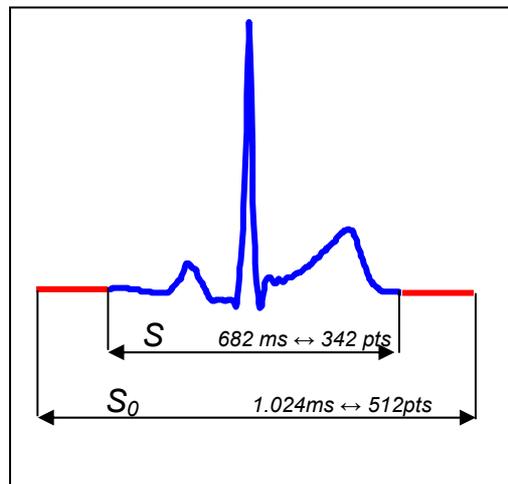


Figure 2 : La transformée en ondelettes orthogonales contraint de travailler avec un signal dont le nombre de points d'échantillonnage est une puissance entière de 2. Le signal S est donc complété de part et d'autre de 0 pour donner S_0 .

II.1.1.b Bibliothèque d'ondelettes

La première étape de la décomposition est la construction de la base d'ondelettes. Si S_0 est le signal à décomposer de longueur N_p (le nombre de points), la base est constituée de N_p

ondelettes orthogonales¹, qui sont toutes déduites de l'ondelette « mère » par translations et dilatations.

Soit φ l'ondelette mère ; la base se construit de la manière suivante :

$$B = \{ \varphi(2^m x \pm n), n \in [1..2^{m-1}], m \in [1..\log_2(N_p)] \} \quad \text{Eq.1}$$

où m et n sont respectivement le *coefficient de dilatation* et *de position* de chacune des ondelettes, et N_p la longueur du signal à modéliser. Les N_p-1 fonctions de la base sont notées $\{\varphi_i\}_{i=[1..N_p-1]}$ dans la suite.

Une telle bibliothèque est présentée sur la Figure 3 ; les ondelettes (ici des Coiflets) qui ont une même dilatation (m constant) sont représentées sur une même ligne.

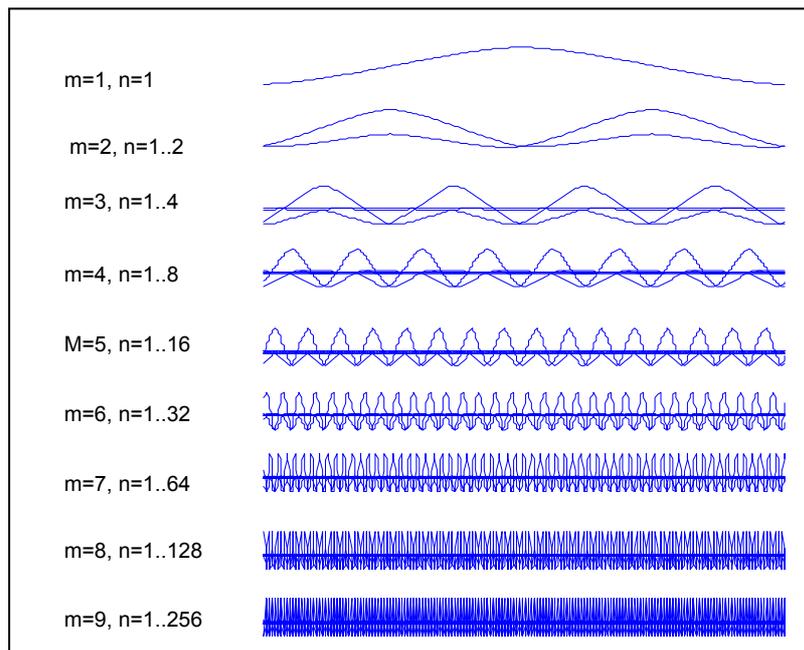


Figure 3 : Famille d'ondelettes utilisée pour la décomposition du signal S_θ . On compte ici 511 ondelettes qui sont toutes orthogonales, et qui constituent ainsi une base orthogonale de l'espace.

¹ En réalité, par construction, la base ne comporte pas 512 mais 511 ondelettes orthogonales, correspondant à autant d'intervalles entre les points.

II.1.1.c Modélisation

La modélisation du signal est peu coûteuse en calcul grâce à la propriété d'orthogonalité des ondelettes évoquée plus haut. Une fois la base construite, la décomposition du signal \mathbf{S}_0 revient à appliquer au vecteur \mathbf{S}_0 la matrice de passage de la base canonique à la base d'ondelettes, ou, en d'autres termes, à calculer les coordonnées du vecteur \mathbf{S}_0 dans la base d'ondelettes :

$$\mathbf{S}_0 = \sum_{i=1}^{N_p-1} \langle \mathbf{S}_0 | \varphi_i \rangle \varphi_i \quad \text{Eq. 2}$$

où $\langle \mathbf{S}_0 | \varphi_i \rangle$ représente la i -ème coordonnées du signal dans la base d'ondelettes.

Ainsi, si l'on décide de choisir $N < N_p - 1$ ondelettes pour modéliser le signal \mathbf{S}_0 , le meilleur modèle Y sera obtenu avec les N ondelettes ayant le plus grand produit scalaire en valeur absolue avec le signal.

$$Y(t) = \sum_{i \in \{A\}} \langle \mathbf{S} | \varphi_i \rangle \varphi_i(t) \quad \text{Eq. 3}$$

où A représente les indices des N plus grands produits scalaires en valeur absolue entre les φ_i et \mathbf{S}_0 ,

L'erreur quadratique moyenne de modélisation s'écrit alors :

$$J = \frac{1}{N_p} \sum_{j=1}^{N_p} (\mathbf{S}(j) - Y(j))^2 \quad \text{Eq. 4}$$

II.1.2 Résultats de la décomposition

L'exemple d'un modèle à $N = 10$ ondelettes du battement précédent est représenté ci-dessous (Figure 4).

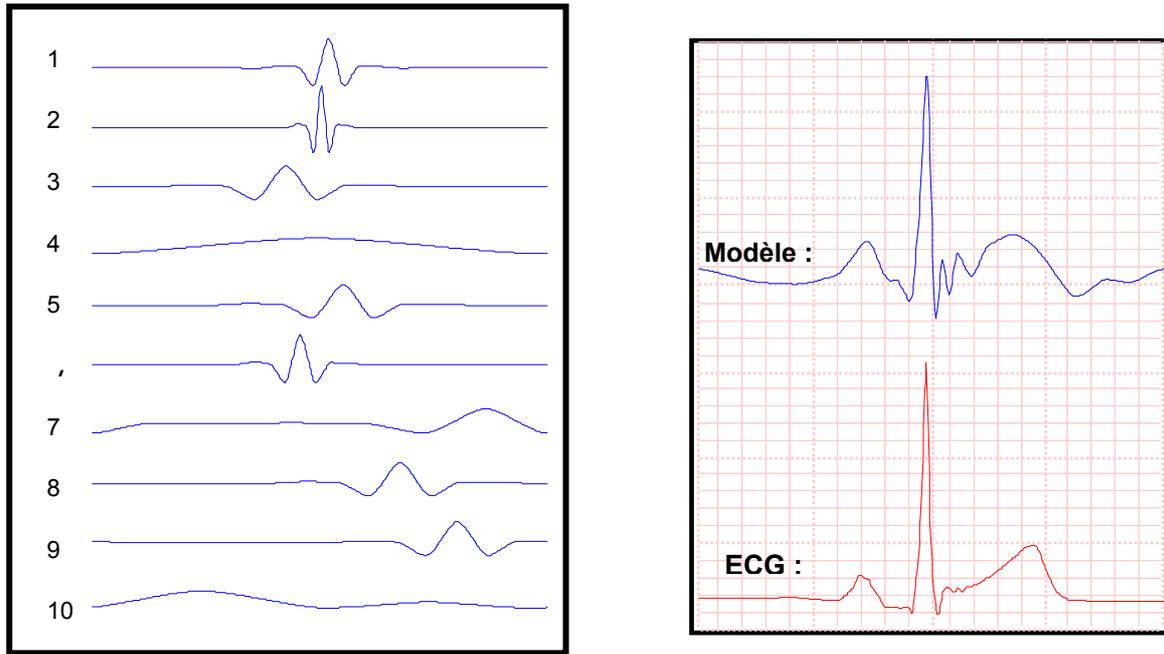


Figure 4 : Le meilleur modèle Y à $N = 10$ Coiflets pour le signal S_0 est représenté en haut à droite. La décomposition est la somme pondérée des 10 ondelettes présentées à gauche.

La modélisation illustrée en Figure 4 n'est pas très satisfaisante : l'erreur quadratique moyenne est de l'ordre de 6.10^{-3} . Pour que le modèle soit mieux représentatif du signal, il serait utile d'augmenter le nombre d'ondelettes utilisées ($N > 10$).

De plus, ce modèle ne répond pas à notre exigence indiquée dans le paragraphe I.2, soit modéliser chaque onde cardiaque par un régresseur particulier. Ici, la somme pondérée de tous les régresseurs peut être comparée au signal original et en fournir une expression analytique plus ou moins satisfaisante selon N , mais *aucun régresseur n'a de signification particulière en termes d'ondes cardiaques*.

Une application de transformées en ondelettes pour l'analyse du signal ECG est proposé par [Thoraval, 1994]. Toutefois le signal qui subit cette transformation n'est pas le signal ECG mais un signal calculé à partir de ce dernier par une application non linéaire qui met en évidence les points d'inflexions de la courbe correspondant à des débuts et des fins d'ondes potentielles. Une chaîne de Markov cachée analyse ensuite les résultats issus de cette décomposition pour repérer les ondes caractéristiques [Senhadji, 1996]. Cette étude montre que la transformée non linéaire associée à une transformée en ondelettes donne de meilleurs résultats qu'une transformée en ondelette seule [Senhadji, 1996], ce qui montre l'importance

d'un travail sur le signal en amont de la reconnaissance qui est l'approche proposée ici, cependant notre volonté d'associer à chaque régresseur (ici des ondelettes) une onde caractéristique n'est pas satisfaite par cette démarche.

Une transformation du signal ECG en ondelette est également proposée par [Bahoura, 1997] pour la détection des ondes QRS, P et T. Les résultats sur le repérage des ondes QRS est de 99.7% de réussite sur la base MIT [MIT-DB, 1997], ce qui est légèrement inférieur à celui obtenu par notre algorithme 99.91%, et les résultats sur les ondes P et T ne sont pas communiqués, à cause vraisemblablement de l'absence de base étiquetée.

Une troisième application des ondelettes couramment proposée en ECG est la compression du signal [Hilton, 1997], [Ahmed, 2000], ce qui n'est pas le but recherché ici dans un premier temps.

II.2 Modélisation par un réseau de neurones à fonctions dorsales

Les réseaux de neurones à fonction d'activation dorsale constituent un deuxième outil d'approximation classiquement utilisé en régression [Dreyfus, 2002].

II.2.1 Principe de la modélisation

Un réseau de neurones est une somme pondérée de fonctions non-linéaires paramétrées, appelées « neurones cachés », des variables de la fonction à modéliser. Chaque neurone caché est généralement la tangente hyperbolique d'une somme pondérée des variables du modèle. Il est commode de représenter graphiquement le modèle, comme nous l'avons fait sur la Figure 5 pour un réseau à une seule variable (le temps), ce qui correspond à la modélisation d'un signal temporel tel que celui de l'électrocardiogramme.

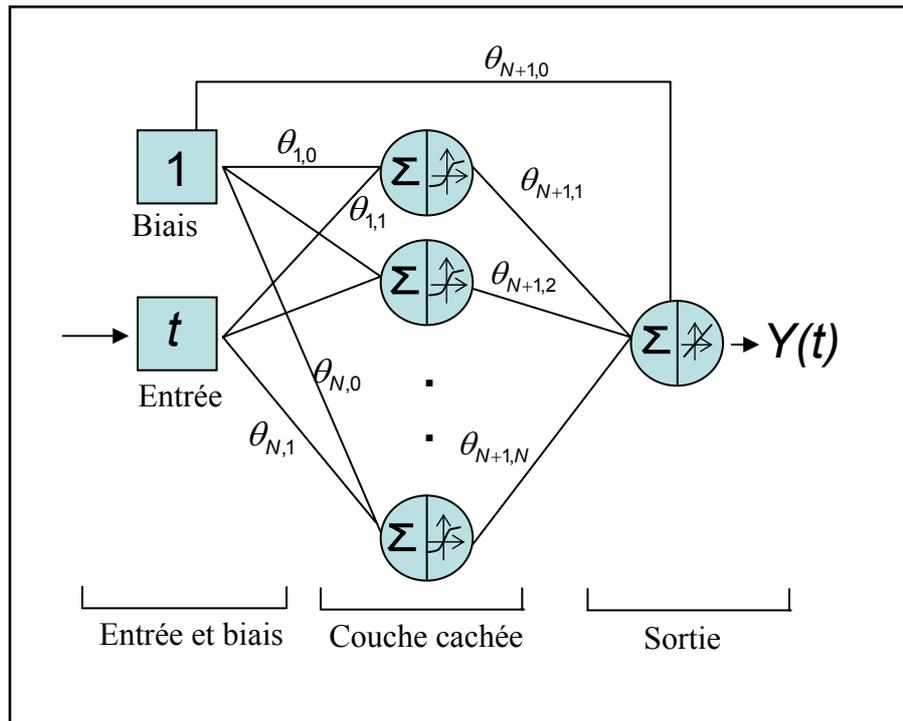


Figure 5 : Réseau de neurones à une couche cachée. Ici pour modéliser un signal d'une seule variable à une dimension, il suffit d'une unique entrée et d'une sortie. La sortie est une somme pondérée des fonctions tangentes hyperboliques des N neurones cachés.

Le modèle Y obtenu s'écrit de la manière suivante :

$$Y(t) = \sum_{i=1}^N \theta_{N+1,i} \tanh(\theta_{i,1} t + \theta_{i,0}) + \theta_{N+1,0} \quad \text{Eq. 5}$$

où N est le nombre de neurones cachés, $\theta_{i,1}$ la pondération de la variable du modèle dans la fonction réalisée par le neurone caché i , $\theta_{N+1,i}$ la pondération du neurone caché i dans la sortie du modèle (indiquée $N+1$), $\theta_{i,0}$ la pondération du biais dans la fonction réalisée par le neurone caché i , et $\theta_{N+1,0}$ la pondération du biais dans la sortie du modèle.

L'estimation des paramètres se fait par optimisation d'une fonction de coût (définie plus loin par l'équation 8), calculée sur un ensemble d'apprentissage. Celui-ci est constitué des couples de valeurs : l'abscisse de chaque point (position temporelle t) et l'ordonnée correspondante. Pour le signal S de N_p points, cette base de données est constituée des N_p couples :

$$\{(1, S(1)), (2, S(2)), \dots, (N_p, S(N_p))\} \quad \text{Eq. 6}$$

II.2.2 Résultats de la modélisation

Après estimation des paramètres θ du réseau, moyennant un bon choix a priori du nombre de neurones cachés $N = 10$, on obtient un modèle satisfaisant au sens de l'erreur quadratique moyenne, de l'ordre de 10^{-4} . Mais, ici encore, les régresseurs, ici les tangentes hyperboliques correspondant aux neurones cachés, ne portent pas d'information en eux-mêmes ; seule leur somme pondérée est significative (Figure 6).

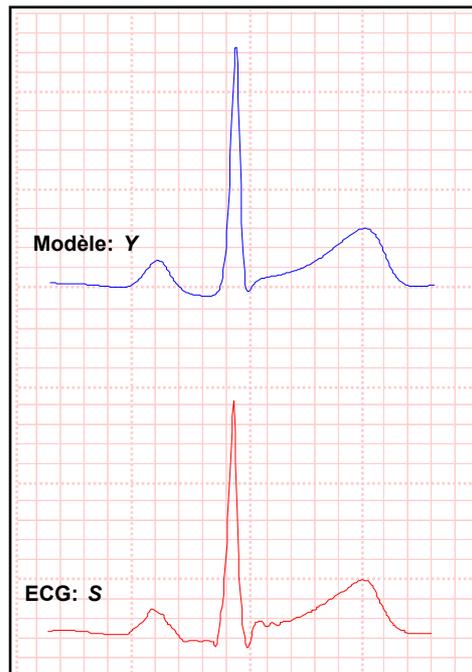


Figure 6 : Modèle obtenu avec un réseau de neurones à fonctions dorsales (tangente hyperbolique) avec 10 neurones cachés. Le modèle est la somme de ces fonctions qui sont autant de régresseurs. Ce modèle est satisfaisant au sens de l'erreur quadratique moyenne, mais, individuellement, ces fonctions n'ont pas de signification.

Les réseaux de neurones à fonction dorsale ne sont donc pas adaptés pour l'étiquetage des ondes caractéristiques. Le fait de ne pas avoir de correspondance entre régresseurs et ondes caractéristiques vient essentiellement de la forme des régresseurs (tangente hyperbolique ici) qui n'est pas adaptée à la forme des ondes.

L'étude d'un réseau à fonction radiale va donner des résultats plus intéressants.

II.3 Modélisation par un réseau de fonctions radiales de base (RBF pour Radial Basis Function)

II.3.1 Principe de la modélisation

La modélisation par RBF est beaucoup plus pertinente pour l'étude qui nous intéresse car on peut choisir comme fonction de base une gaussienne qui, une fois ses paramètres adaptés, possède une forme qui pourrait être une bonne représentation de chacune des ondes cardiaques.

Ici, les paramètres du réseau sont, d'une part, les pondérations des fonctions radiales, et, d'autre part, les paramètres des gaussiennes (moyenne et écart-type). De même que pour les réseaux précédents, le nombre N de fonctions à utiliser pour le modèle doit être fixé a priori (Figure 7).

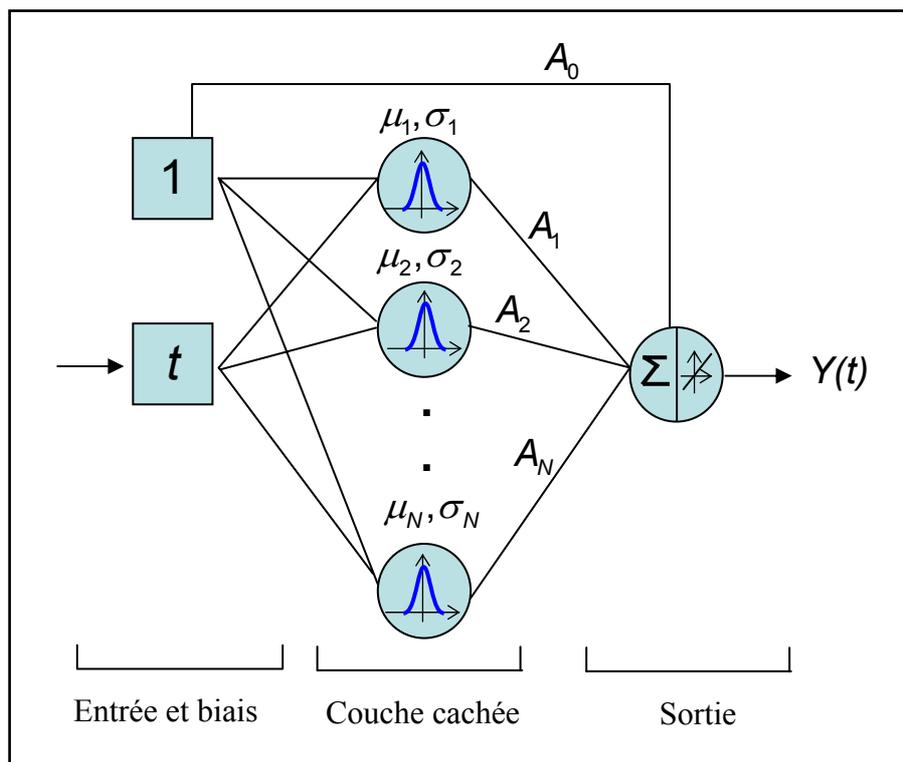


Figure 7 : Réseau de neurones à fonctions radiales. L'architecture est identique à celle du réseau précédent. En revanche, les paramètres sont les paramètres des N gaussiennes, et le poids de chaque gaussienne dans la combinaison linéaire de sortie.

Le modèle obtenu Y peut s'écrire de la manière suivante :

$$Y(t) = \sum_{i=1}^N A_i \exp\left(-\frac{(t - \mu_i)^2}{2\sigma_i^2}\right) + A_0 \quad \text{Eq. 7}$$

où A_i est la pondération de la gaussienne i dans la somme (analogue aux $\theta_{N+1, i}$ du modèle précédent), μ_i et σ_i respectivement la moyenne et l'écart type de la gaussienne i .

Ce modèle Y est particulièrement intéressant car c'est une combinaison linéaire de fonctions gaussiennes : c'est donc une décomposition du signal en une somme de fonctions qui sont proches en forme des ondes caractéristiques à modéliser. On espère ainsi associer à chaque onde caractéristique une fonction gaussienne.

L'optimisation de ces paramètres se fait par la minimisation d'une fonction de coût J calculée sur la base même d'apprentissage que dans le cas des réseaux de neurones. On utilise pour J l'erreur quadratique moyenne :

$$J(\tilde{\mu}, \tilde{\sigma}, \tilde{A}) = \frac{1}{N_p} \sum_{i=1}^{N_p} (S(i) - Y(i))^2 \quad \text{Eq. 8}$$

$$J(\tilde{\mu}, \tilde{\sigma}, \tilde{A}) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left(S(i) - \sum_{j=1}^N A_j \exp\left(-\frac{(i - \mu_j)^2}{2\sigma_j^2}\right) + A_0 \right)^2 \quad \text{Eq. 9}$$

où $\tilde{\mu} = \{\mu\}_{i=[1, N]}$, $\tilde{\sigma} = \{\sigma\}_{i=[1, N]}$ et $\tilde{A} = \{A\}_{i=[1, N]}$

J est non linéaire en les paramètres, ce qui impose, pour la recherche de son minimum, l'utilisation d'un algorithme itératif qui nécessite une initialisation arbitraire des paramètres du réseau [Dreyfus, 2002]. Pour maximiser les chances de trouver le minimum global de J , plusieurs initialisations des poids et donc plusieurs apprentissages sont nécessaires, ce qui multiplie d'autant le temps de calcul.

Afin d'éviter ces multiples apprentissages, il est souhaitable que les valeurs initiales des paramètres soient telles que la fonction de coût soit aussi proche que possible de son minimum.

II.3.2 Initialisation du réseau – Orthogonal Forward Regression

L'idée est donc d'initialiser les paramètres du réseau (écarts-types, moyennes et amplitudes de chacune des gaussiennes) avec des valeurs proches des valeurs optimales. Une telle tâche peut être réalisée par l'algorithme d'« orthogonal forward regression (OFR) » fondé sur l'algorithme d'orthogonalisation de Gram-Schmidt [Chen, 1967], [Oussar, 2000].

Un tel algorithme permet de classer des régresseurs par ordre de pertinence pour représenter un vecteur donné. On va donc utiliser cet algorithme pour classer par ordre de pertinence pour le battement à modéliser des gaussiennes prédéfinies. Les 10 gaussiennes les plus pertinentes permettront l'initialisation du réseau de neurones.

Plus précisément, soit une famille de régresseurs \mathbf{B} , et un signal \mathbf{S} ; l'OFR permet donc de classer par ordre de pertinence les régresseurs de la famille \mathbf{B} pour l'explication de \mathbf{S} , par orthogonalisations successives [Chen, 1967]. Nous montrons ainsi dans le paragraphe suivant comment l'on peut ainsi construire une bibliothèque \mathbf{B} de N_b gaussiennes, et la manière de sélection des N gaussiennes les plus pertinentes pour l'explication du battement cardiaque \mathbf{S} .

II.3.2.a Construction de la bibliothèque \mathbf{B}

Comme dans le cas de la base d'ondelettes construite précédemment, la bibliothèque \mathbf{B} contient des gaussiennes de différentes largeurs centrées, à différents emplacements temporels (Figure 8).

Pour un signal de N_p points, \mathbf{B} est l'ensemble suivant :

$$\mathbf{B} = \left\{ \begin{array}{l} \mathbf{N}(\mu, \sigma), \\ \mu = j.n/2^k, \sigma = n/2^{k-2} \\ j \in [0, 2^k], k \in [1, K_{\max}] \end{array} \right\} \quad \text{Eq. 10}$$

où $\mathbf{N}(\mu, \sigma)$ est la fonction gaussienne d'écart-type σ et de moyenne μ ; K_{\max} est défini de telle manière que le plus petit écart-type σ^{\min} corresponde à la demi largeur de la plus petite onde caractéristique de l'ECG (nous utilisons dans la pratique $\sigma^{\min} = n/2^{K_{\max}-2} = 20$ ms).

Dans la suite des équations, les N_b régresseurs de la bibliothèque \mathbf{B} seront notés $\{\mathbf{B}_i\}_{i=[1..N_b]}$.

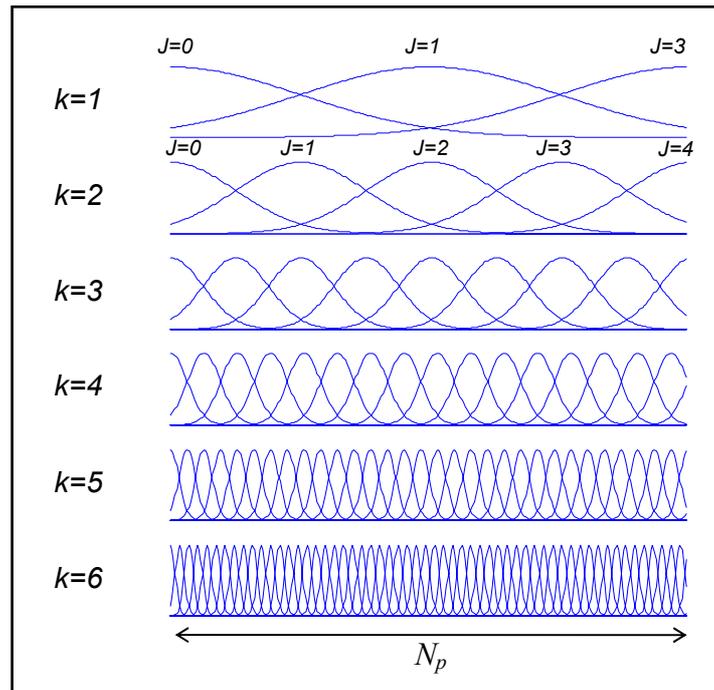


Figure 8 : La bibliothèque de gaussiennes est construite sur le schéma de la base d'ondelettes : elle compte ici 132 gaussiennes pour $N_p=342$, dont les moyennes et les écarts types sont donnés par l'équation 10.

II.3.2.b Représentations et conventions

Le battement à modéliser \mathbf{S} et les régresseurs de la bibliothèque sont des signaux temporels constitués de N_p points d'échantillonnage. Pour l'application de l'algorithme d'OFR, ils sont considérés comme des vecteurs de l'espace à N_p dimensions, appelé habituellement « espace des observations ».

Ici, on considère donc deux représentations ^{II}:

- une représentation *temporelle*, dans laquelle le signal \mathbf{S} et les régresseurs sont des fonctions du temps, cet espace est habituellement appelé l'espace des variables,
- et une représentation *vectorielle*, dans laquelle le signal \mathbf{S} et les régresseurs \mathbf{B} sont des vecteurs de l'espace des observations, de dimension N_p . Pour distinguer ces deux représentations, les vecteurs seront représentés en gras.

La sélection des N régresseurs s'effectue dans l'espace des observations ; elle commence dans l'espace de dimension N_p , et, à chaque itération, l'algorithme opère dans un espace dont la dimension est réduite d'une unité, qui est l'espace orthogonal à celui de l'itération précédente.

^{II} L'importance de cette distinction entre l'espace des variables et l'espace des observations reste superflue ici. Elle est cependant introduite à ce niveau pour faciliter l'analogie avec la méthode présentée dans les paragraphes suivants.

Dans ce qui suit, le numéro de l'itération sera indiqué par un exposant à droite de la lettre représentant le signal ou l'un des régresseurs. Notons qu'un vecteur portant un indice supérieur i s'exprime dans une base orthogonale à celle où s'exprime un vecteur portant un indice supérieur $i-1$; un tel vecteur portera en indice m_i , qui représente son numéro dans la bibliothèque.

Dans la représentation temporelle, une fonction régresseur ne portera que l'indice m_i , suffisant pour indiquer l'itération qui a permis sa sélection. L'erreur de modélisation à l'issue de chaque itération portera, en exposant, le numéro de cette itération.

II.3.2.c Itération 1 de l'algorithme

- Sélection

À la première itération, le signal à modéliser est $\mathbf{S} = \mathbf{S}^1$, et les régresseurs appartiennent à la famille $\mathbf{B} = \mathbf{B}^1 = \{\mathbf{B}_i^1\}_{i=[1..N_b]}$.

Le vecteur de la famille \mathbf{B}^1 qui explique le mieux le signal \mathbf{S}^1 est celui qui réalise le plus petit angle avec ce dernier dans l'espace de dimension N_p . Autrement dit, c'est celui qui a le plus grand cosinus en valeur absolue avec le signal \mathbf{S}^1 .

Soit m_1 l'indice de ce régresseur dans la famille \mathbf{B}^1 :

$$m_1 = \operatorname{argmax}_{k \in [1..N_b]} \left(\frac{|\langle \mathbf{S}^1 | \mathbf{B}_k^1 \rangle|}{\|\mathbf{S}^1\| \cdot \|\mathbf{B}_k^1\|} \right) \quad \text{Eq. 11}$$

On remarque que ce calcul, pour des vecteurs régresseurs ayant tous la même norme, revient à chercher le régresseur qui a le produit scalaire le plus grand avec \mathbf{S}^1 , ce qui est équivalent à rechercher le régresseur le plus corrélé avec \mathbf{S}^1 dans l'espace temporel.

Le régresseur $\mathbf{B}_{m_1}^1$ sélectionné explique \mathbf{S}^1 dans sa direction ; de \mathbf{S}^1 il reste donc à « expliquer » ce qui se trouve dans le sous-espace orthogonal à ce vecteur régresseur.

- Orthogonalisation

On orthogonalise donc les N_b-1 régresseurs restants, et le vecteur \mathbf{S}^1 par rapport à $\mathbf{B}_{m_1}^1$

$$\begin{aligned} \mathbf{B}_k^2 &= \mathbf{B}_k^1 - \langle \mathbf{B}_k^1 | \mathbf{B}_{m_1}^1 \rangle \mathbf{B}_{m_1}^1 & \forall k &= [1, N_b] \setminus \{m_1\} \\ \mathbf{S}^2 &= \mathbf{S}^1 - \langle \mathbf{S}^1 | \mathbf{B}_{m_1}^1 \rangle \mathbf{B}_{m_1}^1 \end{aligned} \quad \text{Eq. 12}$$

La famille $\mathbf{B}^2 = \{\mathbf{B}_i^2\}_{i=[1..N_b] \setminus \{m_1\}}$ est ce qui reste des vecteurs de la famille \mathbf{B}^1 dans l'espace orthogonal à $\mathbf{B}_{m_1}^1$; \mathbf{S}^2 est ce qui reste de \mathbf{S}^1 à expliquer avec les régresseurs de \mathbf{B}^2 c'est-à-dire la partie de \mathbf{S}^1 se trouvant dans le sous-espace orthogonal à $\mathbf{B}_{m_1}^1$ (Figure 9).

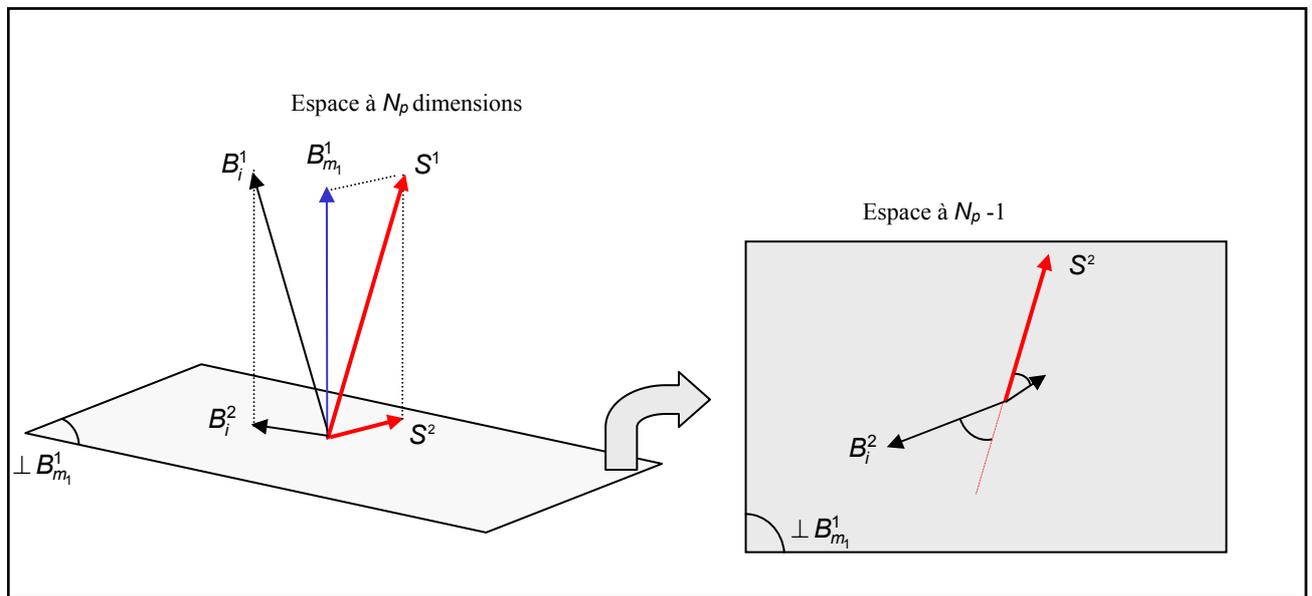


Figure 9: Suite à la sélection du régresseur m_1 qui explique au mieux le signal \mathbf{S} , les régresseurs restants, ainsi que le signal, sont projetés dans l'espace orthogonal à $\mathbf{B}_{m_1}^1$, noté $\perp \mathbf{B}_{m_1}^1$. L'itération 2 s'effectue dans ce nouvel espace de dimension $N_p - 1$.

II.3.2.d Itérations suivantes

Cette première orthogonalisation diminue donc l'espace d'une dimension : il est maintenant de dimension $N_p - 1$; nous disposons de $N_b - 1$ régresseurs pour expliquer le signal \mathbf{S}^2 .

Ainsi, à l'itération $i-1$, nous avons effectué $i-1$ orthogonalisations, et nous avons sélectionné les régresseurs $(\mathbf{B}_{m_1}^1, \mathbf{B}_{m_2}^2, \dots, \mathbf{B}_{m_{i-1}}^{i-1})$. Il nous reste $N_b - i + 1$ régresseurs pour expliquer le signal \mathbf{S}^i dans un espace de dimension $N_p - i + 1$.

- Sélection

Comme précédemment, l'indice m_i du régresseur sélectionné est celui pour lequel la valeur absolue du cosinus avec le signal \mathbf{S}^i est maximale ; m_i s'écrit :

$$m_i = \underset{k \in [1, N_b] \setminus \{m_1, \dots, m_{i-1}\}}{\operatorname{argmax}} \left(\frac{|\langle \mathbf{S}^i | \mathbf{B}_k^i \rangle|}{\|\mathbf{S}^i\| \cdot \|\mathbf{B}_k^i\|} \right) \quad \text{Eq. 13}$$

- Orthogonalisation

Le signal et les régresseurs restants sont ensuite orthogonalisés par rapport à $\mathbf{B}_{m_i}^i$.

$$\begin{aligned} \mathbf{B}_k^{i+1} &= \mathbf{B}_k^i - \langle \mathbf{B}_k^i | \mathbf{B}_{m_i}^i \rangle \mathbf{B}_{m_i}^i \quad \forall k = [1, N_b] \setminus \{m_1, m_2, \dots, m_i\} \\ \mathbf{S}^{i+1} &= \mathbf{S}^i - \langle \mathbf{S}^i | \mathbf{B}_{m_i}^i \rangle \mathbf{B}_{m_i}^i \end{aligned} \quad \text{Eq. 14}$$

- Terminaison

En itérant cet algorithme N fois, on sélectionne donc les N régresseurs de la bibliothèque qui expliquent le mieux le signal. On peut donc écrire, d'après l'équation 14 :

$$\mathbf{S} = \sum_{i=1}^N \langle \mathbf{S}^i | \mathbf{B}_{m_i}^i \rangle \mathbf{B}_{m_i}^i + \mathbf{S}^{N+1} \quad \text{Eq. 15}$$

$$\Leftrightarrow \exists \{ \mathbf{A}_i \}_{i=[1..N]} \quad / \quad \mathbf{S}(t) = \sum_{i=1}^N \mathbf{A}_i \mathbf{B}_{m_i}(t) + \mathbf{E}(t) \quad \text{Eq. 16}$$

où $\mathbf{E}(t)$ est l'erreur de modélisation.

[Chen, 1967] montre que la famille des $\{ \mathbf{A}_i \}_{i=[1..N]}$ est identique à celle obtenue par une optimisation linéaire des moindres carrés des fonctions $\{ \mathbf{B}_{m_i} \}_{i=[1..N]}$.

Dans notre cas, où l'on réalise une modélisation non linéaire, les paramètres $\{ (\mu_{m_i}, \sigma_{m_i}, \mathbf{A}_{m_i}) \}_{i=[1..N]}$ des régresseurs $\{ \mathbf{B}_{m_i} \}_{i=[1..N]}$ constituent un bon point de départ pour initialiser les paramètres du réseau en vue de l'optimisation [Oussar, 2000].

II.3.3 Estimation des paramètres du réseau

Ainsi, le réseau étant initialisé avec des coefficients proches d'une bonne solution, on va maintenant les ajuster grâce à la minimisation de l'erreur quadratique moyenne entre le modèle créé et le signal \mathbf{S} .

Pour un réseau à N gaussiennes, le modèle a l'expression donnée par l'équation 7, l'erreur quadratique moyenne est donc

$$J(\tilde{\mu}, \tilde{\sigma}, \tilde{\mathbf{A}}) = \frac{1}{N_p} \sum_{j=1}^{N_p} (\mathbf{S}(j) - Y(j))^2 \quad \text{Eq. 17}$$

$$J(\tilde{\mu}, \tilde{\sigma}, \tilde{\mathbf{A}}) = \frac{1}{N_p} \sum_{j=1}^{N_p} \left(\mathbf{S}(j) - \sum_{i=1}^N A_i \exp\left(-\frac{(j - \mu_i)^2}{2\sigma_i^2}\right) \right)^2 \quad \text{Eq. 18}$$

L'apprentissage du réseau se fait par minimisation de J par rapport aux paramètres. Pour ce faire, nous utilisons ici quelques itérations de gradient simple, puis des itérations de BFGS (du nom de ses inventeurs Broyden, Fletcher, Goldfarb et Shanno) [Minoux, 1983] qui est un algorithme du second ordre : la direction d'optimisation est calculée à partir d'une estimation des dérivées secondes de J par rapport à ces paramètres, ce qui assure une bonne convergence en fin d'apprentissage (cf. Annexe B).

II.3.4 Résultats de la modélisation

Nous présentons, sur la Figure 10, les résultats pour $N = 10$ gaussiennes, ce qui représente en tout 30 paramètres dans le réseau. Ce choix sera justifié dans la suite.

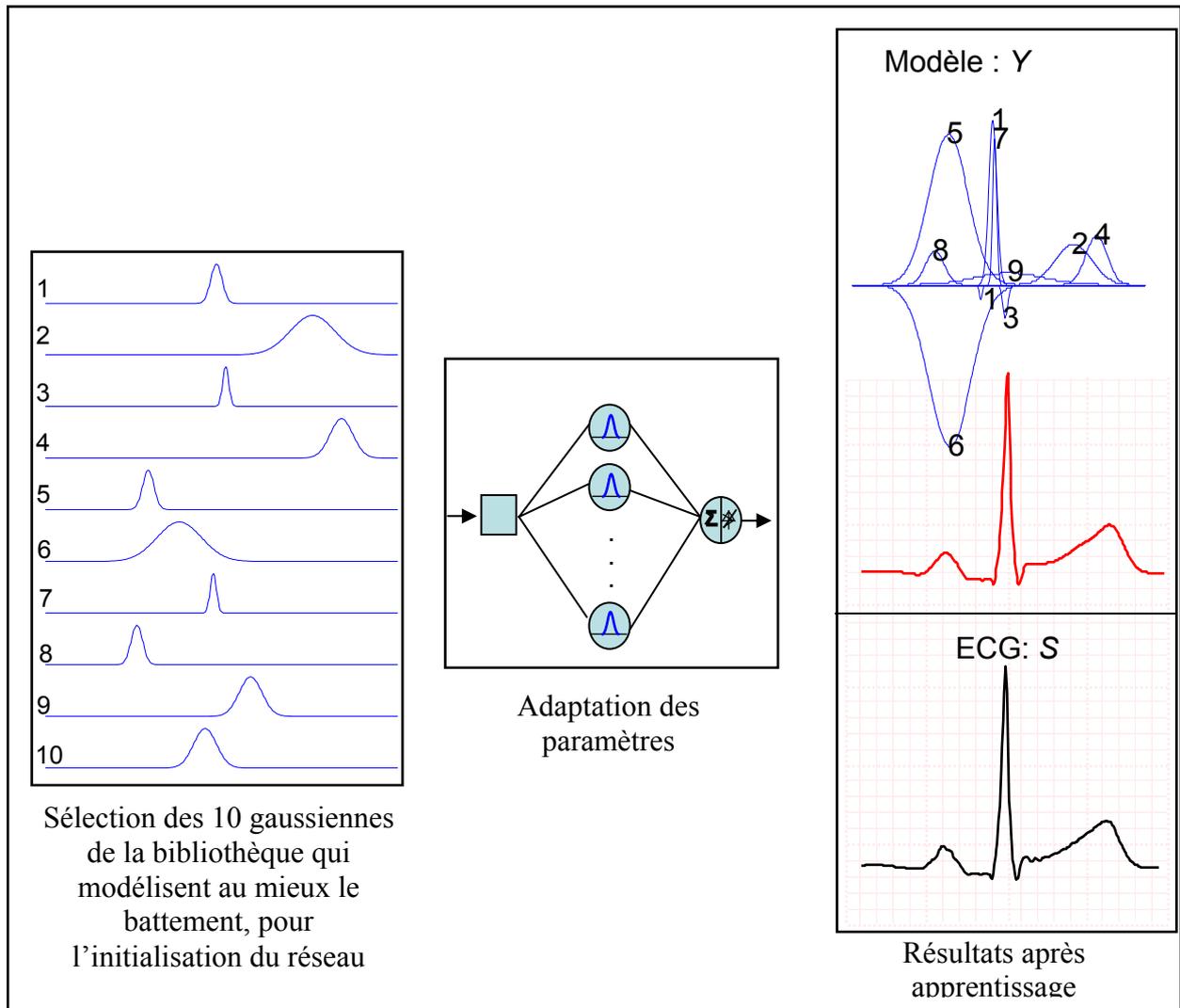


Figure 10 : La décomposition du battement en fonctions gaussiennes s'effectue en deux étapes successives : la sélection des $N = 10$ gaussiennes les plus représentatives du signal dans une bibliothèque préalablement construite, et l'adaptation des paramètres de ces gaussiennes par un réseau de RBF. Le modèle obtenu est présenté à droite au centre, et la décomposition est présentée à droite en haut.

Le modèle obtenu est satisfaisant au sens de l'erreur quadratique moyenne, qui est de l'ordre de $3 \cdot 10^{-5}$, ce qui prouve que l'apprentissage qui a été effectué est correct. Mais ce critère de qualité de la représentation globale de S n'est pas suffisant pour rendre compte d'une identification des ondes caractéristiques du battement. On remarque, en regardant le modèle décomposé sur la Figure 10 en haut à droite, que les gaussiennes sont situées à des emplacements correspondant assez bien aux ondes P, Q, R, S et T, mais qu'il reste deux inconvénients majeurs :

- certaines des gaussiennes (les gaussiennes 5 et 6 dans le cas présent) se compensent et, par conséquent, aucune des deux ne représente une onde caractéristique,
- beaucoup d'ondes sont modélisées par plus d'une gaussienne, comme ici le QRS et l'onde T.

Ces deux remarques vont nous conduire à modifier légèrement l'algorithme.

II.4 Régression orthogonale directe généralisée (GOFR : generalized orthogonal forward regression) par fonctions radiales de base

II.4.1 Principe de l'algorithme

Pour éviter ces problèmes de modélisation par plusieurs régresseurs du modèle d'une même onde caractéristique, nous avons développé méthode originale de *sélection adaptative* : contrairement à l'algorithme précédent dans lequel les régresseurs pertinents sont d'abord sélectionnés, puis ajustés au battement, la procédure présentée ici intègre à chaque étape *sélection* et *ajustement*, régresseur par régresseur : avant chaque orthogonalisation par rapport à un régresseur sélectionné, on effectue une optimisation sommaire des paramètres de celui-ci afin de le rapprocher au mieux du signal. Une fois optimisé, ce nouveau régresseur remplace l'ancien dans la bibliothèque et l'orthogonalisation se fait par rapport à ce dernier.

Nous décrivons ce principe en détail ci-dessous.

S est le signal à modéliser, et nous disposons de la même bibliothèque que précédemment

$$\{B_i\}_{i=1..N_b}$$

Soit Y le modèle construit, et E l'erreur de modélisation. Avant la première itération, le modèle construit est $Y = 0$; l'erreur de modélisation est donc $E^1 = S - 0 = S$.

Ici, la distinction entre l'espace des variables et l'espace des observations introduite au paragraphe II.3.2.b prend tout son sens. Dans la représentation temporelle, le signal, noté S , est constitué des N_p points d'échantillonnage, et les régresseurs sont des fonctions temporelles paramétrées. L'apprentissage, qui porte sur les paramètres de ces régresseurs, est donc réalisé

pour cette représentation, et c'est également dans cette représentation que se calcule l'erreur quadratique moyenne J entre le signal \mathbf{S} et le modèle Y , et l'erreur de modélisation E .

Dans la représentation vectorielle, le signal \mathbf{S} est un vecteur de N_p dimensions, ainsi que chacun des régresseurs. On effectue le calcul des angles pour la sélection itérative des N régresseurs du modèle. Comme précédemment, le numéro de l'itération $i = 1..N$ est indiqué par un exposant. On ajoute ici une (*) à droite du symbole des régresseurs indiquant que ceux-ci ont été ajustés par optimisation.

Ainsi le signal qui permet la sélection du régresseur à la i -ème itération dans l'espace des observations : \mathbf{S}^i , n'est plus relié à l'erreur de modélisation E^i qui permet l'apprentissage de cette bosse une fois sélectionnée. L'hypothèse faite ici est que le \mathbf{S}^i de la représentation vectorielle est proche de la partie de E^i qui est dans l'espace orthogonal aux régresseurs d'indices (m_1, m_{i-1}) (une vérification de cette hypothèse est proposée en Annexe C).

II.4.1.a Première itération

- Sélection

À la première itération, le signal est $\mathbf{S} = \mathbf{S}^1$, et les vecteurs régresseurs sont les $\mathbf{B}^1 = \{\mathbf{B}_i^1\}_{i=1..N_b}$; le vecteur de la famille \mathbf{B}^1 le plus pertinent porte l'indice m_1 , qui s'écrit de la manière suivante :

$$m_1 = \operatorname{argmax}_{k \in [1, N_b]} \left(\frac{|\langle \mathbf{S}^1 | \mathbf{B}_k^1 \rangle|}{\|\mathbf{S}^1\| \cdot \|\mathbf{B}_k^1\|} \right) \quad \text{Eq. 19}$$

Une fois sélectionné, le régresseur \mathbf{B}_{m_1} peut être envisagé comme une fonction temporelle paramétrée utilisée pour modéliser \mathbf{S} ; on optimise ses paramètres pour le faire se « rapprocher » au mieux de \mathbf{S}^{III} avant de chercher un deuxième régresseur.

^{III} Ou plus exactement la partie de \mathbf{S} comprise dans une fenêtre temporelle définie par les paramètres de \mathbf{B}_{m_1} :

μ_{m_1}, σ_{m_1}

- Optimisation du régresseur $B_{m_1}^1$

L'optimisation du régresseur se fait dans la représentation temporelle en minimisant l'erreur quadratique moyenne J entre ce qui est à modéliser: $E^1=S$, et la fonction B_{m_1} , d'écart-type σ_{m_1} , de position temporelle μ_{m_1} , et d'amplitude A_{m_1} .

$$J(\sigma_{m_1}, \mu_{m_1}, A_{m_1}) = \frac{1}{N_p} \sum_{k=1}^{N_p} \left(E^1(k) - A_{m_1} \exp\left(-\frac{(k - \mu_{m_1,1})^2}{2\sigma_{m_1}^2}\right) \right)^2 \quad \text{Eq. 20}$$

Cette optimisation a l'avantage d'être rapide car on n'optimise ici que trois paramètres : les paramètres d'une gaussienne.

A l'issue de cette optimisation, les paramètres $\sigma_{m_1}^*$, $\mu_{m_1}^*$ et $A_{m_1}^*$ du régresseur $B_{m_1}^*$ sont ajustés, et sont solutions du problème d'optimisation.

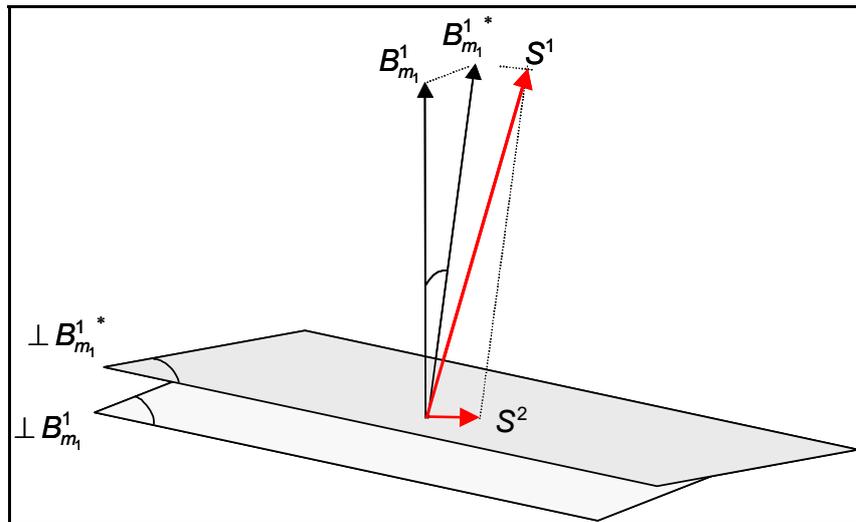


Figure 11 : La première itération de l'algorithme est constituée de trois étapes. 1) la sélection du meilleur régresseur $B_{m_1}^1$ de la bibliothèque obtenue par le calcul des angles que chacun d'eux forme avec le signal. 2) L'ajustement des paramètres du régresseur sélectionné pour obtenir le nouveau régresseur $B_{m_1}^1*$. 3) L'orthogonalisation des régresseurs restant et du signal par rapport à $B_{m_1}^1*$ pour se placer dans un espace orthogonal à $B_{m_1}^1*$, noté $\perp B_{m_1}^1*$. Les itérations suivantes sont construites sur ce même schéma de trois étapes.

- Orthogonalisation

L'orthogonalisation est effectuée par rapport au *régresseur ajusté* $\mathbf{B}_{m_1}^{1*}$ et non par rapport au *régresseur sélectionné* $\mathbf{B}_{m_1}^1$ comme c'était le cas précédemment :

$$\begin{aligned}\mathbf{B}_k^2 &= \mathbf{B}_k^1 - \left\langle \mathbf{B}_k^1 \mid \mathbf{B}_{m_1}^{1*} \right\rangle \mathbf{B}_{m_1}^{1*} & \forall k = [1, N_b] \setminus \{m_1\} \\ \mathbf{S}^2 &= \mathbf{S}^1 - \left\langle \mathbf{S}^1 \mid \mathbf{B}_{m_1}^{1*} \right\rangle \mathbf{B}_{m_1}^{1*}\end{aligned}\tag{Eq. 21}$$

\mathbf{S}^2 et les $\{\mathbf{B}_i^2\}_{i=[1..N_b]\setminus\{m_1\}}$ sont respectivement ce qui reste du signal et des régresseurs dans l'espace orthogonal à $\mathbf{B}_{m_1}^{1*}$.

En revenant à la représentation temporelle, l'erreur de modélisation E^2 à l'itération 2 s'écrit :

$$E^2(t) = \mathbf{S}(t) - \mathbf{Y}(t) = \mathbf{S}(t) - \mathbf{B}_{m_1}^*(t)\tag{Eq. 22}$$

le modèle étant, à cette étape, $\mathbf{Y}(t) = \mathbf{B}_{m_1}^*(t)$

Après la première orthogonalisation, l'algorithme opère dans un espace de dimension $N_p - 1$, et nous disposons de $N_b - 1$ régresseurs pour expliquer le signal \mathbf{S}^2 , ce qui prépare à l'itération suivante (sélection, optimisation, orthogonalisation).

II.4.1.b Itérations suivantes

Supposons réalisées $i - 1$ itérations : nous avons donc effectué $i - 1$ sélections, optimisations, et orthogonalisations, pour obtenir les $i - 1$ régresseurs ajustés $\mathbf{B}_{m_1}^*, \mathbf{B}_{m_2}^*, \dots, \mathbf{B}_{m_{i-1}}^*$.

Le modèle \mathbf{Y} s'écrit, à la fin de l'itération $i-1$:

$$\mathbf{Y}(t) = \sum_{k=1}^{i-1} \mathbf{B}_{m_k}^*(t)\tag{Eq. 23}$$

Et l'erreur de modélisation est :

$$E^i(t) = \mathbf{S}(t) - \mathbf{Y}(t) = \mathbf{S}(t) - \sum_{k=1}^{i-1} \mathbf{B}_{m_k}^*(t)\tag{Eq. 24}$$

Il nous reste donc $N_b - i + 1$ régresseurs pour expliquer le signal E^i dans un espace de dimension $N_p - i + 1$ orthogonal aux $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$.

- Sélection

Comme précédemment, l'indice m_i du régresseur sélectionné est celui pour lequel la valeur absolue du cosinus avec le signal \mathbf{S}^i est maximale ; m_i s'écrit donc :

$$m_i = \underset{k \in [1, N_b] \setminus \{m_1, \dots, m_{i-1}\}}{\operatorname{argmax}} \left(\frac{|\langle \mathbf{S}^i | \mathbf{B}_k^i \rangle|}{\|\mathbf{S}^i\| \cdot \|\mathbf{B}_k^i\|} \right) \quad \text{Eq. 254}$$

Le régresseur $\mathbf{B}_{m_i}^i$ est orthogonal aux $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$.

- Optimisation du régresseur \mathbf{B}_{m_i}

L'optimisation se fait comme précédemment par minimisation de la fonction de coût J par rapport aux paramètres σ_{m_i} , μ_{m_i} et A_{m_i} de \mathbf{B}_{m_i} . J est l'erreur quadratique moyenne entre ce qui reste à modéliser E^i et la gaussienne \mathbf{B}_{m_i} :

$$J(\sigma_{m_i}, \mu_{m_i}, A_{m_i}) = \frac{1}{N_p} \sum_{k=1}^{N_p} \left(E^i(k) - A_{m_i} \exp\left(-\frac{(k - \mu_{m_i,1})^2}{2\sigma_{m_i}^2}\right) \right)^2 \quad \text{Eq. 26}$$

A l'issue de cette optimisation, les paramètres sont $\sigma_{m_i}^*$, $\mu_{m_i}^*$ et $A_{m_i}^*$; la gaussienne optimisée est notée $\mathbf{B}_{m_i}^*$.

- Orthogonalisation

Il est important de noter que, contrairement à $\mathbf{B}_{m_i}^i$ de l'algorithme précédent, le régresseur $\mathbf{B}_{m_i}^{i*}$ n'est pas dans l'orthogonal de l'espace engendré par la famille des vecteurs $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$, car un apprentissage a été réalisé dans l'espace temporel à partir de \mathbf{B}_{m_i} .

Ceci étant, après l'orthogonalisation, on veut exprimer le signal et les régresseurs restants dans un espace orthogonal à l'espace engendré par les $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*}, \mathbf{B}_{m_i}^{1*})$ ^{IV}

À cette étape, il s'agit donc d'orthogonaliser le signal et les régresseurs restants par rapport à la composante du vecteur $\mathbf{B}_{m_i}^{i*}$, notée $\perp \mathbf{B}_{m_i}^{i*}$ ^V, qui se trouve dans l'espace orthogonal aux $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$.

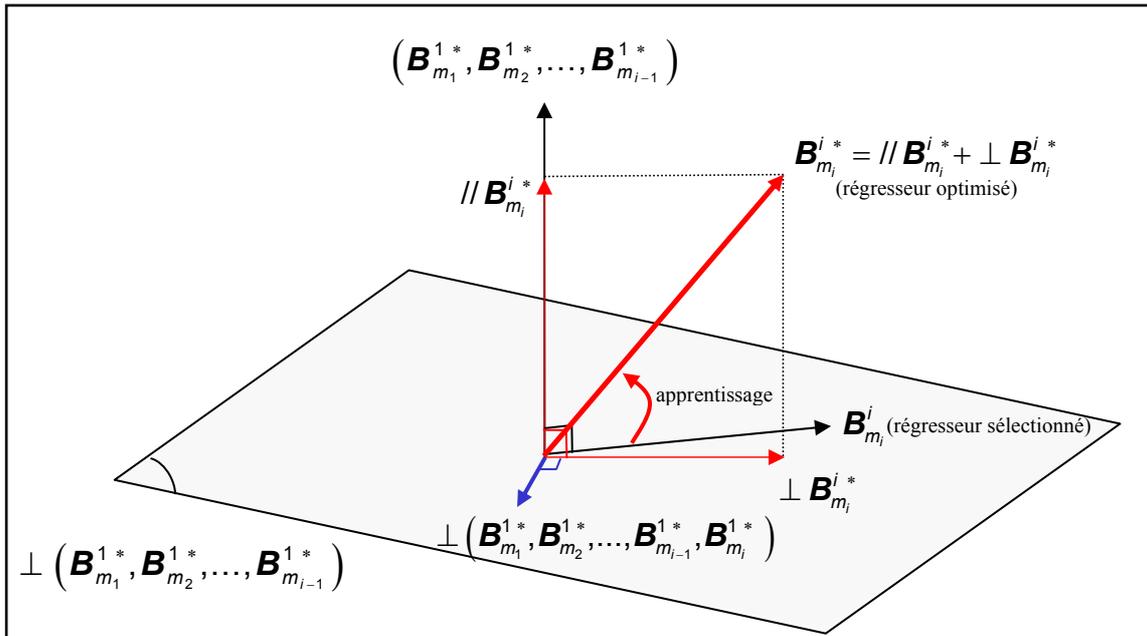


Figure 12: Après $i-1$ itérations de l'algorithme, l'ensemble des régresseurs se trouve dans un espace orthogonal à $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$ noté $\perp (\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$, et la sélection du m_i -ième régresseur se fait dans cet espace. Une fois que le régresseur est sélectionné, ses paramètres sont ajustés pour « représenter » au mieux le signal \mathbf{E}^i ; le régresseur obtenu est noté $\mathbf{B}_{m_i}^{i*}$. Comme il ne se trouve plus dans l'espace $\perp (\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$, on effectue l'orthogonalisation des régresseurs restant par rapport à la composante de $\mathbf{B}_{m_i}^{i*}$ qui se trouve dans l'espace $\perp (\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$, notée $\perp \mathbf{B}_{m_i}^{i*}$. On construit ainsi l'espace orthogonal aux $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*}, \mathbf{B}_{m_i}^{1*})$, nécessaire pour l'itération $i+1$.

^{IV} On remarquera ici qu'il est nécessaire d'exprimer tous les vecteurs dans la même base : la base de départ, dont l'exposant est 1.

^V $\perp \mathbf{B}_{m_i}^{i*}$ s'exprime dans la base orthogonale aux $(\mathbf{B}_{m_1}^{1*}, \mathbf{B}_{m_2}^{1*}, \dots, \mathbf{B}_{m_{i-1}}^{1*})$

La procédure d'orthogonalisation s'effectue alors de manière analogue à ce qui a été vu précédemment :

$$\begin{aligned} \mathbf{B}_k^{i+1} &= \mathbf{B}_k^i - \langle \mathbf{B}_k^i | \perp \mathbf{B}_{m_i}^{i*} \rangle \perp \mathbf{B}_{m_i}^{i*} \quad \forall k = [1, N_b] \setminus \{m_1, \dots, m_i\} \\ \mathbf{S}^{i+1} &= \mathbf{S}^{i+1} - \langle \mathbf{S}^{i+1} | \perp \mathbf{B}_{m_i}^{i*} \rangle \perp \mathbf{B}_{m_i}^{i*} \end{aligned} \quad \text{Eq. 27}$$

- Terminaison

En effectuant N itération de cet algorithme, on obtient N gaussiennes pour modéliser le signal \mathbf{S} . Le modèle Y ainsi obtenu peut s'écrire :

$$Y(t) = \sum_{i=1}^N \mathbf{B}_{m_i}^*(t) \quad \text{Eq. 28}$$

où les $\{\mathbf{B}_{m_i}^*\}_{i=[1,N]}$ sont les gaussiennes de paramètres $\{(\mu_{m_i}^*, \sigma_{m_i}^*, \mathbf{A}_{m_i}^*)\}_{i=[1,N]}$.

II.4.2 Estimation des paramètres du réseau

Comme précédemment, on peut initialiser le réseau de RBF présenté au II.3.3 avec les paramètres $\{(\mu_{m_i}^*, \sigma_{m_i}^*, \mathbf{A}_{m_i}^*)\}_{i=[1,N]}$ et effectuer un apprentissage. En pratique, on remarque que les modifications apportées aux paramètres par ce nouvel apprentissage ne justifient pas le temps qui lui est consacré : les paramètres résultants sont en effet très proches des paramètres initiaux, la variation généralement constatée étant inférieure à 2%.

On se contente donc d'utiliser l'algorithme précédent pour obtenir la décomposition du signal en N gaussiennes.

II.4.3 Résultats de la modélisation

Le résultat de la modélisation du battement précédent avec $N = 10$ gaussiennes est présenté sur la Figure 13. Chaque gaussienne a donc été sélectionnée et optimisée ; un apprentissage final, simultané, des 10 gaussiennes n'a pas été effectué.

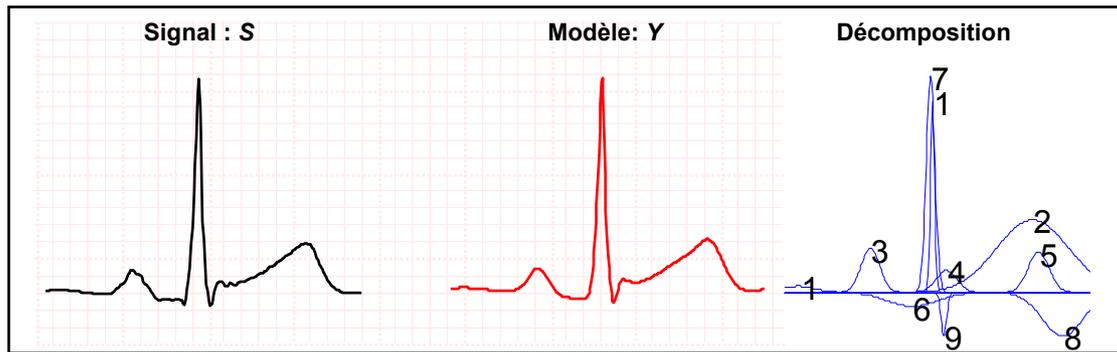


Figure 13 : Modèle obtenu par l'algorithme précédent avec $N=10$ gaussiennes. Ici le modèle est bon au sens de l'erreur quadratique moyenne, et les gaussiennes sont bien situées aux emplacements des ondes caractéristiques. Cependant, on continue à modéliser avec plusieurs gaussiennes une seule onde cardiaque.

Le modèle est un modèle satisfaisant au sens de l'erreur quadratique moyenne, qui est de l'ordre de $3 \cdot 10^{-5}$, et, contrairement au modèle précédent, les gaussiennes sont à présent bien situées, et, de plus, chacune participe à la modélisation d'une onde cardiaque caractéristique. Une décomposition proche de celle proposée ici est décrite dans [Suppappola, 1997], l'auteur qualifie son modèle de modèle « intuitif » de l'ECG. La recherche des gaussiennes utiles à la décomposition n'est cependant pas effectuée par l'algorithme proposé ci-dessus mais par des simples recherches de maxima. Les modèles proposés ici sont très proches de ceux obtenus par l'auteur. Les avantages et inconvénients sont identiques : on continue à modéliser une seule onde avec plusieurs gaussiennes. Par exemple, l'onde R est modélisée avec les gaussiennes 1 et 7, et l'onde T avec les gaussiennes 2, 5 et 8.

L'origine de ce problème réside dans le fait que nous utilisons comme fonction générique une gaussienne, qui est symétrique, alors que les ondes à modéliser sont souvent asymétriques ; il est donc normal d'avoir besoin de plusieurs gaussiennes.

Nous allons montrer que ce problème peut être résolu en remplaçant les gaussiennes par des fonctions appelées « bosses », où une telle asymétrie est paramétrée.

III Régression orthogonale directe généralisée par fonctions bosses

Nous reprenons ici l'algorithme précédent dans lequel les gaussiennes sont remplacées par des fonctions bosses.

III.1 Définition de la bosse

L'idée est donc de disposer d'une fonction paramétrée asymétrique pour modéliser les ondes cardiaques. La fonction que nous présentons ici est constituée de deux demi-gaussiennes reliées par une partie linéaire (Figure 14). Les paramètres de cette fonction sont :

- μ : l'emplacement temporel de la fonction,
- $^1\sigma$: l'écart type de la demi-gaussienne de gauche,
- $^2\sigma$: l'écart type de la demi-gaussienne de droite,
- $^L\sigma$: la longueur de la partie linéaire reliant les demi-gaussiennes,
- A : l'amplitude de la fonction.

Son expression temporelle s'écrit, en fonction de ces paramètres, de la manière suivante :

$$B(t, \mu, ^1\sigma, ^2\sigma, ^L\sigma, A) = \begin{cases} A \exp\left(-\frac{1}{2}\left(\frac{t - (\mu - ^L\sigma/2)}{^1\sigma}\right)^2\right) & \text{si } t \leq \mu - ^L\sigma/2 \\ A & \text{si } \mu - ^L\sigma/2 < t < \mu + ^L\sigma/2 \\ A \exp\left(-\frac{1}{2}\left(\frac{t - (\mu + ^L\sigma/2)}{^2\sigma}\right)^2\right) & \text{si } t \geq \mu + ^L\sigma/2 \end{cases} \quad \text{Eq. 29}$$

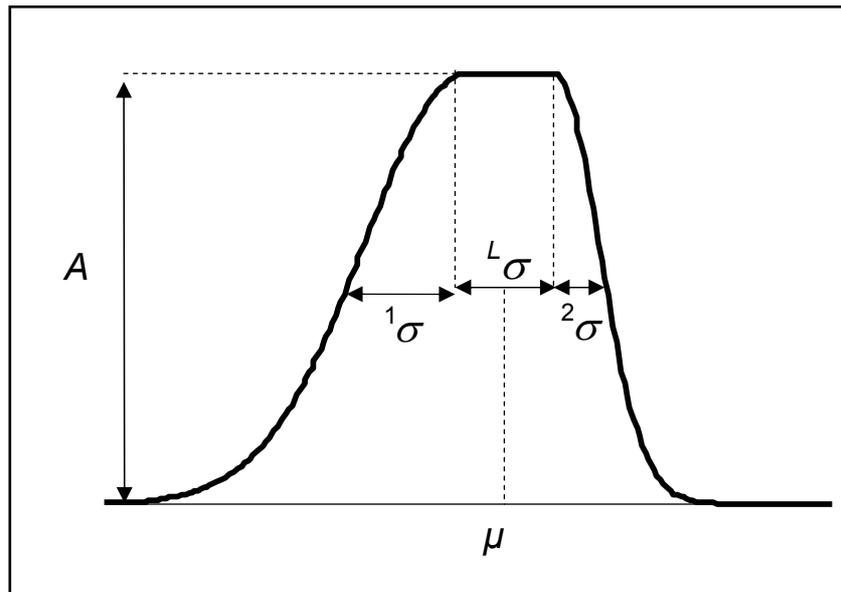


Figure 14 : La fonction bosse est une fonction asymétrique à 5 paramètres ; elle est constituée de deux demi-gaussiennes reliées par une partie linéaire. Elle est continue, et ses dérivées partielles par rapport à ses paramètres sont aussi continues, ce qui est une propriété nécessaire à l'application des algorithmes classiques d'optimisation.

En vue de réaliser l'ajustement de ses paramètres par les algorithmes classiques d'optimisation non linéaire, il est nécessaire que la fonction bosse possède des dérivées partielles continues en chacun des paramètres, ce qui impose les contraintes suivantes :

$${}^1\sigma > 0, {}^2\sigma > 0 \text{ et } {}^L\sigma \geq 0 \quad \text{Eq. 30}$$

Les bosses particulières pour lesquelles ${}^1\sigma = {}^2\sigma$ et ${}^L\sigma = 0$ sont des gaussiennes d'écart-type $\sigma = {}^1\sigma = {}^2\sigma$, et de paramètre μ identique à celui de la bosse.

Disposant d'une nouvelle fonction générique pour la décomposition, nous allons appliquer l'algorithme précédent pour la sélection et l'apprentissage.

III.2 L'algorithme GOFR

L'algorithme présenté précédemment au II.4 permet donc la sélection et l'adaptation de fonctions paramétrées d'une bibliothèque pour modéliser un signal \mathcal{S} , le préalable étant de construire une bibliothèque.

III.2.1 Bibliothèque de bosses

La bibliothèque utilisée est ici composée de bosses symétriques. Elle est, en réalité, identique à celle construite précédemment : les gaussiennes sont des bosses symétriques.

Elle est définie de la manière suivante :

$$B = \left\{ \begin{array}{l} \mathbf{B}(\mu, {}^1\sigma, {}^2\sigma, {}^L\sigma, 1) \\ \mu = j.n/2^k, {}^1\sigma = {}^2\sigma = n/2^{k-2}, {}^L\sigma = 0 \\ j \in [0, 2^k], k \in [1, K_{\max}] \end{array} \right\} \quad \text{Eq. 31}$$

où $\mathbf{B}(\mu, {}^1\sigma, {}^2\sigma, {}^L\sigma, 1)$ est la fonction bosse définie en équation 32, et K_{\max} est tel que le plus petit $\sigma = {}^1\sigma^{\min} = {}^2\sigma^{\min}$ corresponde à la largeur de la plus petite onde caractéristique de l'ECG (nous utilisons dans la pratique $\sigma = {}^1\sigma^{\min} = {}^2\sigma^{\min} = n/2^{K_{\max}-2} = 20$ ms).

III.2.2 Mise en œuvre de l'algorithme

L'algorithme utilisé ici est identique à celui qui a été présenté au II.4.1. Nous allons donc le présenter sous forme d'illustrations des étapes de modélisation du battement \mathbf{S} . Les étapes d'itérations et les équations précédentes étant écrites pour toute famille de régresseurs, elles conservent donc ici tout leur sens.

III.2.2.a Première itération

Au début de la décomposition, le signal est $\mathbf{S} = \mathbf{S}^1$, et la famille de régresseurs est

$B = \mathbf{B}^1 = \{\mathbf{B}_i^1\}_{i=1..N_b}$, représentée sur la Figure 15.

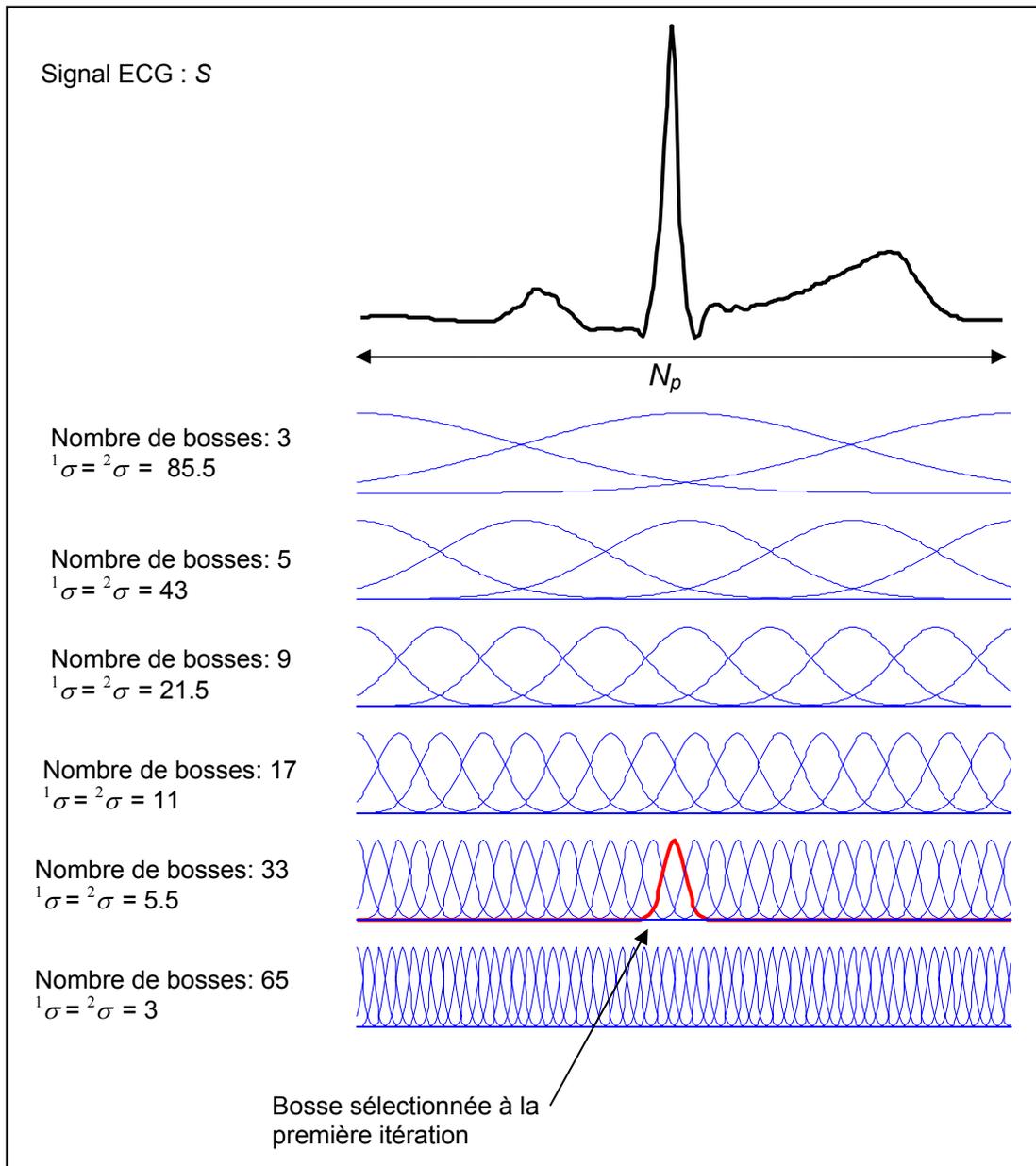


Figure 15 : Signal à modéliser et bibliothèque de bosses symétriques. Cette dernière est composée de 132 bosses pour $N_p=352$. La bosse sélectionnée à la première itération est la bosse numéro 50, représentée en gras sur la Figure.

- Sélection

La première bosse sélectionnée est celle d'indice m_1 tel que :

$$m_1 = \operatorname{argmax}_{k \in [1, N_b]} \left(\frac{\left| \langle \mathbf{S}^1 | \mathbf{B}_k^1 \rangle \right|}{\|\mathbf{S}^1\| \cdot \|\mathbf{B}_k^1\|} \right) \tag{Eq. 32}$$

Ici $m_1 = 50$; la bosse B_{50}^1 est représentée en gras sur les Figures 15 et 16.

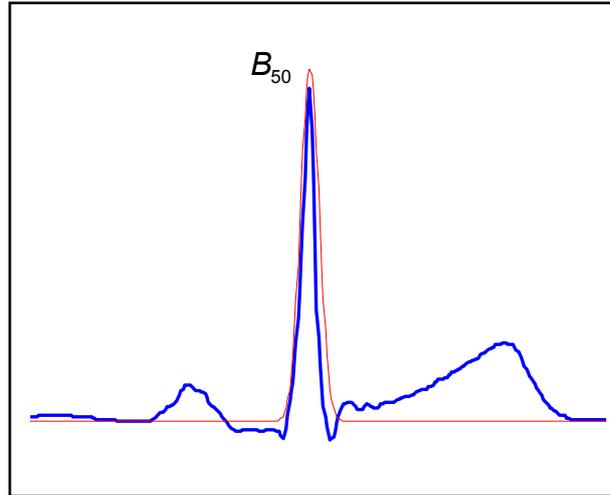


Figure 16 : La première bosse sélectionnée est la bosse numéro 50 de la bibliothèque. À la première itération c'est aussi la fonction de la bibliothèque la plus corrélée avec le signal S .

- Optimisation de la bosse B_{50}

On optimise les paramètres de la bosse $B_{50}(\mu_{50}, {}^1\sigma_{50}, {}^2\sigma_{50}, {}^L\sigma_{50}, A_{50})$ pour qu'elle représente au mieux le signal.

Au début de l'apprentissage, les paramètres sont :

$$\mu_{50} = 166, {}^1\sigma_{50} = {}^2\sigma_{50} = 5,5, {}^L\sigma_{50} = 0 \text{ et } A_{50} = 1.$$

L'optimisation porte sur ces paramètres, qui sont ajustés par minimisation de l'erreur de modélisation, notée E , qui, à la première itération, est le signal lui-même :

$$E^1 = S - 0 = S \quad \text{Eq. 33}$$

L'adaptation du régresseur B_{50} pour qu'il représente au mieux S (ou E^1) serait fondée sur une minimisation de la fonction de coût suivante :

$$J(\mu_{50}, {}^1\sigma_{50}, {}^2\sigma_{50}, {}^L\sigma_{50}, A_{50}) = \frac{1}{N_p} \sum_{k=1}^{N_p} (E^1(k) - B_{50}(k, \mu_{50}, {}^1\sigma_{50}, {}^2\sigma_{50}, {}^L\sigma_{50}, A_{50}))^2 \quad \text{Eq. 34}$$

Notre objectif étant de représenter au mieux la partie de S que B_{50} modélise déjà, on restreint l'intervalle d'apprentissage à la partie du signal où la fonction B_{50} est non nulle.

On définit donc l'intervalle $[\mu_{50} - 3 \sigma_{50}^1 ; \mu_{50} + 3 \sigma_{50}^2]$ en dehors duquel le signal à modéliser est considéré comme nul^{VI}.

On construit alors le signal ${}^{app}E^1$ (Figure 17) tel que :

$${}^{app}E^1(t) = \begin{cases} E^1(t) & \text{si } t \in [\mu_{50} - 3 \sigma_{50}^1 ; \mu_{50} + 3 \sigma_{50}^2] \\ 0 & \text{sinon} \end{cases} \quad \text{Eq. 35}$$

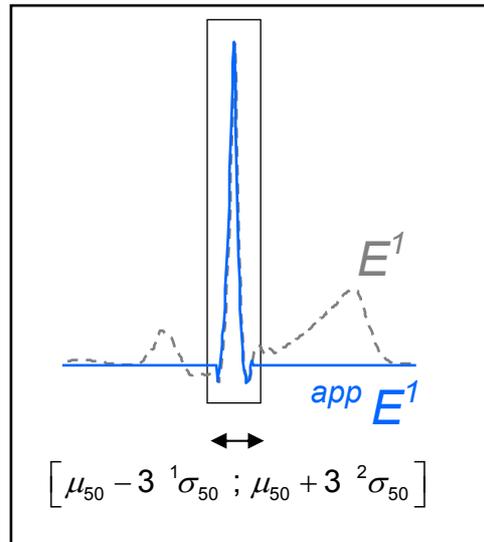


Figure 17 : L'apprentissage de la bosse est effectué pour la fonction temporelle ${}^{app}E^1(t)$ qui est localement égale à $E^1(t)$, mais qui est nulle loin de la bosse sélectionnée. Le principal avantage est de contraindre la bosse sélectionnée à s'adapter à la forme qui a permis sa sélection, et éviter ainsi qu'elle modélise une autre partie du signal.

La fonction de coût à optimiser devient alors :

$$J(\mu_{50}, \sigma_{50}^1, \sigma_{50}^2, \sigma_{50}^L, A_{50}) = \frac{1}{N_p} \sum_{k=1}^{N_p} \left({}^{app}E^1(k) - B_{50}(k, \mu_{50}, \sigma_{50}^1, \sigma_{50}^2, \sigma_{50}^L, A_{50}) \right)^2 \quad \text{Eq. 36}$$

^{VI} Notons que le fait d'imposer les valeurs nulles en dehors de l'intervalle $[\mu_{50} - 3 \sigma_{50}^1 ; \mu_{50} + 3 \sigma_{50}^2]$ procure un avantage supplémentaire : nous « forçons » la bosse à rester sur la ligne de base, et à ne pas modéliser la moyenne du signal.

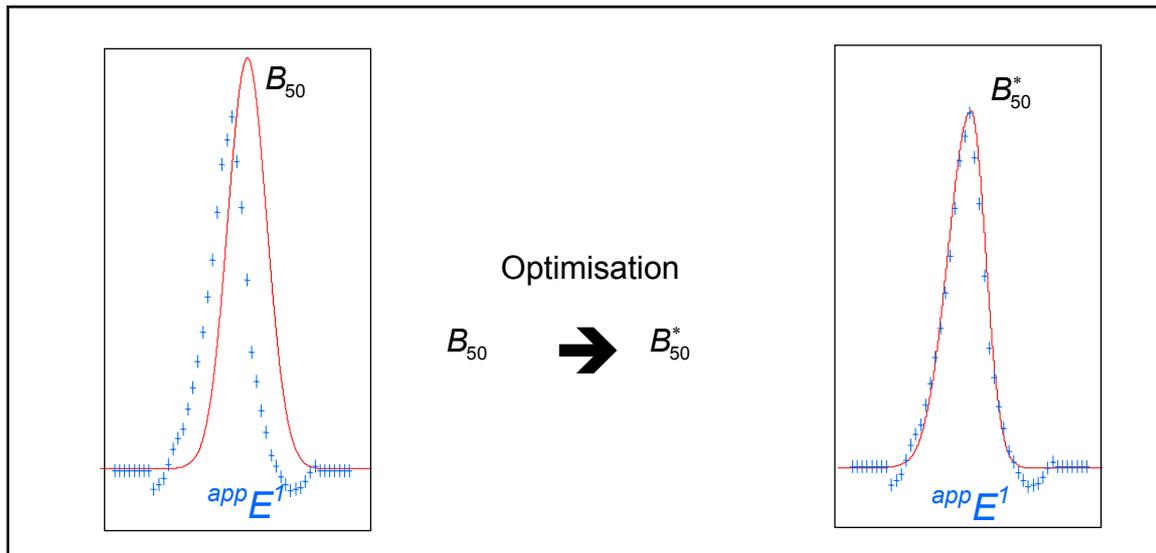


Figure 18 : Les points d'apprentissage pour adapter la bosse sont représentés en bleu. Les points à zéro, en dehors de l'intervalle $[\mu_{50} - 3 \sigma_{50}^1 ; \mu_{50} + 3 \sigma_{50}^2]$ ne sont pas représentés sur cette figure. Après optimisation, la bosse B_{50}^* « colle » au signal d'apprentissage $^{app}E^1$.

L'optimisation est réalisée par 40 pas de gradient simple [Minoux, 1983], et 60 pas de BFGS [Minoux, 1983]. Ces deux optimisations sont réalisées sous les contraintes présentées précédemment (Eq. 30) ; les algorithmes sont détaillés en Annexe B.

Les paramètres en fin d'optimisation de la bosse B_{50}^* sont :

$$\mu_{50}^* = 165,9 ; \sigma_{50}^{1*} = 4,6 ; \sigma_{50}^{2*} = 3,0 ; \sigma_{50}^{L*} = 0 \text{ et } A_{50}^* = 0,8.$$

- Orthogonalisation

La bosse B_{50} est remplacée dans la bibliothèque par la bosse B_{50}^* , et l'orthogonalisation du signal S^1 et du reste de la bibliothèque s'effectue par rapport à cette dernière, qui est considérée comme un vecteur B_{50}^{1*} .

On construit donc le signal S^2 et la famille de régresseur $\{B_i^2\}_{i=[1..N_b] \setminus \{50\}}$ dans l'orthogonal de B_{50}^{1*} d'après les équations 21 et 22.

L'erreur de modélisation à cette étape est :

$$E^2(t) = S(t) - B_{50}^*(t) \tag{Eq. 37}$$

et le modèle :

$$Y(t) = B_{50}^*(t) \quad \text{Eq. 38}$$

III.2.2.b Itérations suivantes et terminaison

On va ainsi sélectionner et apprendre six bosses ($N = 6$) ; ce choix est justifié par le nombre d'ondes caractéristiques cardiaques : 5 (P, Q, R, S et T), ce qui autorise un régresseur pour modéliser du bruit. Pour $N = 6$, le modèle comporte en tout 30 paramètres^{VII}.

Il s'écrit de la manière suivante :

$$Y(t) = \sum_{i=1}^N B_{m_i}^*(t) \quad \text{Eq. 39}$$

La Figure 19 représente le modèle obtenu et la décomposition.

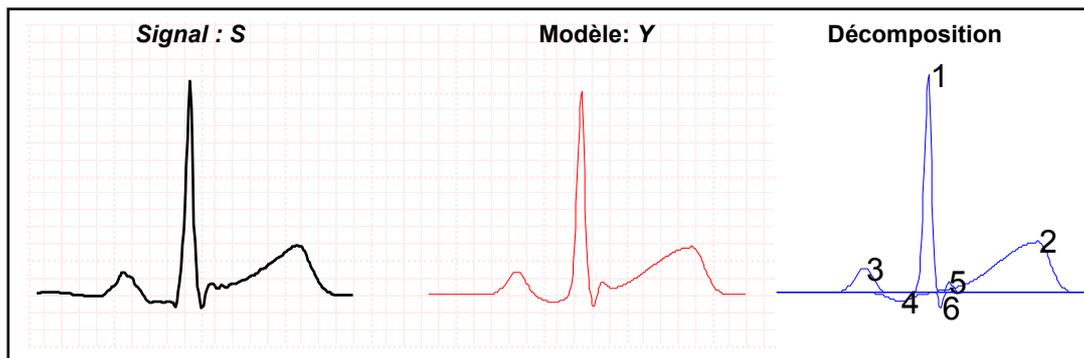


Figure 19 : Le modèle obtenu est très pertinent pour notre étude : chaque régresseur modélise une seule onde cardiaque, et toutes les ondes caractéristiques sont représentées.

Ce modèle est légèrement moins bon au sens de l'erreur quadratique moyenne que le modèle RBF (erreur de l'ordre de 10^{-4}) ; cependant, il est bien plus pertinent pour notre étude. En effet, pour ce battement, *chaque onde caractéristique est modélisée par un unique régresseur*, et 5 des 6 régresseurs modélisent effectivement une onde caractéristique.

Ce type de modèle est tout à fait approprié pour faciliter l'étape suivante, qui consiste en l'étiquetage des ondes.

^{VII} Ce qui justifie a posteriori le choix de retenir 10 gaussiennes (fonctions à 3 paramètres) ou 10 ondelettes (fonctions également à 3 paramètres) dans les modélisations précédentes, de façon à comparer des modèles qui possèdent tous 30 paramètres.

III.3 Résultats de l'algorithme

Il est difficile de présenter des résultats d'ensemble pour cet algorithme de modélisation car le critère de « bonne modélisation » du battement cardiaque est qualitatif. Nous présentons donc ici des exemples de modélisations pour en illustrer la qualité ; les limites de la modélisation apparaîtront d'elles-mêmes dans le chapitre consacré à l'étiquetage. Nous justifions cependant ici l'importance d'une bonne détection de la ligne de base pour la qualité de la modélisation en bosse.

III.3.1 Battements normaux

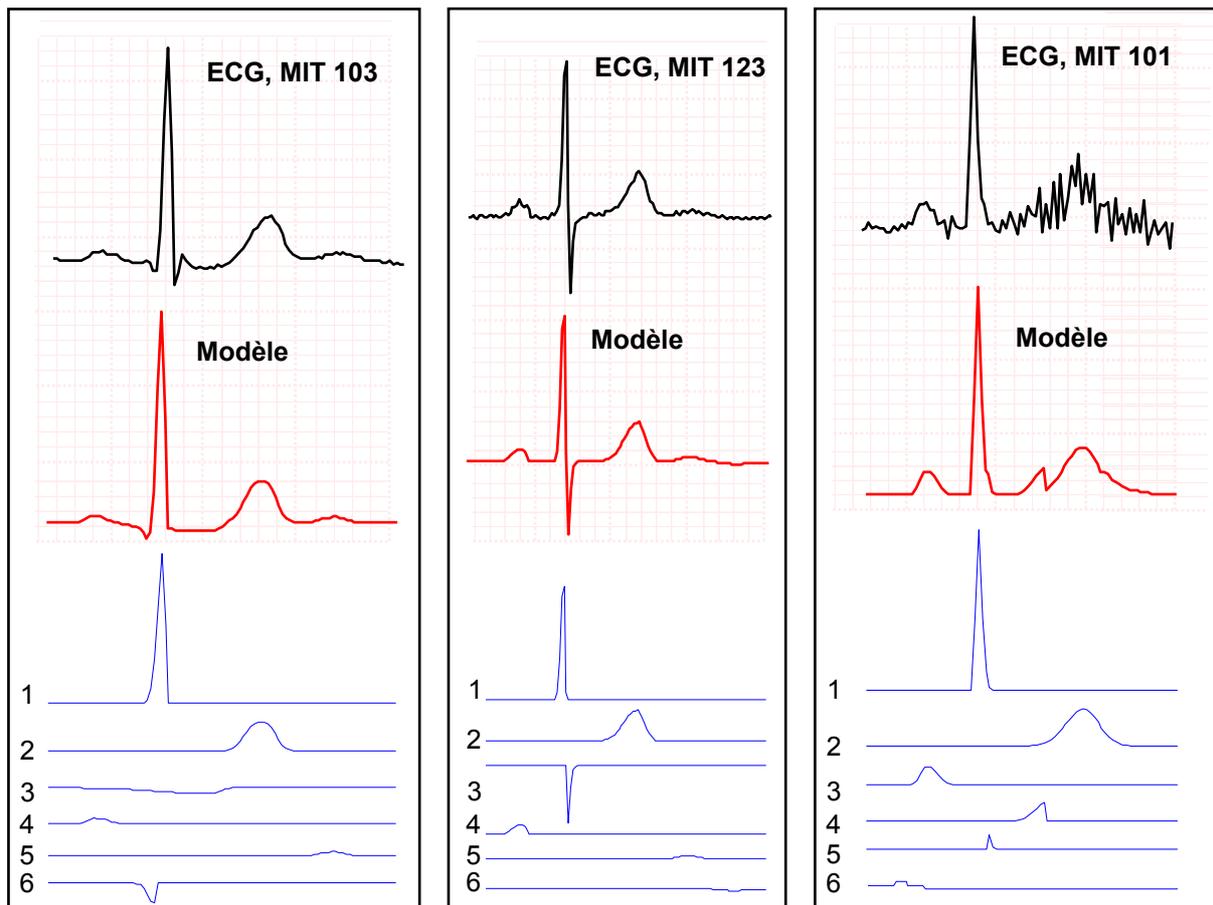


Figure 20 : 3 modélisations de battements normaux choisis dans la base MIT. Le premier ressemble au battement modélisé au paragraphe précédent ; le second est un battement biphasique, dont les ondes R et S sont quasiment de même amplitude ; le troisième est très bruité sur l'onde T par du bruit de haute fréquence. On note ici que pour chacun de ces battements, chaque onde cardiaque est modélisée par un régresseur au moins et l'essentiel du tracé est conservé.

III.3.2 Battements pathologiques

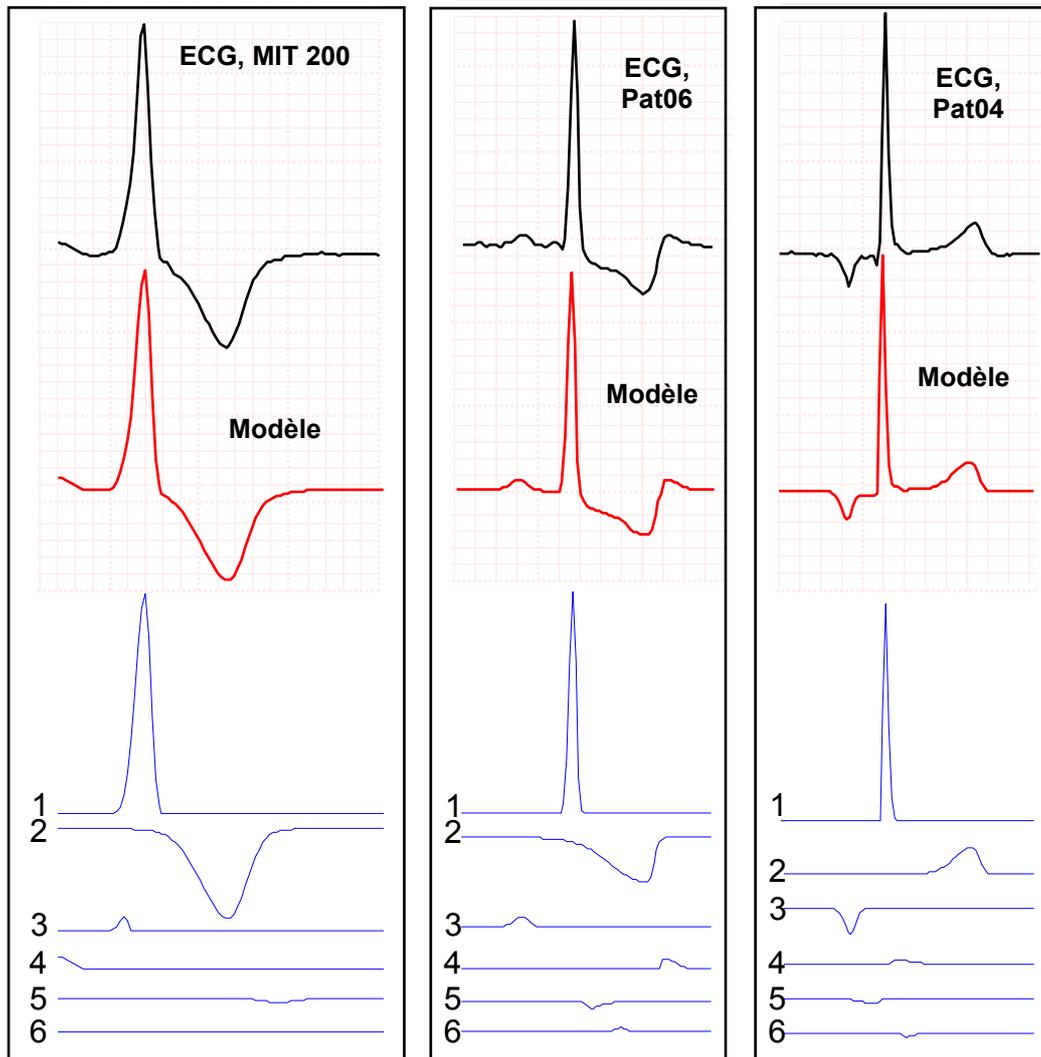


Figure 21: Le premier battement modélisé ici est une extrasystole ventriculaire ; les bosses sélectionnées par l'algorithme correspondent aux différentes ondes cardiaques. Ce type de battement ne comporte que deux ondes d'intérêt. Le deuxième battement présente une anomalie au niveau de l'intervalle entre l'onde S et l'onde T, qui est anormalement décalé vers le bas. Cette anomalie est modélisée ici par une bosse spécifique ce qui permettra son identification ultérieure. Le troisième battement est issu d'un foyer ectopique auriculaire : on remarque en effet que l'onde P est inversée par rapport à un battement normal. Ici encore, cette anomalie est modélisée par une bosse spécifique.

III.3.3 Importance de la ligne de base

Notre objectif, fixé au paragraphe I.2, était d'obtenir une bijection entre les régresseurs utilisés pour modéliser un battement et les ondes caractéristiques de celui-ci. Nous avons vu que l'algorithme développé construit des modèles où cette bi-univocité est quasiment respectée.

La bonne estimation de la ligne de base, par la méthode décrite dans le chapitre 4, est un facteur important pour l'obtention de cette décomposition pertinente. En effet, l'algorithme décompose l'ECG en bosses et en creux (fonctions bosses positives et négatives), à partir d'une ligne de référence. Les cardiologues utilisent la ligne isoélectrique du cœur, pour définir les creux et les bosses, qui possèdent ainsi une signification physiologique : onde P, Q, R S et T.

Les exemples de modélisation présentés Figure 20 et 21 aboutissent à l'objectif recherché car la ligne de base est alignée sur la ligne isoélectrique du cœur : elle est identique à celle qui est utilisée par le cardiologue.

Supposons maintenant que cette ligne soit mal définie, comme indiqué sur la Figure 22 : la première bosse sélectionnée est différente de celle sélectionnée précédemment, cette bosse n'a plus de signification pour notre analyse, et la modélisation qui va suivre ne sera pas de bonne qualité.

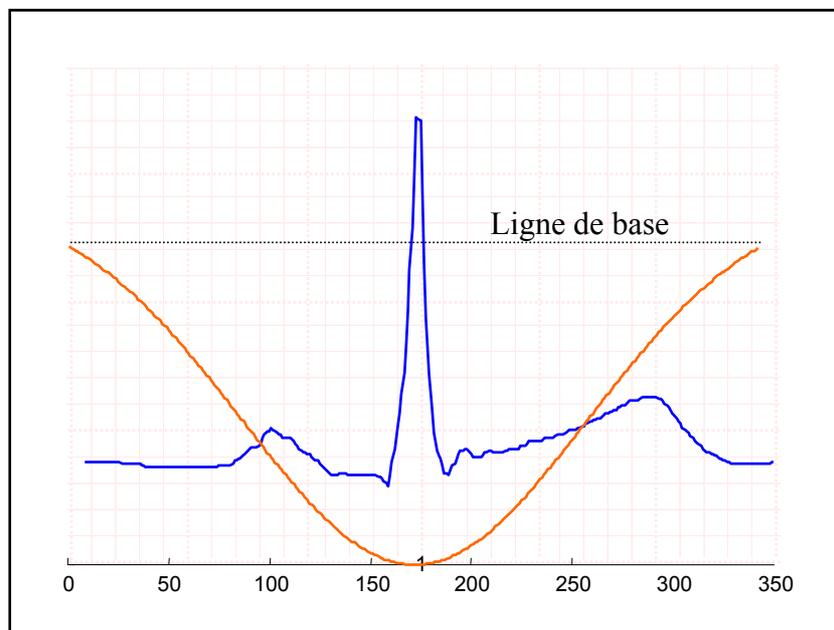


Figure 22 : Comme la ligne de base est particulièrement mal estimée dans ce cas, et que la sélection du régresseur de la bibliothèque se fait implicitement par rapport à cette ligne, le premier régresseur ne représente aucune onde caractéristique, et modélise simplement la moyenne du signal.

L'explication peut aussi être donnée dans la représentation vectorielle de dimension N_p (Figure 23). Désignons comme précédemment par \mathbf{S}^1 le vecteur à modéliser et par $\{\mathbf{B}_i^1\}_{i=[1..N_b]}$ les bosses de la bibliothèque : changer la ligne de base revient à ajouter un vecteur \mathbf{L}_B à \mathbf{S}^1 , et donc à changer les produits scalaires entre le signal et les régresseurs, ce qui modifie donc l'ordre de sélection.

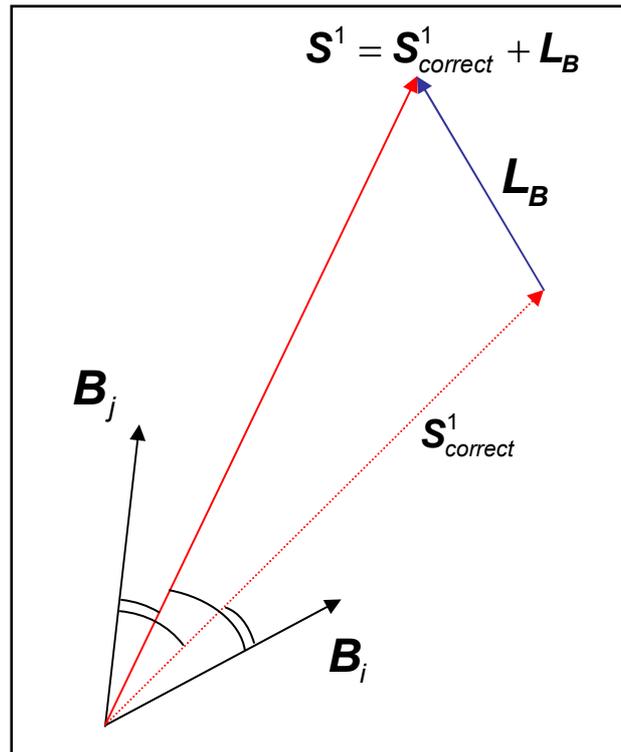


Figure 23 : Une mauvaise estimation de la ligne de base peut s'interpréter dans la représentation vectorielle par l'ajout d'un vecteur \mathbf{L}_B au vecteur $\mathbf{S}^1_{correct}$ représentant le battement pour lequel la ligne de base est correctement estimée. Dans ce cas, les produits scalaires entre \mathbf{S}^1 et les $\{\mathbf{B}_i^1\}_{i=[1..N_b]}$ sont changés, et les régresseurs sélectionnés ne répondent plus à notre objectif. Cet exemple illustre bien, *a posteriori*, l'importance d'une bonne estimation de la ligne de base.

En revanche, quand la ligne de base n'est « pas trop » éloignée de la ligne isoélectrique du cœur, la modélisation compense l'erreur grâce à une bosse, très large, qui fait office de biais.

Le fait que la ligne de base soit un facteur important dans le succès de la modélisation justifie *a posteriori* le temps de calcul qui lui est consacré dans le traitement d'un enregistrement.

Résumé :

Nous avons présenté dans ce chapitre une décomposition mathématique des battements cardiaques sur une famille de régresseurs spécifiques appelés « bosses ».

La bosse est une fonction de 5 paramètres ajustables, qui, contrairement aux régresseurs classiquement utilisés en modélisation (ondelettes, RBF, ...), est construite de manière à « ressembler » au mieux aux ondes caractéristiques du battement cardiaque.

Cette fonction spécifique, associée à un algorithme de décomposition efficace, permet d'associer à chaque onde caractéristique du battement (P, Q, R, S et T) un (ou parfois deux) régresseurs, auquel sera attribuée ensuite une étiquette « médicale ».

La modélisation d'un battement en N bosses nécessite d'effectuer N fois les trois étapes suivantes :

- la sélection d'une bosse dans une bibliothèque préalablement construite,
- l'adaptation des paramètres de la bosse par une optimisation non linéaire sous contraintes pour l'ajuster au mieux au battement cardiaque,
- l'orthogonalisation du reste de la bibliothèque et du signal par rapport à ce régresseur ajusté.

L'efficacité de cette méthode, testée sur les bases de données classiques de l'électrocardiographie, sera démontrée, dans la suite du mémoire, dans le chapitre consacré à l'étiquetage des ondes.

Par ailleurs, la méthode de *sélection adaptative* (GOFR) présentée ici constitue un outil qui convient bien à un traitement du signal fondé sur une approche *cognitive*, dans la mesure où il permet une utilisation performante de régresseurs spécifiques du problème considéré.

