

# Application au traitement du signal radar

Dans les chapitres précédents, nous nous sommes consacré à l’aspect modélisation et nous avons étendu le modèle classique de chaînes de Markov cachées à bruit indépendant à différents modèles prenant en compte diverses propriétés du signal. La validation des résultats a été faite au travers d’expérimentations et les modèles étaient utilisés pour obtenir des segmentations d’image et de signaux. Dans ce chapitre, nous allons voir comment la segmentation peut être utilisée, comme résumé du signal, par une autre application : celle de la détection de cibles dans les signaux radar. Le détecteur original que nous allons présenté utilise les propriétés statistiques du signal reçu ; plus exactement, il compare le signal reçu d’une distance et d’un angle de visée donnés aux signaux provenant du voisinage. Ce détecteur devra connaître les propriétés statistiques du voisinage, l’intérêt de la segmentation sera alors d’obtenir des voisinages homogènes. Nous commençons à rappeler dans ce chapitre les principes du traitement du signal radar.

## 6.1 Prérequis en traitement du signal radar

Un radar est constitué de deux antennes, une antenne émettrice et une antenne réceptrice. Des signaux reçus par cette dernière, on peut déterminer trois attribus : l’azimut, la distance et la vitesse du milieu réfléchissant. Dans le cas d’antennes tournantes, l’azimut correspond à la direction de pointage du radar. Quant à la distance et à la vitesse, des traitements supplémentaires que nous détaillerons, sont nécessaires pour les déterminer. D’autres détails techniques du radar, tels que la formation de faisceaux pour les radars à antennes fixes, sont présentés dans [28, 72, 90].

### 6.1.1 Radar à impulsions

Un radar à impulsions émet des impulsions ayant, chacune, une certaine durée  $\tau$ . Le train d’impulsions est émis pendant une durée appelée “durée de cohérence”, qui sera notée  $T_{\text{coh}}$ . Le temps écoulé entre deux impulsions, appelé “période” ou durée de récurrence, sera noté  $T_{\text{rec}}$ .

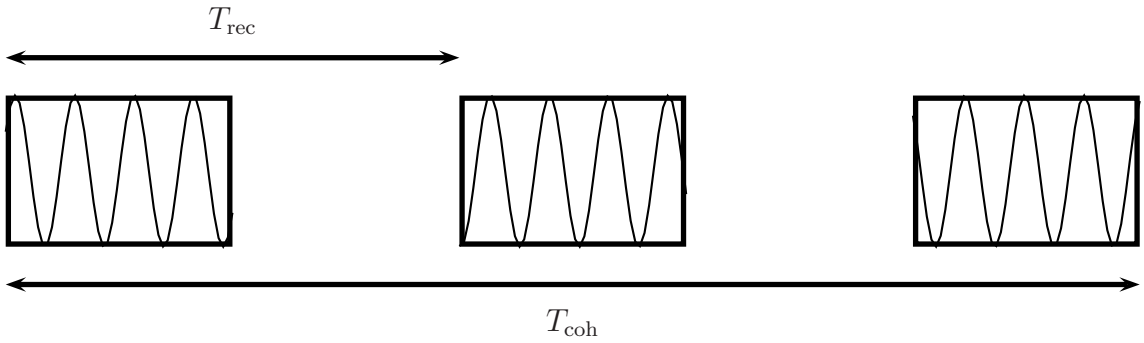


FIG. 6.1 – Schéma d'émission d'un radar à impulsions

Dans le cas du radar à impulsion classique, l'impulsion est portée par un signal sinusoïdal de fréquence  $f_0$  appelée fréquence porteuse. Ainsi, le signal émis complexe  $S_e$  en fonction du temps est donné par :

$$S_e(t) = s_e(t) \exp(2i\pi f_0 t), \text{ où } s_e(t) = \begin{cases} 1 & \text{si } t \in \bigcup_{k=0}^{\frac{T_{\text{coh}}}{T_{\text{rec}}}} [kT_{\text{rec}}, kT_{\text{rec}} + \tau], \\ 0 & \text{sinon.} \end{cases}$$

Chaque signal émis réfléchit sur un certain milieu situé à une distance  $D$  du radar et de vitesse radiale  $v$  par rapport au radar qui sera de signe positif lorsque le milieu se rapproche du radar. On notera que ce milieu peut posséder plusieurs composantes cinétiques ; c'est le cas notamment si le milieu est un fluide. De part l'éloignement au radar, il s'ensuit un décalage en temps entre le signal émis et le signal reçu. Si  $t_0$  est l'instant d'émission du signal, le signal réfléchi correspondant sera reçu à l'instant  $t = t_0 + \Delta t$ , où :

$$\Delta t = \frac{2D}{c},$$

$c$  désignant la vitesse de la lumière.

Le signal réfléchi subit également un décalage en fréquence dû à la vitesse du milieu. Si  $f_0$  est la fréquence porteuse du signal émis, la fréquence porteuse du signal réfléchi est  $f = f_0 + \Delta f$ , où :

$$\Delta f = \frac{2v}{\lambda},$$

$\lambda$  étant la longueur d'onde correspondant à la fréquence  $f_0$ . Ce phénomène physique est appelé "effet Doppler".

Ainsi, le signal reçu à l'instant  $t$  sera la sommation de signaux émis à des instants antérieurs et réfléchis sur des milieux plus ou moins éloignés du radar. La composante du signal reçu correspondante au signal émis à l'instant  $t_0$  a pour expression :

$$A \exp(i\varphi) s_e(t_0) \exp(2i\pi(f_0 + \Delta f)t) + b(t),$$

où  $A \exp(i\varphi)$  est un terme d'atténuation dû au milieu réfléchissant et à la propagation de l'onde et  $b(t)$  est un bruit additif, souvent supposé gaussien. Cette composante a été réfléchi sur un milieu situé à une distance  $D = \frac{c\Delta t}{2}$  et se déplaçant à une vitesse  $v = \frac{\lambda\Delta f}{2}$ .

Du signal reçu, nous avons besoin de connaître la nature du milieu situé à l'instant  $D$ . Pour cela, nous devons isoler la composante correspondante. Ceci se fait grâce à un filtrage appelé "filtrage adaptatif". Notons  $S_r(t)$  le signal complexe reçu à l'instant  $t$ , il est montré dans [72] que la contribution du signal réfléchi par le milieu situé à une distance  $D = \frac{c\Delta t}{2}$  est approximée par chacune des intégrales :

$$S_r(\Delta t, k) = \int_{kT_{\text{rec}}}^{(k+1)T_{\text{rec}}} S_r(u) \bar{S}_e(u - \Delta t, f_0) du, \quad (6.1)$$

où  $k \in \left\{0, \dots, \frac{T_{\text{coh}}}{T_{\text{rec}}} - 1\right\}$ . A  $k$  fixé, la fonction  $\Delta t \rightarrow S_r(\Delta t, k)$  est couramment appelée "chirp". Le calcul de toutes les intégrales  $S_r(\Delta t, k)$  fournit un échantillonnage en temps du signal reçu provenant de la distance  $D$ . Nous l'appellerons "échantillon In-Phase Quadrature" (IQ). Cet échantillon nous est utile pour déterminer le spectre de ce signal. Ce spectre appelé "spectre Doppler", nous permet ensuite de déterminer les composantes cinétiques du milieu. Conformément au théorème d'échantillonnage de Shannon, la largeur de bande du spectre Doppler est égale à l'inverse du pas d'échantillonnage, ainsi toute fréquence située hors de la bande  $\left[f_0 - \frac{1}{2T_{\text{rec}}}, f_0 + \frac{1}{2T_{\text{rec}}}\right]$  ne peut être détectée. Soit  $(z_1, \dots, z_m)$  l'échantillon IQ, le spectre Doppler complexe est donné par :

$$\forall f \in \left[f_0 - \frac{1}{2T_{\text{rec}}}, f_0 + \frac{1}{2T_{\text{rec}}}\right], S_{\text{C}}(f) = \frac{1}{\sqrt{2\pi m}} \sum_{k=1}^m z_k \exp(2i\pi k T_{\text{rec}}(f - f_0)), \quad (6.2)$$

et la puissance spectrale Doppler correspondante par :

$$\forall f \in \left[f_0 - \frac{1}{2T_{\text{rec}}}, f_0 + \frac{1}{2T_{\text{rec}}}\right], S(f) = \frac{1}{2\pi m} \left| \sum_{k=1}^m z_k \exp(2i\pi k T_{\text{rec}}(f - f_0)) \right|^2. \quad (6.3)$$

Les fréquences  $f$  pour lesquelles on détermine cette puissance seront appelées "fréquences Doppler".

Les radars utilisés pour les applications sont des radar à impulsions particuliers appelés "radar à compression d'impulsions". Dans un radar à compression d'impulsion, chaque impulsion émise est modulée en fréquence au lieu d'être un simple signal sinusoïdal. Les éléments physiques vus précédemment restent applicables, cependant, le filtrage adaptatif pour un  $k$  fixé produit un "chirp" qui décroît plus rapidement lorsqu'on s'éloigne de son maximum que dans le cas du radar à impulsion classique. Il s'ensuit un gain en résolution distance. La figure 6.2 présente le filtrage adaptatif d'une impulsion classique et d'une impulsion modulée en fréquence. Dans cet exemple, on suppose que  $S_r(t) = S_e(t - \Delta t)$ , ce qui correspond à la condition idéale d'un unique corps réfléchissant situé à la distance  $D = \frac{c\Delta t}{2}$ . Nous représentons uniquement les parties réelles des chirps.

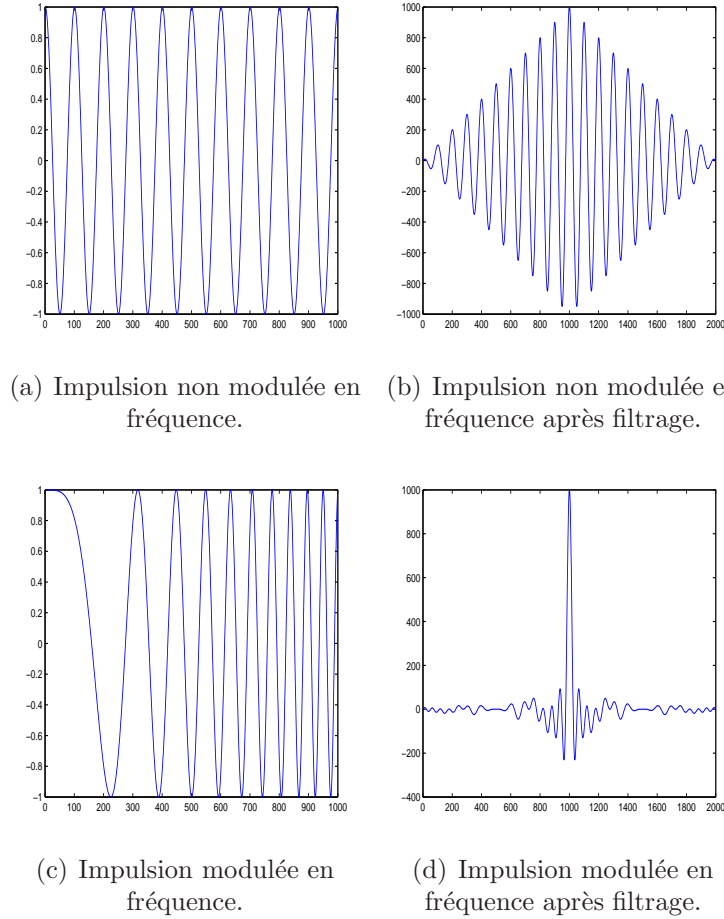


FIG. 6.2 – Filtrage adaptatif de deux types d'impulsion.

### 6.1.2 Principe de la détection TFAC

Le filtrage adaptatif présenté précédemment nous permet de déterminer les puissances reçues en fonction de la distance, ainsi que la répartition de cette puissance dans le spectre Doppler. Cependant, l'intérêt premier d'un radar est de pouvoir détecter automatiquement certains corps réfléchissant appelés "cibles". Ces cibles sont des objets matériels tels que des bateaux ou des êtres humains, ils ont la propriété de réfléchir plus le signal que le milieu environnant. Nous présentons dans cette sous-section, la technique permettant de détecter automatiquement ces cibles. Celle-ci est basée sur les tests d'hypothèses et est appelée "détection à Taux de Fausses Alarmes Constant" (TFAC). Notons ( $H_0$ ) l'hypothèse "absence de cible" et ( $H_1$ ) l'hypothèse "présence de cible". On appellera "case" le triplet (distance, azimut, fréquence Doppler), le couple (distance, azimut) sera appelé case distance-azimut. La case pour laquelle on cherche à déterminer la présence d'une cible sera appelée "case sous test". Le signal reçu d'une case donnée est la quantité  $S_C(f)$  obtenue après filtrage adaptatif et transformée de Fourier pour un azimut, un  $\Delta t$  et une fréquence  $f$  donnés. Soit  $r$  sa partie réelle. Cette quantité suit une certaine loi de probabilité qui diffère selon qu'il y ait une cible ou non. Notons  $p(r|H_0)$  la densité de probabilité de  $r$  en absence de cible et  $p(r|H_1)$  sa densité de probabilité en présence de cible. La probabilité de l'hypothèse  $H_i$  conditionnellement à l'observation  $r$

est calculée par la règle de Bayes :

$$p(Hi|r) \propto p(r|Hi)p(Hi).$$

Nous décidons la présence d'une cible si :

$$\frac{p(H1|r)}{p(H0|r)} > 1,$$

ce qui est équivalent à :

$$\frac{p(r|H1)}{p(r|H0)} > \frac{p(H0)}{p(H1)}.$$

L'a priori  $\lambda = \frac{p(H0)}{p(H1)}$  est déterminé de façon à maintenir la probabilité de fausses alarmes constante. Cette probabilité de fausses alarmes est égale à :

$$\mathbb{P}_{\text{fa}} = \int_{\frac{p(r|H1)}{p(r|H0)} > \lambda} p(r|H0) dr.$$

On définit également la probabilité de détection par :

$$\mathbb{P}_{\text{d}} = \int_{\frac{p(r|H1)}{p(r|H0)} > \lambda} p(r|H1) dr.$$

On remarque que plus  $\lambda$  est petit et mieux on détecte mais on a de fausses alarmes. Il nous faut donc trouver un compromis entre taux de fausses alarmes bas et taux de détection élevé. La qualité d'un détecteur est mesuré par la probabilité de détection en fonction de la probabilité de fausses alarmes. Dans la pratique, on exige un taux de fausses alarmes égal à  $10^{-6}$  et le taux de détection s'échelonne entre 0.5 et 0.9 selon les applications.

Examinons, pour illustrer, le cas où le signal complexe reçu en absence de cible suit une loi normale complexe circulaire centrée de variance  $\sigma_{\mathbb{C}}^2$ , hypothèse faite dans la plupart des cas. La quantité  $\sigma_{\mathbb{C}}^2$  est la puissance du signal reçu en absence de cible. Sous cette condition,  $r$

suit une loi normale réelle centrée de variance  $\sigma^2 = \frac{\sigma_{\mathbb{C}}^2}{2}$ . Ainsi :

– sous (H0) :

$$r = b,$$

où  $b$  suit une loi normale réelle centrée et de variance  $\sigma^2$ ;

– sous (H1) :

$$r = a + b,$$

où  $a \in \mathbb{R}^+$ .

Le rapport  $SNR = \frac{a^2}{2\sigma^2}$  est appelé rapport signal sur bruit (SNR). Dans ce cas, la décision "présence d'une cible" est équivalente à :

$$\frac{r}{\sigma} > \frac{\sigma}{a} \log(\lambda) + \frac{a}{2\sigma}.$$

Par conséquent, pour une probabilité de fausses alarmes et un rapport signal sur bruit donnés, on a :

$$\log \lambda = \sqrt{2SNR}\phi^{-1}(1 - \mathbb{P}_{fa}) - SNR,$$

où  $\phi$  est la fonction de répartition d'une loi normale centrée réduite.

On montre alors que  $\mathbb{P}_d$  est donné en fonction de  $\mathbb{P}_{fa}$  par :

$$\mathbb{P}_d = 1 - \phi\left(\phi^{-1}(1 - \mathbb{P}_{fa}) - \sqrt{2SNR}\right).$$

Dans la pratique, l'écart-type  $\sigma$  est estimé à partir des signaux reçus de cases ayant même azimut et même fréquence Doppler et dont les distances sont voisines de celle de la case sous test. On y estime également la moyenne  $m$  et on retranche au signal reçu de la case sous test cette moyenne. On décide alors la présence d'une cible si  $\frac{|r - m|}{\sigma} >_{H1} \alpha$ , où  $\alpha = \frac{\sigma}{a} \log(\lambda) + \frac{a}{2\sigma}$ . Le détecteur ainsi construit est optimal lorsque le bruit est gaussien. Ainsi, les courbes de performances présentées dans la figure 6.3 sont les meilleures que l'on puisse avoir. Cependant, on remarque que pour un rapport signal sur bruit égal à 2 et une probabilité de fausses alarmes égale à  $10^{-4}$ , la probabilité de détection est inférieure à 0.1. Ainsi, même dans des conditions idéales, les performances de ce détecteur ne sont pas si bonnes. Cette remarque nous a motivé à concevoir un détecteur TFAC original. Dans le détecteur que nous proposons, la détection s'effectue directement sur les échantillons IQ, ces derniers contenant toute la connaissance disponible sur le spectre Doppler.

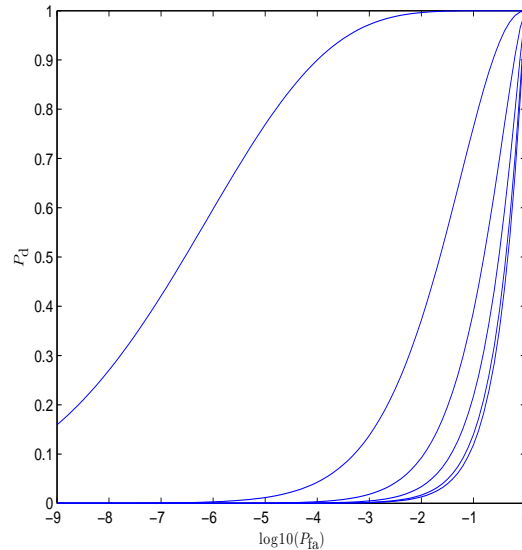


FIG. 6.3 – Courbes de performance  $\mathbb{P}_d$  fonction de  $\log_{10} \mathbb{P}_{fa}$  pour différents rapport signal sur bruit : 0.005, 0.02, 0.125, 0.5, 2 et 12.5 de bas en haut.

## 6.2 Détection à partir des données IQ

Nous présentons dans cette section un détecteur TFAC original. Contrairement au détecteur TFAC classique qui n'utilise uniquement l'intensité rétrodiffusée, celui-ci utilise tout un échantillon IQ obtenu à une distance et un azimut donné. Pour définir notre détecteur, nous nous inspirons de la détection TFAC classique. Le détecteur classique utilise la quantité  $\frac{|r - m|}{\sigma}$  qui s'interprète comme une distance entre le signal  $r$  de la case sous test et la moyenne  $m$  des cases environnantes. Dans notre nouveau détecteur, nous définirons la distance entre l'échantillon IQ observé et la moyenne des échantillons IQ environnants. La notion de distance entre échantillon IQ n'est pas triviale. En effet, la distance entre deux échantillons IQ ayant même densité spectrale devra être nulle. On supposera dans la suite que les échantillons IQ sont réalisations de processus gaussiens complexes circulaires stationnaires du second ordre et dont la famille de covariance est sommable. Ainsi la densité spectrale est la transformée de Fourier de la matrice de covariance de l'échantillon IQ. La distance entre deux échantillons IQ sera donc définie comme la distance entre leur matrice de covariance respective. Il s'agira donc d'une distance entre distributions de probabilité, qui, comme nous le verrons, n'est pas euclidienne. Une fois la notion de distance définie, on définira la notion de moyenne d'une famille de matrices de covariance, nous verrons en particulier que la notion de moyenne dépend de la distance que l'on a définie.

### 6.2.1 Distance entre distributions d'une même famille paramétrique

Considérons une famille de densité de probabilité  $\Lambda = \{y \in \mathcal{Y} \rightarrow p(y; \theta) : \theta \in \Theta\}$  par rapport à une mesure de référence  $\nu$  sur  $\mathcal{Y}$ . L'ensemble des paramètres  $\Theta \subset \mathbb{R}^k$  sera supposé ouvert non vide de  $\mathbb{R}^k$ . Nous allons détailler ici la construction de la distance entre deux distributions  $p(\cdot; \theta_1)$  et  $p(\cdot; \theta_2)$  telle qu'elle a été faite par C. R. Rao [3]. Tout d'abord, il est courant de considérer l'information de Kullback :

$$\bar{K}(\theta_1, \theta_2) = \int \log \left( \frac{p(y; \theta_1)}{p(y; \theta_2)} \right) p(y; \theta_1) d\nu(y),$$

comme une "métrique" entre distributions. Cependant, celle-ci n'est pas symétrique. Afin de définir la distance de Rao, considérons le développement limité de la fonction  $\tilde{\theta} \rightarrow \bar{K}(\theta, \tilde{\theta})$  au voisinage de  $\theta$  :

$$\bar{K}(\theta, \theta + d\theta) = \frac{1}{2} (d\theta)^T I_Y(\theta) d\theta + o(\|d\theta\|^2),$$

où  $I_Y(\theta)$  est la matrice d'information de Fisher. Ainsi, pour deux valeurs de paramètres  $\theta$  et  $\theta + d\theta$  suffisamment proches l'une de l'autre, à une constante près, la distance entre les distributions  $p(\cdot; \theta)$  et  $p(\cdot; \theta + d\theta)$  peut être approximée par  $(d\theta)^T I_Y(\theta) d\theta$ . On remarque que cette forme quadratique est symétrique et elle permet de définir la métrique différentielle :

$$dl = \sqrt{(d\theta)^T I_Y(\theta) d\theta}, \quad (6.4)$$

appelée "distance de Rao". On remarquera en reprenant le raisonnement de la sous-section 1.2.1 du chapitre 1 que cette métrique est invariante par changement de paramétrage  $\eta = g(\theta)$  et ne dépend donc que de l'espace fonctionnel  $\Lambda$ . L'élément  $dl$  sera interprété comme l'élément

de longueur de  $\Lambda$ . Considérons une courbe de dimension 1 dans  $\Lambda$  d'extrémités  $p(\cdot; \theta_1)$  et  $p(\cdot; \theta_2)$ , celle-ci est paramétrée par :

$$t \in [t_1, t_2] \rightarrow \gamma_\Lambda(t) \in \Lambda,$$

où  $\gamma_\Lambda(t_j) = p(\cdot; \theta_j)$ . A cette courbe, on peut lui faire correspondre une courbe  $\theta$  de  $\Theta$  par  $\gamma_\Lambda = \varphi \circ \theta$  où  $\varphi$  est l'application qui à  $\theta \in \Theta$  associe  $p(\cdot; \theta)$ . La longueur de cette courbe est égale à :

$$L_\theta(\theta_1, \theta_2) = \int_{t_1}^{t_2} \sqrt{(\theta'(t))^T I_Y(\theta(t)) (\theta'(t))} dt. \quad (6.5)$$

Toute courbe minimisant l'intégrale (6.5) est appelée géodésique et la distance entre  $p(\cdot; \theta_1)$  et  $p(\cdot; \theta_2)$  est par définition la longueur d'une géodésique, elle est définie par :

$$D(\theta_1, \theta_2) = \min_{\theta} \int_{t_1}^{t_2} \sqrt{(\theta'(t))^T I_Y(\theta(t)) (\theta'(t))} dt. \quad (6.6)$$

On remarquera que nous avons utilisé l'information de Kullback entre deux distributions afin de définir la métrique différentielle de Rao. Dans [9], il y figure différentes mesures de divergence entre lois dont la différentielle seconde donne la même métrique différentielle.

Expliquons maintenant le lien entre la métrique de Rao et les mesures de Jeffreys abordées au chapitre 1, ceci peut justifier l'usage de la métrique de Rao plutôt qu'une autre divergence symétrique entre lois. Lorsque l'espace  $\Lambda$  est muni de l'élément de longueur  $dl$  exprimé dans le paramétrage  $\Theta$  par (6.4), l'élément vectoriel de longueur  $\vec{dl}$  a pour expression :

$$\vec{dl} = \begin{pmatrix} dl_1 \\ \vdots \\ dl_k \end{pmatrix} = (I_Y(\theta))^{\frac{1}{2}} \begin{pmatrix} d\theta_1 \\ \vdots \\ d\theta_k \end{pmatrix}.$$

On en déduit alors que l'élément de volume de  $\Lambda$  a pour expression :

$$d\tau = \sqrt{\det I_Y(\theta)} d\theta_1 \dots d\theta_k.$$

On reconnaît alors la mesure de Jeffreys. Plus exactement, choisir  $\theta$  selon une loi de Jeffreys sur  $\Theta$  est équivalent à choisir la fonctionnelle  $p(\cdot; \theta)$  selon une loi uniforme sur  $\Lambda$ . On retrouve ainsi le caractère non informatif de la loi uniforme. On remarquera d'ailleurs, si  $\Theta$  est discret, l'ensemble  $\Lambda$  l'est aussi et peut être identifié à  $\Theta$ .

Nous allons calculer dans la section suivante la distance entre deux lois gaussiennes centrées et circulaires. On définira la distance entre deux matrices de covariance comme la distance entre les lois gaussiennes centrées circulaires correspondantes.

### 6.2.2 Distance entre lois normales complexes centrées circulaires

Considérons la famille des lois gaussiennes complexes circulaires centrées paramétrées par la matrice de covariance  $\Sigma$ . Dans la suite, cette famille sera confondue avec celle des matrices hermitiennes définies positives. La densité de telles lois est :

$$z \in \mathbb{C}^n \rightarrow \frac{1}{\pi^n \det \Sigma} \exp(-\bar{z}^T \Sigma^{-1} z).$$



En utilisant le développement limité de l'information de Kullback, on montre que la forme quadratique différentielle associée à la métrique de Rao est :

$$dl^2 = \text{tr} (\Sigma^{-1} d\Sigma \Sigma^{-1} d\Sigma).$$

Soient  $\Sigma^{(1)}$  et  $\Sigma^{(2)}$  deux matrices hermitiennes définies positives. Considérons une courbe  $t \in [0, 1] \rightarrow \Sigma(t)$  dans l'espace des paramètres telle que  $\Sigma(0) = \Sigma^{(1)}$  et  $\Sigma(1) = \Sigma^{(2)}$ . La longueur de cette courbe est donnée par :

$$\int_0^1 \sqrt{\text{tr} (\Sigma^{-1}(t) \Sigma'(t) \Sigma^{-1}(t) \Sigma'(t))} dt,$$

et la distance entre deux lois gaussiennes circulaires de covariances respectives  $\Sigma^{(1)}$  et  $\Sigma^{(2)}$  est alors donnée par :

$$D(\Sigma^{(1)}, \Sigma^{(2)}) = \min_{\Sigma} \int_0^1 \sqrt{\text{tr} (\Sigma^{-1}(t) \Sigma'(t) \Sigma^{-1}(t) \Sigma'(t))} dt.$$

Calculons cette distance. Pour cela, on effectue le changement de variable  $\tilde{\Sigma} = (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma (\Sigma^{(1)})^{-\frac{1}{2}}$  où  $\Sigma^\alpha = \exp(\alpha \log(\Sigma))$ . La métrique différentielle s'écrit alors :

$$dl^2 = \text{tr} (\tilde{\Sigma}^{-1} d\tilde{\Sigma} \tilde{\Sigma}^{-1} d\tilde{\Sigma}),$$

et donc  $D(\Sigma^{(1)}, \Sigma^{(2)}) = D(\text{Id}, (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}})$  et  $t \rightarrow \Sigma(t)$  est une géodésique entre  $\Sigma^{(1)}$  et  $\Sigma^{(2)}$  si et seulement si  $t \rightarrow (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma(t) (\Sigma^{(1)})^{-\frac{1}{2}}$  est une géodésique entre  $\tilde{\Sigma}^{(1)} = \text{Id}$  et  $\tilde{\Sigma}^{(2)} = (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}}$ .

Par le changement de variable  $\tilde{\Sigma} = P \Delta \bar{P}^T$ , où  $\Delta$  est diagonale à valeurs propres réelles positives et  $P$  est une matrice orthogonale, on a de même :

$$dl^2 = \text{tr} (\Delta^{-1} d\Delta \Delta^{-1} d\Delta),$$

qui est indépendant de  $t \rightarrow P(t)$ . Soient  $\lambda_j$  les éléments diagonaux de  $\Delta$ , valeurs propres de  $(\Sigma^{(1)})^{-\frac{1}{2}} \Sigma (\Sigma^{(1)})^{-\frac{1}{2}}$  et  $x_j = \log(\lambda_j)$ . On a :

$$dl^2 = \sum_{j=1}^n \left( \frac{d\lambda_j}{\lambda_j} \right)^2 = \sum_{j=1}^n dx_j^2.$$

La métrique différentielle correspond donc à une métrique euclidienne dans l'espace des logarithmes des valeurs propres de  $\Delta$ . On en déduit que la géodésique dans l'espace des valeurs propres est unique et donnée par  $t \rightarrow t \log(\mu_j)$ , où les  $\mu_j$  sont les valeurs propres de  $(\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}}$ .

La distance entre les matrices hermitiennes définies positives  $\Sigma^{(1)}$  et  $\Sigma^{(2)}$  est finalement donnée par :

$$D(\Sigma^{(1)}, \Sigma^{(2)}) = \sqrt{\sum_{j=1}^n (\log(\mu_j))^2}. \quad (6.7)$$

Cette distance entre matrices hermitiennes définies positives a été définie dans un contexte différent par C. L. Siegel [111], nous l'appellerons donc distance de Siegel. De l'expression de la géodésique dans l'espace des logarithmes des valeurs propres, on en déduit que  $\log(\Delta(t)) = t \log(\Delta^{(2)})$  et donc  $\Delta(t) = (\Delta^{(2)})^t$ . Ainsi toute géodésique  $t \rightarrow \tilde{\Sigma}(t)$  a pour expression :

$$\tilde{\Sigma}(t) = P(t)\bar{P}(1)^T \left( \tilde{\Sigma}^{(2)} \right)^t P(1)\bar{P}(t)^T.$$

On en déduit finalement que toute géodésique entre  $\Sigma^{(1)}$  et  $\Sigma^{(2)}$  dans l'espace des matrices hermitiennes définies positives a pour expression :

$$t \in [0, 1] \rightarrow (\Sigma^{(1)})^{\frac{1}{2}} P(t)\bar{P}(1)^T \left[ (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}} \right]^t P(1)\bar{P}(t)^T (\Sigma^{(1)})^{\frac{1}{2}},$$

où  $t \in [0, 1] \rightarrow P(t)$  est une fonction différentiable à valeurs dans l'ensemble des matrices orthogonales telle que les colonnes de  $P(1)$  soient les vecteurs propres de  $(\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}}$ . Dans la suite, nous choisirons la géodésique telle que pour tout  $t$ ,  $P(t) = P(1)$ , ainsi son expression est :

$$t \in [0, 1] \rightarrow (\Sigma^{(1)})^{\frac{1}{2}} \left[ (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}} \right]^t (\Sigma^{(1)})^{\frac{1}{2}}. \quad (6.8)$$

### 6.2.3 Moyenne de matrices hermitiennes définies positives

Nous définissons la moyenne de matrices hermitiennes définies positives, également utile dans notre détecteur TFAC. La moyenne dépend de la distance et est donnée par la définition suivante :

**Définition 6.2.1** (Moyenne de deux éléments d'une variété différentielle). *Soit  $\Lambda$  un espace métrique muni de la distance  $D$ . Une moyenne de deux éléments  $f$  et  $g$  est l'élément  $m = m(f, g)$  d'une géodésique vérifiant :*

$$D(f, m) = D(m, g) = \frac{1}{2} D(f, g).$$

Si la géodésique est unique, la moyenne est unique. Dans le cas contraire, on doit se fixer une géodésique pour définir la moyenne. Reprenons le cas des lois gaussiennes complexes circulaires centrées qui est identifié à l'espace des matrices hermitiennes définies positives. La moyenne de  $\Sigma^{(1)}$  et de  $\Sigma^{(2)}$  relative à la géodésique (6.8) est égale à :

$$M = (\Sigma^{(1)})^{\frac{1}{2}} \left[ (\Sigma^{(1)})^{-\frac{1}{2}} \Sigma^{(2)} (\Sigma^{(1)})^{-\frac{1}{2}} \right]^{\frac{1}{2}} (\Sigma^{(1)})^{\frac{1}{2}}.$$

Dans le cas monovarié, on retrouve la moyenne géométrique. Nous appellerons alors cette moyenne "moyenne géométrique" de deux matrices hermitiennes définies positives.

Pour définir la moyenne de plusieurs éléments, nous généralisons la notion d'isobarycentre. Il y a plusieurs manières de généraliser la notion d'isobarycentre. La première est celle de D. Petz [93] qui utilise la propriété suivante :

**Proposition 6.2.1.** *Soient  $M_1, \dots, M_n$ ,  $n$  points d'un espace vectoriel et soit  $G$  leur isobarycentre. Notant  $G_j$  l'isobarycentre des points  $(M_i)_{i \neq j}$ , alors  $G$  est également isobarycentre des points  $G_1, \dots, G_n$ .*

Bien que l'ensemble des matrices hermitiennes définies positives ne soit pas un espace vectoriel, il est possible d'étendre la définition de barycentre en utilisant cette propriété. Ainsi, il propose la méthode récursive suivante pour le calcul de la moyenne de  $N + 1$  matrices hermitiennes définies positives  $\Sigma^{(1)}, \dots, \Sigma^{(N+1)}$  :

1. soit  $M_N$  l'opérateur qui à  $N$  matrices hermitiennes définies positives associe sa moyenne ;
2. on construit la suite de polygône  $T_k$  suivante :
  - (a)  $T_0 = (\Sigma^{(1)}, \dots, \Sigma^{(N+1)})$  ;
  - (b)  $T_{n+1} = (G_{n+1}^{(1)}, \dots, G_{n+1}^{(N+1)})$ , où  $G_{n+1}^{(j)} = M_N (G_n^{(1)}, \dots, G_n^{(j-1)}, G_n^{(j+1)}, \dots, G_n^{(N+1)})$ .

La suite des polygônes  $T_n$  converge vers un singleton. La moyenne  $M_{N+1}$  est alors définie comme la limite de cette suite.

Cependant, dans la pratique nous ne calculerons pas la moyenne de cette manière, l'algorithme ainsi construit étant de complexité  $\mathcal{O}(N!)$ . T. Ando et R. Mathias [4] propose une autre définition qui utilise l'associativité du barycentre et qui permet de calculer la moyenne pour  $2^n$  matrices. Pour cela, si on sait calculer la moyenne de 2 matrices, la moyenne de 4 matrices est définie comme la moyenne de la moyenne des 2 premières et de la moyenne des 2 dernières et ainsi de suite. Cependant, la moyenne obtenue est différente de la précédente et ne généralise pas convenablement la notion de barycentre.

Nous allons plutôt utiliser la définition de H. Kärcher [62]. Cette définition généralise la propriété suivante :

$$\overrightarrow{GM_1} + \dots + \overrightarrow{GM_N} = 0, \quad (6.9)$$

où  $G$  est l'isobarycentre des points  $M_1, \dots, M_N$ .

Considérons un espace muni d'une métrique différentielle. Un point  $G$  est un barycentre de Kärcher de  $N$  points  $M_1, \dots, M_N$  s'il satisfait la relation :

$$\sum_{k=1}^N \left. \frac{d\gamma_k(t)}{dt} \right|_{t=0} = 0, \quad (6.10)$$

où  $\gamma_k$  est la géodésique de  $G$  à  $M_k$ .

Dans le cas de l'ensemble des matrices hermitiennes définies positives muni de la géodésique définie par (6.8), soit  $G$  l'isobarycentre des  $N$  matrices  $\Sigma^{(1)}, \dots, \Sigma^{(N)}$ , la géodésique  $\gamma_k$  est donnée par :

$$\gamma_k(t) = G^{\frac{1}{2}} \left[ G^{-\frac{1}{2}} \Sigma^{(k)} G^{-\frac{1}{2}} \right]^t G^{\frac{1}{2}}.$$

Ainsi :

$$\left. \frac{d\gamma_k(t)}{dt} \right|_{t=0} = G^{\frac{1}{2}} \log \left( G^{-\frac{1}{2}} \Sigma^{(k)} G^{-\frac{1}{2}} \right) G^{\frac{1}{2}}.$$

On en déduit que la moyenne  $G$  de Kärcher de  $\Sigma^{(1)}, \dots, \Sigma^{(N)}$  vérifie :

$$\sum_{k=1}^N \log \left( G^{-\frac{1}{2}} \Sigma^{(k)} G^{-\frac{1}{2}} \right) = 0.$$

Cette moyenne peut se calculer récursivement de la manière suivante [7] :

1. initialisation de la suite  $G_0$ ;
2. à partir de  $G_m$ , on calcule  $G_{m+1} = G_m^{\frac{1}{2}} \exp \left( -\epsilon \sum_{k=1}^N \log \left( G_m^{-\frac{1}{2}} \Sigma^{(k)} G_m^{-\frac{1}{2}} \right) \right) G_m^{\frac{1}{2}}$ ,  
avec  $-1 < \epsilon < 0$ .

La limite de cette suite est la moyenne de Kärcher.

### 6.2.4 Principe du détecteur TFAC

Dans cette sous-section, nous décrivons notre détecteur TFAC. En chaque case distance-azimut, nous observons des échantillons IQ qui sont des vecteurs complexes. Avant de tester la présence de cible, nous commençons par estimer les matrices de covariances des échantillons. L'estimation de ces matrices de covariance sera expliquée dans la section suivante. Une fois ces matrices estimées, nous calculons les moyennes des matrices des cases environnantes, puis nous calculons la distance de la matrice de la case sous test à cette moyenne. Le test s'écrit alors :

Si  $D(M, \Sigma_{\text{test}}) > \lambda$ , nous décidons la présence d'une cible,

où  $M$  est la moyenne,  $\Sigma_{\text{test}}$  la matrice de covariance sous test et  $\lambda$  un seuil déterminé pour maintenir le taux de fausses alarmes.

Ce détecteur TFAC pourra également utiliser une segmentation du signal obtenue à l'aide d'un modèle de chaînes de Markov cachées. La segmentation bayésienne apparaît ainsi comme un traitement en amont du signal. Une fois cette segmentation obtenue, on considérera uniquement les cases environnantes dans la même classe que la case sous test, les échantillons IQ considérés pour le calcul de la moyenne auront alors des propriétés statistiques voisines.

La section suivante présente la segmentation bayésienne à partir des données observées qui sont les échantillons IQ. Ces derniers étant de grande taille (256 valeurs complexes), nous aurons besoin de réduire l'espace d'observation tout en conservant le maximum d'information. La technique de réduction de l'espace d'observation est également présentée dans la section suivante.

## 6.3 Segmentation et prétraitement des données radar

L'algorithme de Burg présenté dans la sous-section suivante permet de réduire l'espace d'observation afin de pouvoir utiliser les modèles de chaînes de Markov cachées pour la segmentation des données radar. Les vecteurs complexes de taille réduite obtenus par l'algorithme de Burg seront ensuite utilisés pour l'estimation des matrices de covariance des échantillons IQ par l'algorithme de Gohberg-Semencul [35].

### 6.3.1 Algorithme de Burg et estimation des covariances

#### Algorithme de Burg

Soit  $Y = (Y_n)_{n \in \mathbb{Z}}$  un processus centré et stationnaire du second ordre à valeurs dans  $\mathbb{C}$  de covariance complexe  $\gamma(k) = \mathbb{E}(Y_n \bar{Y}_{n-k})$ . L'ensemble des variables aléatoires de carré

intégrable est muni du produit scalaire  $\langle X, Z \rangle = \mathbb{E}(X\bar{Z})$ , on notera  $\|\cdot\|$  la norme associée. On a :

$$Y_n = \underbrace{\left( Y_n + \sum_{k=1}^p a_k^{(p)} Y_{n-k} \right)}_{\text{Orthogonal à } Y_{n-1}, \dots, Y_{n-p}} + \underbrace{\left( -\sum_{k=1}^p a_k^{(p)} Y_{n-k} \right)}_{\text{Proj}(Y_n | Y_{n-1}, \dots, Y_{n-p})},$$

où  $\text{Proj}(Y_n | Y_{n-1}, \dots, Y_{n-p})$  est la projection orthogonale de  $Y_n$  sur le sous-espace vectoriel engendré par  $Y_{n-1}, \dots, Y_{n-p}$ . Les coefficients  $a_k^{(p)}$  sont appelés coefficients auto-régressifs et les coefficients  $\mu_p = a_p^{(p)}$  sont appelés coefficients d'auto-corrélation partielle ou coefficients de réflexion. Ce sont les coefficients de réflexion estimés par l'algorithme de Burg que l'on observera lors de la segmentation, la réalisation de  $Y$  correspond alors aux données IQ d'une case distance-azimut.

Les coefficients de réflexion sont solutions de l'équation matricielle de Yule-Walker suivante :

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \dots & \dots & \gamma(p-1) \\ \bar{\gamma}(1) & \gamma(0) & \gamma(1) & \dots & \gamma(p-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \dots & \dots & \dots & \bar{\gamma}(1) & \gamma(0) \end{pmatrix} \times \begin{pmatrix} a_p^{(p)} \\ \vdots \\ a_1^{(p)} \end{pmatrix} = - \begin{pmatrix} \gamma(p) \\ \vdots \\ \gamma(1) \end{pmatrix}. \quad (6.11)$$

Ces équations peuvent se résoudre par l'algorithme de Durbin-Levinson suivant :

1. initialisation :

$$\mu_1 = a_1^{(1)} = -\frac{\gamma(1)}{\gamma(0)},$$

$$\epsilon_1 = \stackrel{\text{def}}{=} \|Y_2 - \text{Proj}(Y_2 | Y_1)\|^2 = \gamma(0) (1 - |\mu_1|^2) ;$$

2. itération :

$$a_{p+1}^{(p+1)} = \mu_{p+1} = -\frac{\gamma(p+1) + \sum_{k=1}^p a_k^{(p)} \gamma(p+1-k)}{\epsilon_p},$$

$$\epsilon_{p+1} = \stackrel{\text{def}}{=} \|Y_{p+2} - \text{Proj}(Y_{p+2} | Y_1, \dots, Y_{p+1})\|^2 = \epsilon_p (1 - |\mu_{p+1}|^2),$$

et pour  $k \in \{1, \dots, p\}$ ,

$$a_k^{(p+1)} = a_k^{(p)} + \mu_{p+1} \bar{a}_{p+1-k}^{(p)}. \quad (6.12)$$

Contrairement à l'algorithme de Durbin-Levinson, l'algorithme de Burg ne nécessite pas de connaître les covariances. De plus, ces dernières seraient mal estimées à partir des échantillons IQ, ceux-ci n'étant pas assez grands. Cependant, l'algorithme de Burg ne calcule pas les coefficients de réflexion de manière exacte mais les estime par une méthode semblable à celle des moindres carrés.

Détaillons l'algorithme de Burg. Soit  $(y_1, \dots, y_M)$  un échantillon IQ en une case distance-azimut donnée. On définit les erreurs de filtrage (directes) et de lissage (rétrogrades) pour  $1 \leq p \leq M - 1$  par :

$$f_p(n) = y_n - \text{Proj}(y_n | y_{n-1}, \dots, y_{n-p}) = y_n + \sum_{k=1}^p a_k^{(p)} y_{n-k},$$

$$b_p(n) = y_{n-p} - \text{Proj}(y_{n-p} | y_{n-p+1}, \dots, y_n) = y_{n-p} + \sum_{k=1}^p \bar{a}_k^{(p)} y_{n-p+k}.$$

L'algorithme de Burg consiste à minimiser :

$$U^{(p)} = \sum_{n=p+1}^M [|f_p(n)|^2 + |b_p(n)|^2].$$

Pour  $p = 1$ , on doit chercher  $\mu_1 = a_1^{(1)}$  minimisant :

$$U^{(1)} = \sum_{n=2}^M [|y_n + \mu_1 y_{n-1}|^2 + |y_{n-1} + \bar{\mu}_1 y_n|^2],$$

$\mu_1$  est alors donné par :

$$\mu_1 = -2 \frac{\sum_{n=2}^M y_n \bar{y}_{n-1}}{\sum_{n=2}^M (|y_{n-1}|^2 + |y_n|^2)}. \quad (6.13)$$

En utilisant la formule (6.12), on montre que :

$$f_{p+1}(n) = f_p(n) + \mu_{p+1} b_p(n-1),$$

$$b_{p+1}(n) = b_p(n-1) + \bar{\mu}_{p+1} f_p(n),$$

ainsi  $\mu_{p+1}$  doit minimiser :

$$U^{(p+1)} = \sum_{n=p+2}^M [|f_p(n) + \mu_{p+1} b_p(n-1)|^2 + |b_p(n-1) + \bar{\mu}_{p+1} f_p(n)|^2],$$

il est alors donné par :

$$\mu_{p+1} = -2 \frac{\sum_{n=p+2}^M f_p(n) \bar{b}_p(n-1)}{\sum_{n=p+2}^M (|f_p(n)|^2 + |b_p(n-1)|^2)}. \quad (6.14)$$

Ainsi l'algorithme de Burg fonctionne de la manière suivante :

1. initialisation :

- calcul de  $\mu_1$  en utilisant la formule (6.13) ;
- calcul des erreurs de filtrage et de lissage pour  $n \geq 2$  :
 
$$f_1(n) = y_n + \mu_1 y_{n-1},$$

$$b_1(n) = y_{n-1} + \bar{\mu}_1 y_n ;$$

2. itération :

- calcul de  $\mu_{p+1}$  en utilisant (6.14) ;
- calcul des erreurs de filtrage et de lissage pour  $n \geq p + 2$  :
 
$$f_{p+1}(n) = f_p(n) + \mu_{p+1} b_p(n-1),$$

$$b_{p+1}(n) = b_p(n-1) + \bar{\mu}_{p+1} f_p(n) ;$$

3. estimation de la variance  $\gamma(0)$  puis calcul des erreurs de prédictions :

$$\epsilon_1 = \gamma(0) (1 - |\mu_1|^2),$$

$$\epsilon_{p+1} = \epsilon_p (1 - |\mu_{p+1}|^2) ;$$

4. calcul des  $a_k^{(p)}$  par (6.12).

La variance  $\gamma(0)$  correspond à la puissance reçue de la case distance-azimut et les coefficients de réflexion caractérisent la forme du spectre Doppler.

On peut montrer que si  $\mu_{p+1} = 0$ , alors pour tout  $k \geq p + 1$ ,  $\mu_k = 0$ , le processus  $Y$  est alors un processus auto-régressif d'ordre  $p$  (AR(p)). Dans la pratique, on imposera un ordre  $p$  et on considérera que pour  $k \geq p + 1$ ,  $\mu_k = 0$ . Dans ce cas, pour tout  $n \geq p$ , le processus  $Y$  satisfait pour  $n \geq p$  :

$$\text{Proj}(Y_n | Y_{n-1}, \dots, Y_{n-p}) = \text{Proj}(Y_n | (Y_m)_{m \leq n-1}),$$

et  $B_n = Y_n - \text{Proj}(Y_n | (Y_m)_{m \leq n-1})$  est une séquence indépendante centrée et de variance  $\epsilon_p$  appelée bruit d'innovation. Comme :

$$B_n = Y_n + \sum_{k=1}^p a_k^{(p)} Y_{n-k},$$

alors la densité spectrale de  $Y$  correspondant au spectre Doppler s'écrit :

$$\forall f \in \left[ f_0 - \frac{1}{2T_{\text{rec}}}, f_0 + \frac{1}{2T_{\text{rec}}} \right], S(f) = \frac{\epsilon_p}{2\pi} \times \frac{1}{\left| 1 + \sum_{k=1}^p a_k^{(p)} \exp(2ik\pi T_{\text{rec}}(f - f_0)) \right|^2}. \quad (6.15)$$

Dans la pratique, il se peut que l'ordre  $p$  imposé soit trop petit et que  $\mu_{p+1}$  ait une valeur non négligeable. Ainsi, dans [6], il est proposé un algorithme de Burg régularisé de façon à imposer une décroissance rapide des  $\mu_p$ . Dans cette nouvelle version de Burg, la quantité  $U^{(p)}$  à minimiser est remplacée par :

$$E^{(p)} = U^{(p)} + \frac{1}{N-p} \left[ \gamma_0 \int_{-\frac{1}{2}}^{\frac{1}{2}} |A^{(p)}(\lambda)|^2 d\lambda + \gamma_1 \int_{-\frac{1}{2}}^{\frac{1}{2}} \left| \frac{dA^{(p)}}{d\lambda}(\lambda) \right|^2 d\lambda \right],$$

où  $\gamma_0$  et  $\gamma_1$  sont deux coefficients de régularisation et

$$A^{(p)}(\lambda) = 1 + \sum_{k=1}^p a_k^{(p)} e^{2i\pi k\lambda}.$$

Le calcul des coefficients de réflexion s'effectue de manière similaire et est détaillé dans [6].

Dans les expérimentations, nous prendrons  $p = 10$ ,  $\gamma_0 = 0$  et  $\gamma_1 = 10000$ .

Nous voyons ainsi que l'algorithme de Burg permet de transformer les vecteurs IQ de chaque case distance-azimut en vecteurs complexes de taille réduite égale à  $p$ .

A partir des coefficients de réflexion, il est possible de calculer la famille de covariance grâce à l'algorithme de Gohberg-Semencul que nous présentons dans le paragraphe suivant.

### Algorithme de Gohberg-Semencul

Nous présentons maintenant le calcul de la covariance  $\gamma(k) = \mathbb{E}(Y_n \bar{Y}_{n-k})$  à partir des coefficients de réflexion  $(\mu_1, \dots, \mu_p)$ , de la puissance spectrale  $\gamma(0)$ , des erreurs de prédictions  $(\epsilon_1, \dots, \epsilon_{p-1})$  et des coefficients auto-régressifs. Ce calcul s'effectue récursivement de la manière suivante :

1. initialisation :

$$\gamma(1) = -\gamma(0)\mu_1 ;$$

2. itération :

pour  $m \leq p - 1$  :

$$\gamma(m+1) = - \left[ \epsilon_m \mu_{m+1} + \sum_{k=1}^m a_k^{(m)} \gamma(m+1-k) \right],$$

et pour  $m \geq p$  :

$$\gamma(m+1) = - \sum_{k=1}^p a_k^{(p)} \gamma(m+1-k).$$

La connaissance de  $(\gamma(0), \dots, \gamma(p))$  suffit pour déterminer les coefficients auto-régressifs jusqu'à l'ordre  $p$  grâce aux formules de Durbin-Levinson. Remarquant alors que pour  $m \geq p$ , le calcul de  $\gamma(m+1)$  n'utilise uniquement les coefficients auto-régressifs jusqu'à l'ordre  $p$ , ainsi la connaissance de  $(\gamma(0), \dots, \gamma(p))$  suffit pour déterminer toute la covariance. Ainsi, dans les expérimentations, lorsque l'on devra calculer les moyennes de matrices et la distance entre matrices, nous n'utiliserons uniquement les matrices de covariance suivantes :

$$\Sigma = \begin{pmatrix} \gamma(0) & \dots & \gamma(p) \\ \vdots & \vdots & \vdots \\ \bar{\gamma}(p) & \dots & \gamma(0) \end{pmatrix}.$$



### 6.3.2 Modèles CMC utilisés

Nous précisons ici le modèle que nous allons utiliser pour la segmentation. En chaque case distance-azimut  $(m, n)$ , nous observons un vecteur de coefficients de réflexion  $\mu^{(m,n)} = (\mu_1^{(m,n)}, \dots, \mu_p^{(m,n)})$  ainsi que la puissance  $r^{(m,n)} = \gamma^{(m,n)}(0)$ . L'observation est donc la réalisation  $(\mu^{(m,n)}, r^{(m,n)})_{(m,n) \in \mathcal{S}}$  d'un champ aléatoire bi-dimensionnel indexé sur  $\mathcal{S} = \{1, \dots, N_D\} \times \{1, \dots, N_A\}$ , où  $N_D$  est le nombre de distances et  $N_A$  le nombre d'azimuts considérés. Ce processus est transformé en processus mono-dimensionnel de taille  $N = N_D \times N_A$  en utilisant un parcours azimut par azimut. La réalisation du processus mono-dimensionnel est donc :

$$((\mu^{(1,1)}, r^{(1,1)}), (\mu^{(2,1)}, r^{(2,1)}), \dots, (\mu^{(N_D,1)}, r^{(N_D,1)}), (\mu^{(1,2)}, r^{(1,2)}), \dots, (\mu^{(N_D, N_A)}, r^{(N_D, N_A)})).$$

On notera désormais  $(\mu^{(n)}, r^{(n)})$  la  $n^{\text{ième}}$  marginale de cette réalisation mono-dimensionnelle. Cette réalisation est celle du processus observé que l'on notera  $Z = (\mu^{(n)}, r^{(n)})_{1 \leq n \leq N}$ . Le processus caché sera noté  $X = (X_n)_{1 \leq n \leq N}$  et chaque  $X_n$  prendra ses valeurs dans un ensemble fini  $\mathcal{X} = \{\omega_1, \dots, \omega_K\}$ . Le modèle considéré est celui des chaînes de Markov cachées à bruit indépendant. La distribution de  $(X, Z)$  sera alors donnée par :

$$p(x_{1:N}, z_{1:N}) = p(x_1)p(z_1|x_1) \prod_{n=1}^{N-1} p(x_{n+1}|x_n)p(z_{n+1}|x_{n+1}).$$

Quant à la loi d'observation  $p(z_n|x_n)$ , elle sera donnée par :

$$p(z_n|x_n) = p(r^{(n)}|x_n) \times \prod_{k=1}^p p\left(\left|\mu_k^{(n)}\right| \mid x_n\right) p\left(\arg\left(\mu_k^{(n)}\right) \mid x_n\right),$$

où  $p(r^{(n)}|x_n = \omega_j)$  est la distribution de l'inverse d'une loi gamma de paramètres  $a_j$  et  $b_j$  :

$$p(r^{(n)}|x_n = \omega_j) = \frac{1}{\Gamma(a_j)} \frac{1}{b_j^{a_j}} \frac{1}{y^{a_j} + 1} \exp\left(-\frac{1}{b_j y}\right), \quad (6.16)$$

$p\left(\left|\mu_k^{(n)}\right| \mid x_n = \omega_j\right)$  est la distribution d'une loi gamma de paramètres  $\alpha_j$  et  $\beta_j$  et

$p\left(\arg\left(\mu_k^{(n)}\right) \mid x_n = \omega_j\right)$  est la distribution d'une loi de Von Mises Fisher de paramètre de direction  $\mu_{0,j}$  et de concentration  $\kappa_j$ .

## 6.4 Expérimentations

Nous commençons cette section par comparer notre détecteur TFAC avec le détecteur TFAC classique. Dans un second temps, nous utiliserons la segmentation obtenue par le modèle de chaînes de Markov cachées dans notre algorithme de détection. Les données traitées sont issues du radar HF Wera utilisé par la société Actimar (Brest) dans le cadre du PREI Decimall en collaboration avec Thalès et financé par la DGA. Ce radar a une résolution de 1 kilomètre par case distance et nous disposons de 120 distances. La résolution en azimut est de 0.02 radians et nous disposons de 32 azimuts consécutifs.

### 6.4.1 Comparaison qualitative des deux détecteurs

Pour comparer les deux détecteurs, nous n'utiliserons pas de courbes  $(\mathbb{P}_{fa}, \mathbb{P}_d)$ , étant donné que pour estimer la probabilité de détection pour une probabilité de fausses alarmes égale à  $10^{-6}$ , il faudrait plus de 1000000 cases distance-azimut et les algorithmes de détection ne pourraient fonctionner à cause du manque de mémoire vive. On va plutôt regarder le gain en terme de distance pour les cases où nous savons qu'il y a une cible. Pour les deux détecteurs, nous calculerons les spectres complexes et la densité spectrale pour 512 fréquences consécutives dans  $\left[ f_0 - \frac{1}{2T_{rec}}, f_0 + \frac{1}{2T_{rec}} \right]$ .

Détaillons comment nous procédons pour les deux détecteurs.

– Cas du TFAC classique :

1. soit  $r(n, \theta, k)$  le signal réel reçu de la case de distance  $n \in \{1, \dots, 120\}$ , d'azimut  $\theta \in \{1, \dots, 32\}$  et de fréquence  $k \in \{1, \dots, 512\}$ . Calculer les moyennes  $m(n, \theta, k)$  et écart type  $\sigma(n, \theta, k)$  des signaux reçus des cases, dites environnantes, de distance  $n' \in \{n - 5, \dots, n - 10\} \cup \{n + 5, \dots, n + 10\}$ , d'azimut  $\theta$  et de fréquence  $k$ . Puis calculer les distances  $\frac{|r(n, \theta, k) - m(n, \theta, k)|}{\sigma(n, \theta, k)}$ ;
2. calculer, pour chaque case distance-azimut, la quantité

$$D_{TFAC}(n, \theta) = \max_{1 \leq k \leq 512} \frac{|r(n, \theta, k) - m(n, \theta, k)|}{\sigma(n, \theta, k)},$$

qui sera appelée par la suite "distance". Le test de présence d'une cible s'écrit :

$$D_{TFAC}(n, \theta) > \alpha.$$

– Cas du TFAC amélioré :

1. pour chaque case de distance  $n$  et d'azimut  $\theta$ , calculer les moyennes des matrices de covariance des échantillons IQ provenant des cases, dites environnantes, d'azimut  $\theta$  et de distance  $n' \in \{n - 5, \dots, n - 10\} \cup \{n + 5, \dots, n + 10\}$ ;
2. calculer ensuite la distance de Siegel  $D_{TFAC}^{Siegel}(n, \theta)$  de la matrice de covariance issue de la case  $(n, \theta, k)$  à la moyenne correspondante.

Afin de pouvoir comparer correctement les deux détecteurs, nous nous donnons une case de référence de distance  $n_0$  et d'azimut  $\theta_0$ . Nous divisons ensuite chaque distance  $D_{TFAC}(n, \theta)$  (resp.  $D_{TFAC}^{Siegel}(n, \theta)$ ) par  $D_{TFAC}(n_0, \theta_0)$  (resp.  $D_{TFAC}^{Siegel}(n_0, \theta_0)$ ). Nous choisissons pour case de référence celle située à 57 kilomètres du radar et ayant un azimut de  $13.13^\circ$  avec l'axe de visée du radar. De la figure 6.6, nous voyons qu'en utilisant le détecteur classique, de nombreuses valeurs atypiques apparaissent ce qui peut générer des fausses alarmes. Nous avons pu mettre en évidence grâce aux vérités terrain la présence de deux cibles, la première est située à une distance de 54 kilomètres et un azimut de  $-9.78^\circ$  et la deuxième est située à une distance de 53 kilomètres et un azimut de  $-67.08^\circ$ . Les figures 6.7 et 6.8 représentent les distances de Siegel et les distances obtenues par le détecteur TFAC classique après normalisation pour les azimuts  $-9.78^\circ$  et  $-67.08^\circ$ .

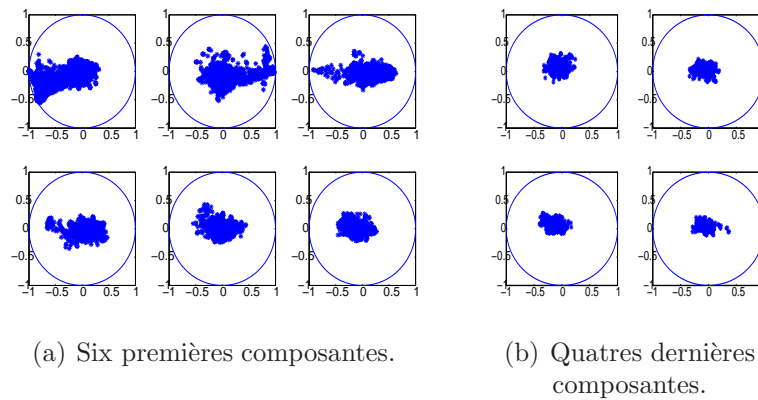
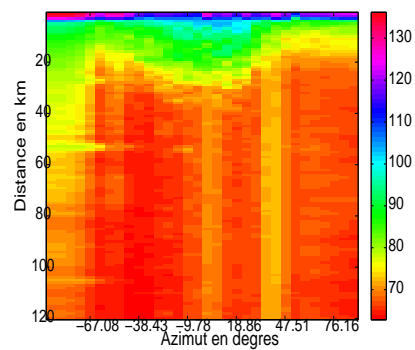
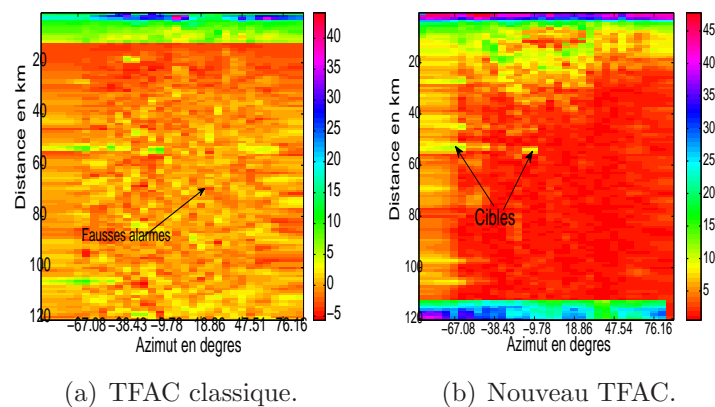
FIG. 6.4 – Représentation des  $120 \times 32$  coefficients de réflexion dans le plan complexe.FIG. 6.5 – Intensité rétrodiffusée en chaque case distance-azimut, l'intensité est mesurée en décibels, soit  $10 \log(\gamma(0))$  (labels des couleurs à droite des graphiques).

FIG. 6.6 – Cartes distance-azimut représentant les distances après normalisation obtenues par le détecteur TFAC classique et celles obtenues par le nouveau détecteur TFAC.

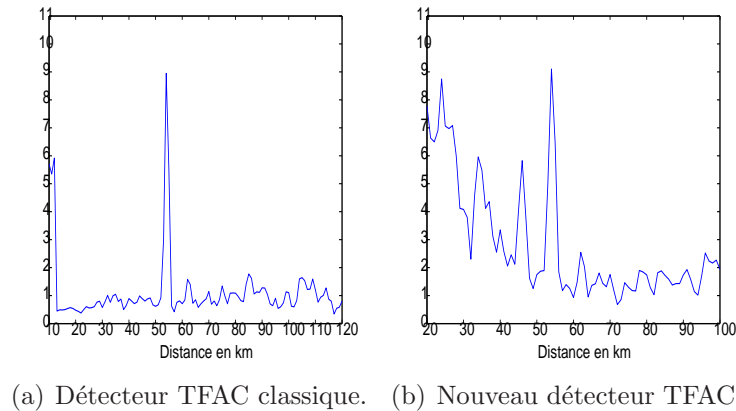


FIG. 6.7 – Comparaison de la détection pour l’azimut de  $-9.78^\circ$ . A gauche, on représente les distances obtenues par le TFAC classique fonction de la case distance et à droite les distances de Siegel. Les résultats sont présentés après normalisation.

Dans la figure 6.7, on voit que pour les deux types de détecteurs, les résultats sont similaires. La cible est bien détectée dans les deux cas, car le rapport signal sur bruit est élevé. Cependant, ce qui différencie un bon détecteur d’un mauvais détecteur est qu’il parvient à détecter pour des faibles rapports signal sur bruit.

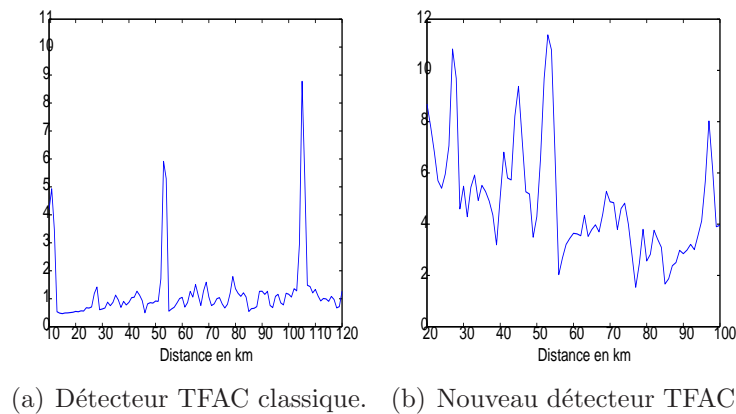


FIG. 6.8 – Comparaison de la détection pour l’azimut de  $-67.08^\circ$ .

Les résultats présentés dans la figure 6.8 sont ceux relatifs à la deuxième cible qui est plongée dans un fort bruit ambiant. On peut constater une nette amélioration en utilisant le nouveau détecteur. La distance obtenue passe de la valeur de 5.92 avec l’ancien détecteur à 11.38 avec le nouveau détecteur. Ainsi, si on avait utilisé dans la détection un seuil de 10, le nouveau détecteur aurait détecté alors que l’ancien n’aurait pas détecté la cible.

### 6.4.2 Détection utilisant une segmentation bayésienne

Nous proposons d'utiliser une segmentation pour affiner la détection. Plus exactement, lorsque l'on calculera la moyenne des matrices, nous n'utiliserons uniquement celles appartenant à la même classe que la case sous test. Nous comparerons la détection en utilisant notre nouveau détecteur sans segmentation et avec segmentation. Nous utiliserons le modèle de chaîne de Markov cachée à bruit indépendant décrit dans la sous-section 6.3.2. La figure 6.9 présente la segmentation de la carte distance-azimut en 4 classes. La figure 6.10 présente l'appartenance des différents coefficients de réflexion à une classe. On peut alors remarquer que les coefficients de réflexion de la classe "jaune" tendent rapidement vers 0 lorsque l'on se rapproche de l'ordre  $p = 10$ . Ainsi, la classe "jaune" correspond à du bruit blanc. Pour finir, nous avons représenté dans la figure 6.11 les densités de lois gamma correspondant aux densités estimées de l'inverse de la puissance pour chacune des classes. Nous y avons également superposé les histogrammes de l'inverse de la puissance en ne considérant que les cases d'appartenance à une même classe. Ceci permet de mettre en évidence le bon comportement de l'algorithme ICE utilisé pour l'estimation.

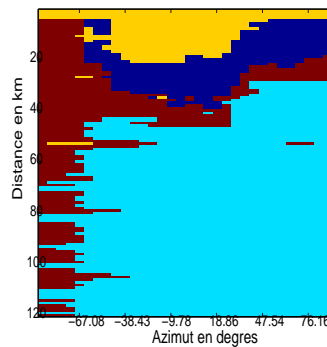


FIG. 6.9 – Segmentation de la carte distance-azimut.

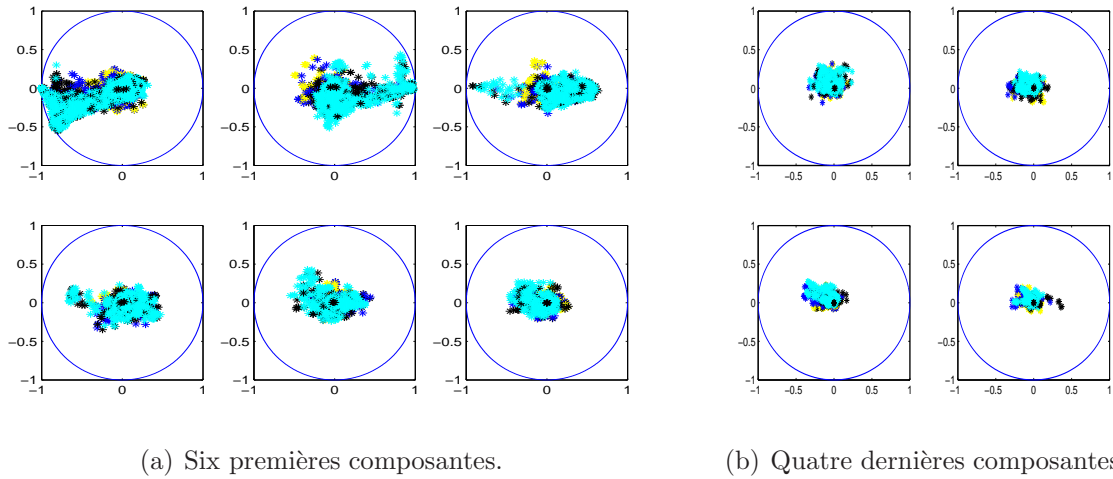


FIG. 6.10 – Classification des différents coefficients de réflexion, la couleur “noire” correspond à la classe brune sur la segmentation.

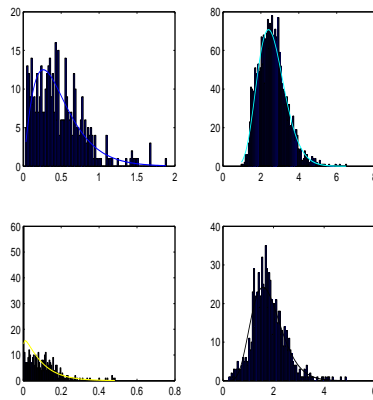


FIG. 6.11 – Comparaison des histogrammes des inverses des puissances pour chaque classe avec les densités estimées (loi gamma).

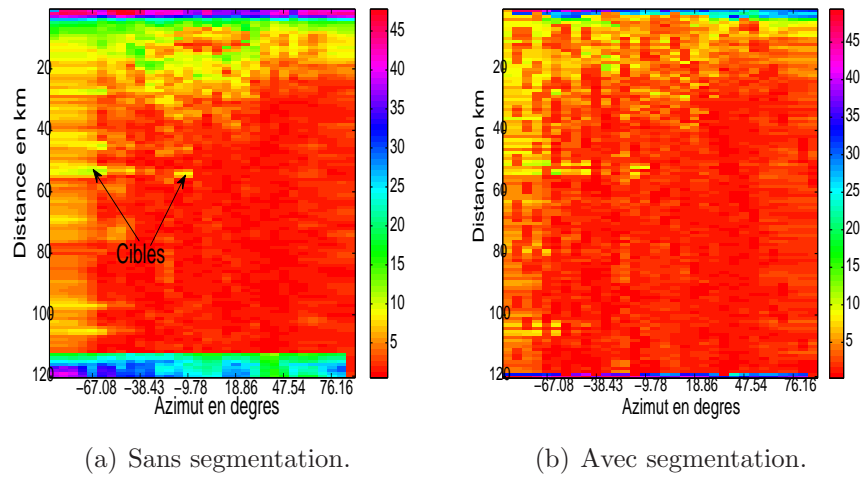


FIG. 6.12 – Distances de Siegel sans et avec segmentation. Avec segmentation, la moyenne est calculée pour les cases dans la même classe que la case sous test.

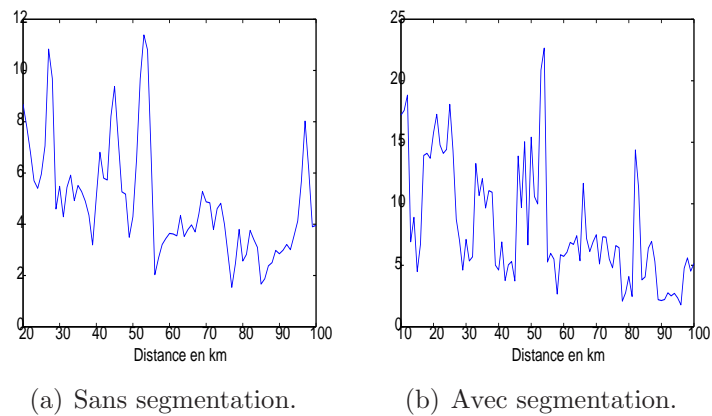


FIG. 6.13 – Distance de Siegel sans et avec segmentation pour l'azimut  $-67.08^\circ$ .

De la figure 6.12 on constate une amélioration du contraste autour de la cible située à 53 kilomètres et  $-67.08^\circ$ . La figure 6.13 présente la distance de Siegel pour l'azimut de  $-67.08^\circ$  sans segmentation et avec segmentation. De cette dernière figure, on constate que l'apport de la segmentation est important. D'une part, la distance de Siegel passe de 11.38 à 22.65 mais d'autre part certains maxima certainement responsables de fausses alarmes sont atténués. La segmentation permet ainsi de rendre plus robuste notre détecteur. Sans la segmentation, le signal d'une des cases environnantes pourrait avoir un comportement statistique très différent des signaux des autres cases environnantes. La moyenne des matrices de covariance serait alors non représentative de l'environnement de la case sous test et la détection en serait affectée. La segmentation permet ainsi d'éliminer du calcul de la moyenne les cases pour lesquelles la statistique du signal serait trop différente de celle de l'ambiance. Ainsi, lors de la détection, on évite de comparer la statistique du signal sous test avec celle du signal provenant d'une cible.

## Conclusion

Dans ce chapitre, nous avons vu comment la segmentation bayésienne pouvait être utilisée comme un pré-traitement des données par d'autres applications. L'application concernée dans ce chapitre est celle de la détection de cible à partir d'un signal radar. Le détecteur que nous avons implémenté utilise les propriétés statistiques du signal reçu. Pour cela, nous avons défini une distance entre lois de probabilité. Cette distance a été ensuite calculée dans le cas des lois normales complexes centrées et circulaires. Parmi les perspectives à envisager, on peut étendre la distance entre matrices de covariance à celle entre lois normales complexes circulaires non centrées. On peut également envisagé le cas non gaussien en utilisant notamment les travaux de M. Calvo et J. M. Oller [26] dans lesquels il généralise au cas de lois elliptiques. Le cas des lois  $\Gamma$  a également été traité dans l'article [107] de F. Reverter et J. M. Oller. Parmi les autres perspectives de ces travaux, on peut également prévoir l'extension du détecteur TFAC médian classique, qui au lieu de calculer les moyennes des cases environnantes, calcule les médianes des cases environnantes. Ce détecteur est alors plus robuste aux valeurs atypiques. En effet, une valeur très grande dans un échantillon ne modifie pas la médiane. Ainsi l'idée serait d'étendre la notion de médiane pour des matrices hermitiennes définies positives en utilisant une propriété de la médiane. Concernant le modèle utilisé pour la segmentation, on peut envisager l'utilisation de lois  $K$  [17] pour l'intensité et l'utilisation de lois sur le polydisque unité  $\{\mu \in \mathbb{C}^p : |\mu_1|^2 + \dots + |\mu_p|^2 = 1\}$  au lieu d'un produit de lois de Von Mises-Fisher. L'introduction de la dépendance longue peut être envisagée également lorsque les données sont par exemple issues de radar de haute résolution, auquel cas cette dépendance longue due à la structure du fouilli de mer (pseudo-périodicité due aux vagues) ne peut être négligée.



# Conclusion et perspectives

Le contexte général de notre étude était l'estimation automatique, à partir des variables observées  $Y = (Y_s)_{s \in \mathcal{S}}$ , des réalisations des variables inobservées  $X = (X_s)_{s \in \mathcal{S}}$ . Le modèle classique par chaînes de Markov cachées (CMC) est parmi les modèles les plus utilisés et les plus efficaces pour accomplir une telle estimation lorsque le nombre de variables est trop grand pour que l'on puisse utiliser la loi  $p(x, y)$  du couple  $(X, Y)$  dans toute sa généralité. Cependant, ce modèle a ses limites à cause de la simplicité de la loi  $p(y|x)$ , qui modélise le "bruit". Ce qui le rend difficilement justifiable dans un certain nombre de situations réelles. Les différentes contributions originales de notre mémoire concernent différentes extensions du modèle CMC. Ces différentes extensions s'appuient sur les modèles génériques que sont les "chaînes de Markov triplets" (CMT, [103]) et les "chaînes triplets partiellement de Markov" (CTPM, [97]). Concernant les CMT, leur pouvoir modélisant a commencé à être exploité récemment dans la thèse de P. Lanchantin [65]. En particulier, les CMT permettent de modéliser les CMC non stationnaires. Nous avons proposé un certain nombre de modélisations particulières se rapportant à cette problématique. Nous avons montré que les modèles classiques par chaînes semi-markoviennes cachées (CSMC) sont des CMT particulières. Dans ce contexte, nous avons proposé une CSMC originale, autorisant des temps de calcul bien moindres que ceux nécessaires à l'utilisation des CSMC classiques. Ensuite, les CSMC ont été étendues au cas où la chaîne cachée n'est pas stationnaire. Enfin, en développant les idées proposées dans la thèse de N. Brunel [21, 23], nous avons introduits dans les nouveaux modèles les copules, ce qui autorise une grande richesse de la modélisation de la loi  $p(y|x)$ . Pour chacun des nouveaux modèles nous avons proposé une méthode adéquate d'estimation des paramètres, de type ICE, et l'intérêt des segmentations bayésiennes non supervisées associées a été validé par des expérimentations informatiques.

Concernant les CTPM, leur potentiel modélisant a été décrit dans [97, 98] ; cependant, contrairement aux CMT, poser le modèle général ne conduit pas immédiatement aux méthodes exploitables. Afin de pouvoir traiter des bruits  $p(y|x)$  à mémoire longue, nous avons proposé, en collaboration avec P. Lanchantin, un modèle particulier dans lequel la chaîne cachée est markovienne, et le bruit est gaussien. De telles "chaînes de Markov cachées par du bruit à mémoire longue" (CMC-ML) sont alors exploitables et permettent d'estimer la chaîne cachée par les méthodes bayésiennes classiques [66]. Nous avons ensuite étendu ces CMC-ML aux modèles dans lesquels seule la copule reste gaussienne, les lois marginales du bruit pouvant être quelconques. Enfin, une extension non triviale du principe général de la méthode d'estimation des paramètres ICE nous a permis de proposer des méthodes de segmentation non supervisées, dont le bon comportement a été constaté par des expérimentations. Notons qu'il y a plusieurs types de bruit à mémoire longue comme les processus FARIMA ou les bruits gaussiens fractionnaires, et les modèles proposés sont utilisables dès que l'on connaît la forme de la fonction de covariance. Nos modèles s'appliquent ainsi, de manière automatique, dans

différents problèmes de détection de changements de régime dans des phénomènes aléatoires à mémoire longue.

En guise de perspective, on peut envisager de généraliser les modèles considérés dans cette thèse à des modèles de dépendance plus complexes tels que les arbres ou autres modèles graphiques. De telles généralisations peuvent se faire facilement dans les arbres de Markov, qui peuvent être vus comme des chaînes “hiérarchisées”. Notons également certains résultats déjà existant, concernant la non stationnarité ou l’utilisation des copules, dans le cadre des champs de Markov [11, 12, 13].

# Annexe A

## Fonctions eulériennes et fonctions de Bessel

### A.1 Fonctions eulériennes Gamma et Beta

**Définition A.1.1** (Fonction Gamma et Beta). *La fonction  $\Gamma$  est définie sur  $\mathbb{C}^+ = \{z \in \mathbb{C} : \operatorname{Re}(z) > 0\}$  par :*

$$\Gamma(z) = \int_0^{+\infty} t^{z-1} \exp(-t) dt. \quad (\text{A.1})$$

*La fonction  $\beta$  est définie sur  $\mathbb{C}^{+2}$  par :*

$$\beta(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)} \quad (\text{A.2})$$

**Propriétés 1.** *On a les propriétés suivantes :*

1. *Pour tout entier  $n \geq 1$ ,  $\Gamma(n) = (n-1)!$  et  $\Gamma(z) = (z-1)\Gamma(z-1)$ .*
2. *On a la formule de Stirling pour tout  $z \in \mathbb{C}^+$  :*

$$\Gamma(z) \sim^{+\infty} \sqrt{2\pi} (z-1)^{z-\frac{1}{2}} e^{-(z-1)}.$$

3. *Pour tous réels  $a > 0$  et  $b > 0$ , on a :*

$$\beta(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt.$$

4. *On a propriété asymptotique :*

$$\Gamma(z) = \lim_{n \rightarrow +\infty} \frac{n! n^z}{z(z+1)\dots(z+n)}.$$

5.  *$\Gamma$  se prolonge à une fonction méromorphe sur  $\mathbb{C}$  dont les pôles sont les éléments de  $\mathbb{Z}^-$ .*

6. Le produit infini de cette fonction méromorphe est donné par :

$$\frac{1}{\Gamma(z)} = z \exp(\gamma z) \prod_{k=1}^{+\infty} \left(1 + \frac{z}{k}\right) \exp\left(-\frac{z}{k}\right),$$

où  $\gamma = \lim_{n \rightarrow +\infty} \sum_{k=1}^n \frac{1}{k} - \ln(n)$  est la constante d'Euler.

7. On a  $\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)}$ , pour tout  $z \in \mathbb{C}^+$ .

8. Pour tout  $z \in \mathbb{C}^+$ , la formule des compléments est donnée par :

$$\Gamma(z)\Gamma\left(z + \frac{1}{n}\right) \dots \Gamma\left(z + \frac{n-1}{n}\right) = (2\pi)^{\frac{n-1}{2}} n^{\frac{1}{2}-nz} \Gamma(nz).$$

**Corollaire A.1.1.** La fonction “digamma” est définie sur  $\mathbb{C} - \mathbb{Z}^-$  par  $\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$ , celle-ci est développable en série de Laurent sur  $\mathbb{C} - \mathbb{Z}^-$  et :

$$\psi(z) = -\frac{1}{z} - \gamma + \sum_{k=1}^{+\infty} \frac{z}{k(z+k)}. \quad (\text{A.3})$$

## A.2 Fonctions de Bessel modifiées

Les deux fonctions de Bessel modifiées de première espèce et de seconde espèce constituent une base de l'espace vectoriel des solutions d'une équation différentielle linéaire du second ordre appelée “équation de Bessel modifiée”.

**Définition A.2.1.** Soit  $\nu \in \mathbb{R}$ , on appelle équation de Bessel modifiée l'équation différentielle :

$$x^2 y'' + xy' - (x^2 + \nu^2)y = 0, \quad (\text{A.4})$$

où  $y$  est une fonction de  $\mathbb{R}$  dans  $\mathbb{R}$ .

Une base de l'ensemble des solutions est donnée par les deux fonctions de Bessel modifiées :

**Définition A.2.2.** On appelle fonction de Bessel modifiée de première espèce notée  $I_\nu$  l'unique solution de (A.4) telle que  $x^{-\nu} I_\nu$  soit développable en série entière et

$$\lim_{x \rightarrow 0} x^{-\nu} I_\nu(x) = \frac{1}{2^\nu \Gamma(\nu + 1)}.$$

Celle-ci est donnée par le développement de Laurent :

$$I_\nu(x) = \left(\frac{x}{2}\right)^\nu \sum_{p=0}^{+\infty} \frac{\left(\frac{x}{2}\right)^{2p}}{p! \Gamma(\nu + p + 1)}.$$

On appelle fonction de Bessel modifiée de seconde espèce (notée  $K_\nu$ ) la fonction définie par :

$$K_\nu(x) = \frac{\pi}{2} \frac{I_{-\nu}(x) - I_\nu(x)}{\sin(\pi\nu)}.$$

De plus c'est l'unique solution de l'équation de Bessel modifiée telle que :

$$\begin{aligned}
& - \left[ \frac{2 \sin(\pi\nu)}{\pi} K_\nu + I_\nu \right] x^\nu \text{ soit développable en série entière.} \\
& - \lim_{x \rightarrow 0} \left[ \frac{2 \sin(\pi\nu)}{\pi} K_\nu(x) + I_\nu(x) \right] x^\nu = \frac{2^\nu}{\Gamma(1-\nu)}.
\end{aligned}$$

Les deux propositions suivantes nous permettent d'écrire les fonctions de Bessel sous forme intégrale :

**Proposition A.2.1.** *Les deux fonctions de Bessel  $I_\nu$  et  $K_\nu$  s'écrivent sous les formes intégrales suivantes :*

$$I_\nu(x) = \frac{x^\nu}{2^\nu \Gamma(\nu + 1) \int_0^\pi \sin^{2\nu}(\theta) d\theta} \times \int_0^\pi \sin^{2\nu}(\theta) \exp(x \cos(\theta)) d\theta, \quad (\text{A.5})$$

$$K_\nu(x) = \frac{1}{2} \int_0^{+\infty} u^{\nu-1} \exp\left(-\frac{x}{2}\left(u + \frac{1}{u}\right)\right) du = \frac{1}{2} \int_{\mathbb{R}} \cosh(\nu t) \exp(-x \cosh t) dt. \quad (\text{A.6})$$

*Preuve.* Utiliser les caractérisations des fonctions de Bessel. □