Détection et classification de cibles multispectrales dans l'infrarouge

Florian Maire

ONERA / Telecom SudParis

Soutenue le 14 Février 2014 devant le jury composé de :

M. Jean-Michel MARIN	Université Montpellier 2	Rapporteur
Mme. Stéphanie ALLASSONNIÈRE	Ecole Polytechnique	Examinateur
M. Jacques BLANC-TALON	DGA - MRIS	Examinateur
M. Christian ROBERT	Université Paris-Dauphine	Examinateur
M. Olivier WINTENBERGER	Université Paris 6	Examinateur
M. Randal DOUC	Telecom SudParis	Directeur de thèse
M. Eric MOULINES	Télécom ParisTech	Directeur de thèse
Mme. Sidonie LEFEBVRE	ONERA	Encadrante

Plan de la présentation

Plan de la présentation

Plan de la présentation

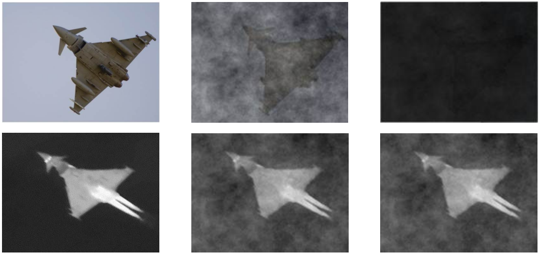
Contexte

- Surveillance de sites sensibles contre des menaces aériennes

- Utilisation d'images infrarouge

Visible

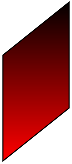
Infrarouge



jour / dégagé   jour / nuageux   nuit / nuageux

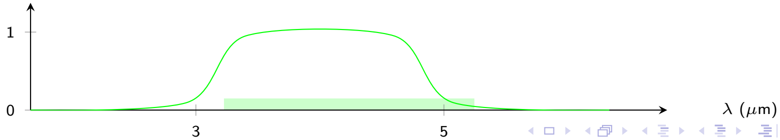
Capteurs mono/multispectral

détecteur infrarouge **monospectral**

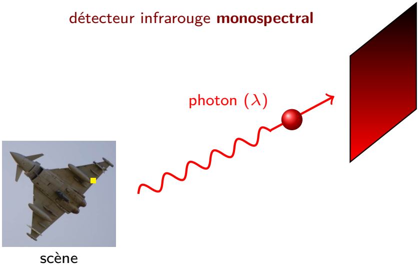


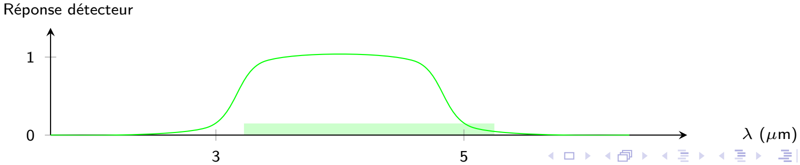
scène

Réponse détecteur

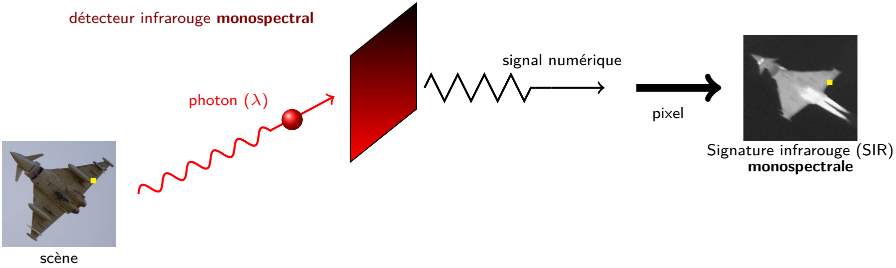


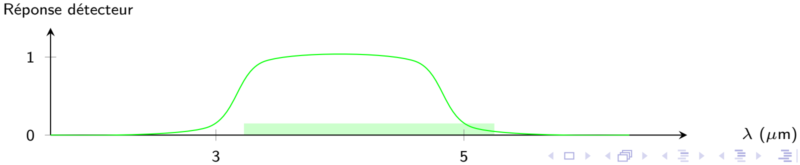
Capteurs mono/multispectral



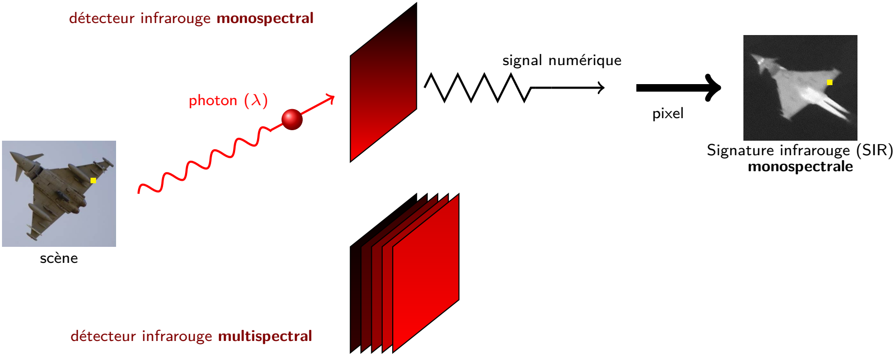


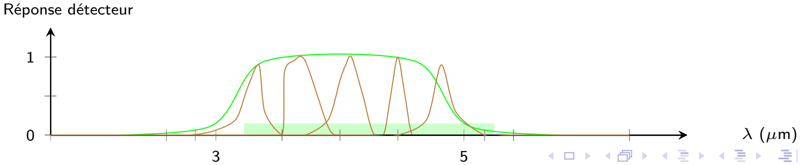
Capteurs mono/multispectral



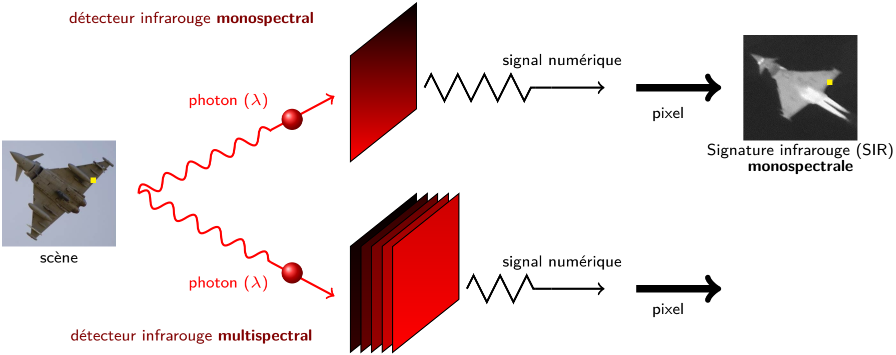


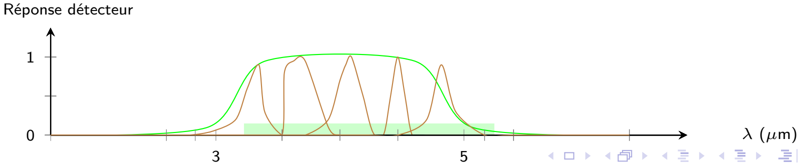
Capteurs mono/multispectral



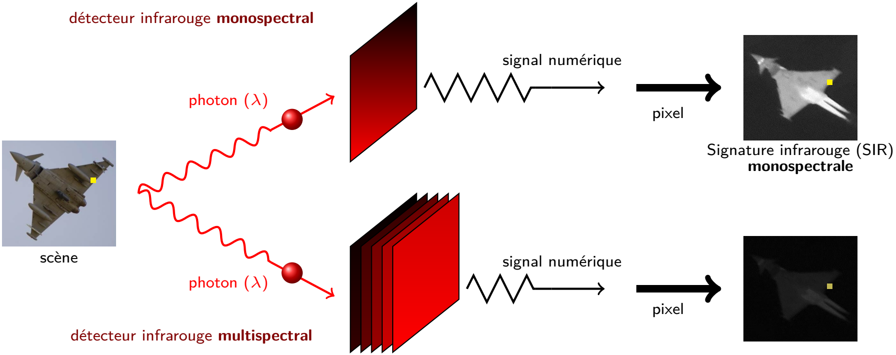


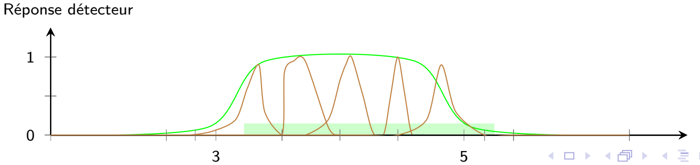
Capteurs mono/multispectral



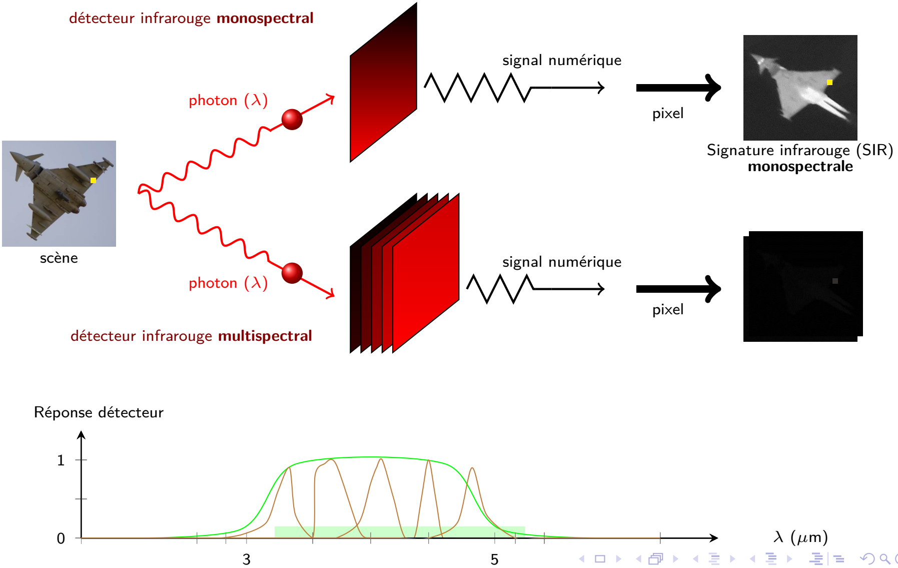


Capteurs mono/multispectral

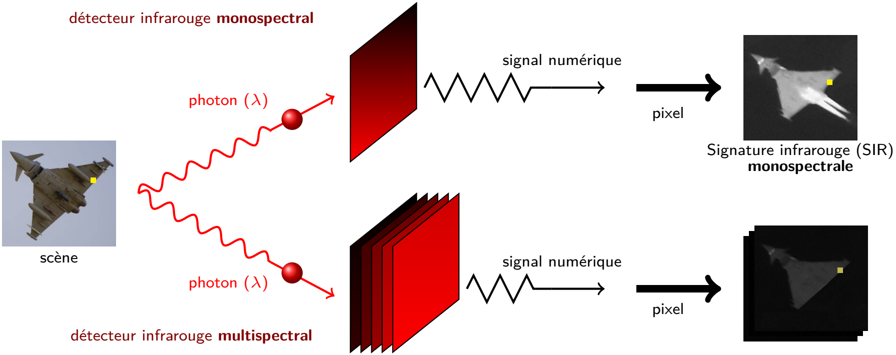


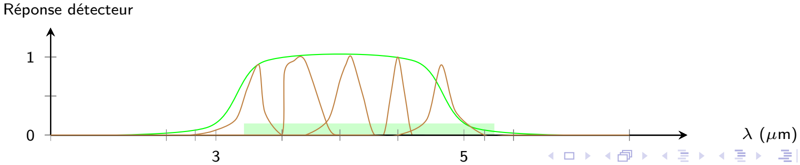


Capteurs mono/multispectral

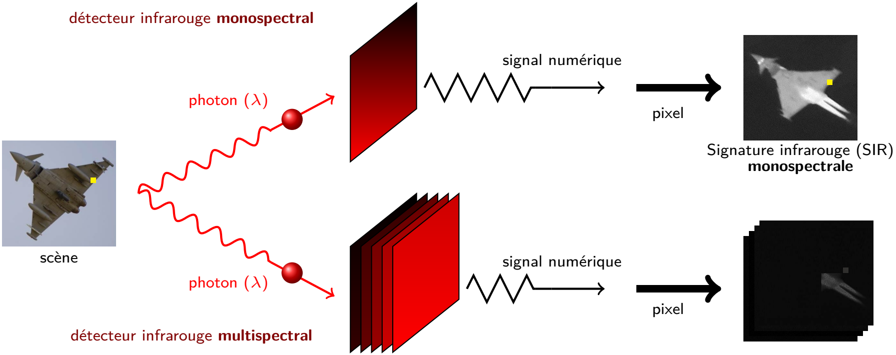


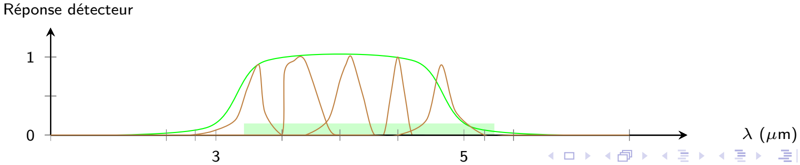
Capteurs mono/multispectral



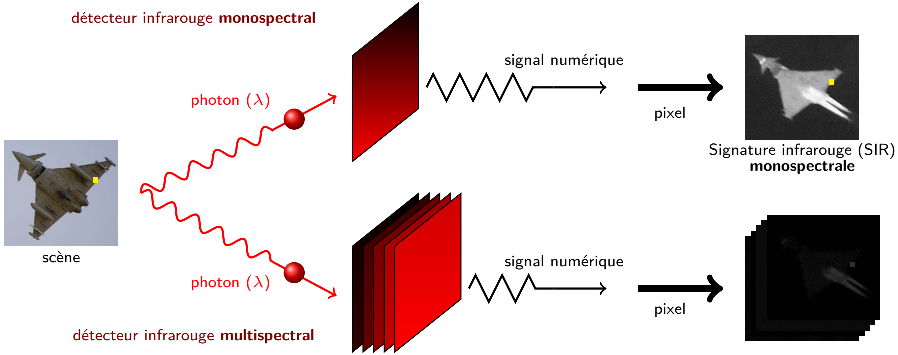


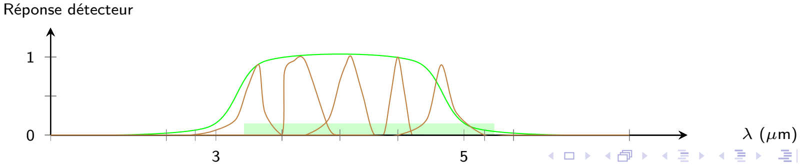
Capteurs mono/multispectral



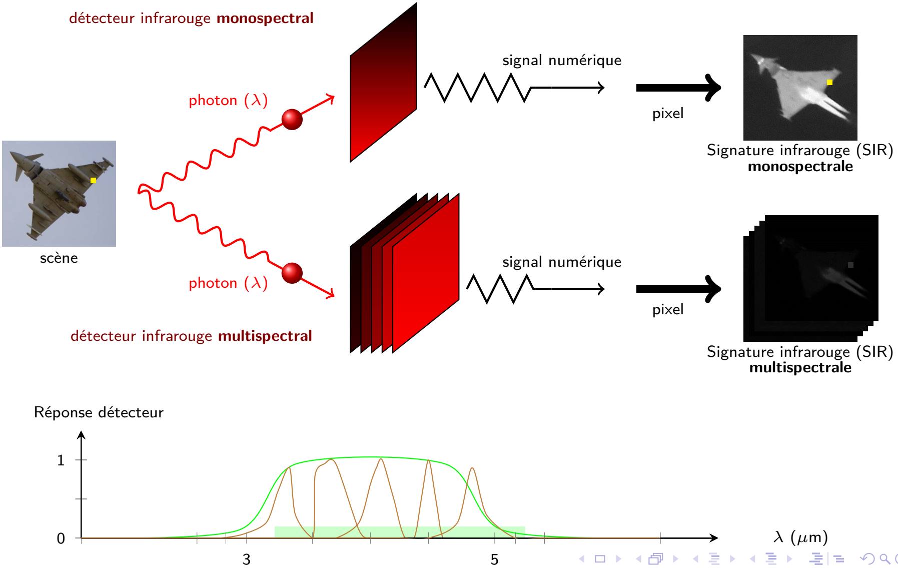


Capteurs mono/multispectral



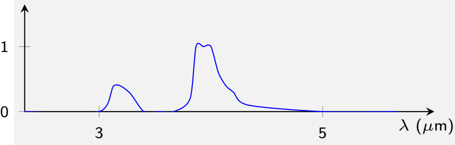


Capteurs mono/multispectral

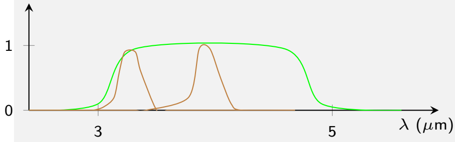


Intérêt du multispectral?

Rayonnement infrarouge hypothétique d'un avion



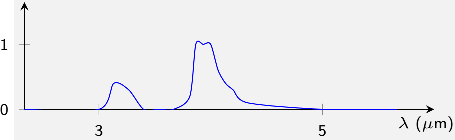
Réponse détecteur monospectral et multispectral



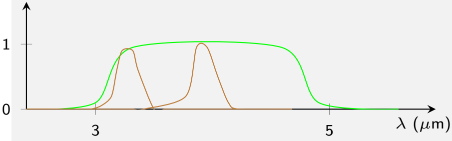
⇒ (i) **Augmenter le rapport signal à bruit**

Intérêt du multispectral?

Rayonnement infrarouge hypothétique d'un avion

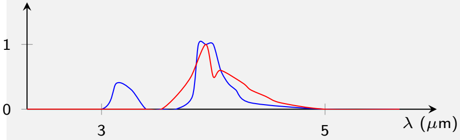


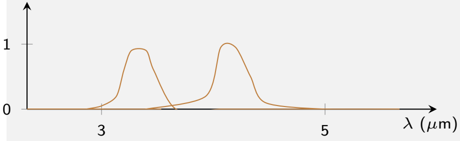
Réponse détecteur monospectral et multispectral



⇒ (i) **Augmenter le rapport signal à bruit**

Rayonnement infrarouge hypothétique de 2 avions **différents**





⇒ (ii) **Discriminer différentes cibles**

Objectifs de la thèse

(i) **détecter** des aéronefs à partir de signature infrarouge,

Objectifs de la thèse

(i) **détecter** des aéronefs à partir de signature infrarouge,

(ii) **classifier** des aéronefs à partir de leur signature infrarouge,

Objectifs de la thèse

(i) **détecter** des aéronefs à partir de signature infrarouge,

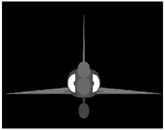
(ii) **classifier** des aéronefs à partir de leur signature infrarouge,

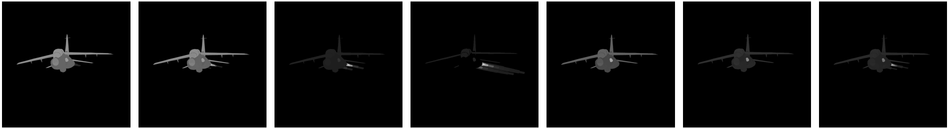
(iii) **spécifier les bandes infrarouge** préférables pour ces applications.

Besoin de données...

- Signatures d'aéronefs *expérimentales* impossibles à obtenir
- Simulateur de signatures mono/multispectral développé par l'ONERA

Simulation d'une SIR monospectrale $\Rightarrow$



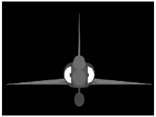


Simulation d'une SIR multispectrale (7 bandes)

Données correspondant au scénario...

- Distance cible/capteur $\Rightarrow$ **faible résolution** :

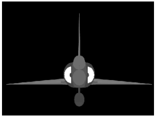
SIR haute résolution
1024×1024



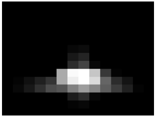
Données correspondant au scénario...

- Distance cible/capteur $\Rightarrow$ **faible résolution** :

SIR haute résolution
1024×1024



$\longrightarrow$

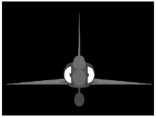


SIR faible résolution
16×16

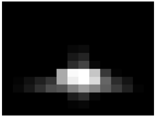
Données correspondant au scénario...

- Distance cible/capteur $\Rightarrow$ **faible résolution** :

SIR haute résolution
1024×1024

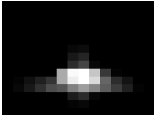


$\longrightarrow$



SIR faible résolution
16×16

- Fond de ciel $\Rightarrow$ **présence de bruit** :

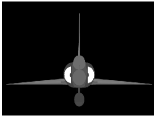


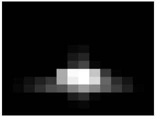
$\longrightarrow$

Données correspondant au scénario...

- Distance cible/capteur $\Rightarrow$ **faible résolution** :

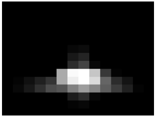
SIR haute résolution
1024×1024

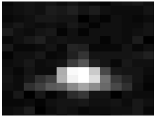




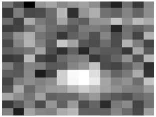
SIR faible résolution
16×16

- Fond de ciel $\Rightarrow$ **présence de bruit** :





σ_1



σ_2

Variabilité des signatures

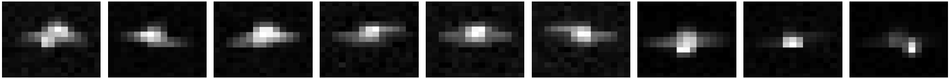
- Incertitude paramètres scénario $\Rightarrow$ variabilité des signatures

Variabilité des signatures

- Incertitude paramètres scénario $\Rightarrow$ variabilité des signatures
- Deux origines de la dispersion :

Variabilité des signatures

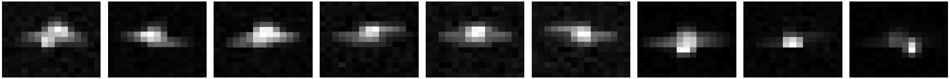
- Incertitude paramètres scénario $\Rightarrow$ variabilité des signatures
- Deux origines de la dispersion :
  - (i) **Dispersion géométrique (spatiale)**



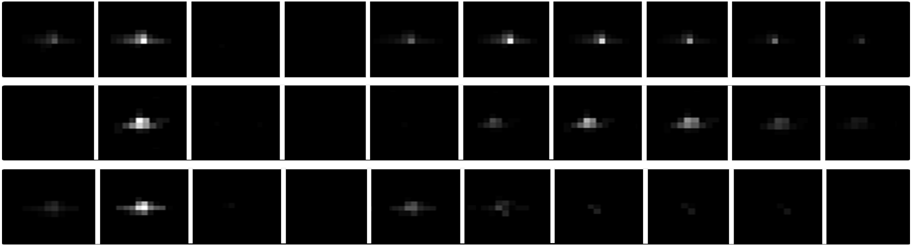
Variabilité des signatures

- Incertitude paramètres scénario $\Rightarrow$ variabilité des signatures
- Deux origines de la dispersion :

(i) **Dispersion géométrique (spatiale)**



(ii) **Dispersion photométrique (spectrale)**



Challenges

Les méthodes statistiques doivent :

Challenges

Les méthodes statistiques doivent :

(i) intégrer simultanément les **dispersions spectrales et spatiales**

Challenges

Les méthodes statistiques doivent :

(i) intégrer simultanément les **dispersions spectrales et spatiales**

(ii) être adaptées aux données de (très) **faible résolution** et **bruitées**

Challenges

Les méthodes statistiques doivent :

(i) intégrer simultanément les **dispersions spectrales et spatiales**

(ii) être adaptées aux données de (très) **faible résolution** et **bruitées**

(iii) permettre un traitement des données de façon **séquentielle** et en ***quasi*-temps réel**

Plan de la présentation

Position du problème

- Observations $\{Y_n,\ n \in \mathbb{N}\}$ v.a. sur $(\mathsf{Y}, \mathcal{Y}, \mathbb{P})$ où

$$\begin{cases} Y_n \mid Y_n \in \mathcal{H}_0 \sim \mathbb{P}_0 \\ Y_n \mid Y_n \in \mathcal{H}_1 \sim \mathbb{P}_1 \end{cases}$$

- Test à deux hypothèses $\Phi : \mathsf{Y} \to \{\mathcal{H}_0, \mathcal{H}_1\}$ tel que :
  - (i) $P_{\mathrm{FA}}(\Phi) = \mathbb{P}[\Phi(Y) \in \mathcal{H}_1 \mid Y \in \mathcal{H}_0]$ soit minimale
  - (ii) $P_{\mathrm{D}}(\Phi) = \mathbb{P}[\Phi(Y) \in \mathcal{H}_1 \mid Y \in \mathcal{H}_1]$ soit maximale

- Hypothèses : $\mathbb{P}_0$ connue, $\mathbb{P}_1$ inconnue

$\Rightarrow$ Approche détection d'anomalie

Détecteur RX (Reed & Xiaoli, 1990)

- Images multispectrales à K bandes de dimension P
- $Y \in \mathrm{Y},\ Y = \{y_j \in \mathbb{R}^K,\ 1 \leq j \leq P\}$
- Modèle de fond
  (i) indépendant spatialement : $\mathbb{P}_0 = \bigotimes_{j=1}^{P} \mu_0$
  (ii) corrélé spectralement : $\mu_0 = \mathcal{N}(0_K, \Sigma)$
- Distance de Mahalanobis :

$$\forall\, y \in \mathbb{R}^K, \quad D_{\mathrm{M}}(y, \mu_0) = (y - \mu_0)^{\mathsf{T}} \Sigma^{-1} (y - \mu_0)$$

- Détecteur RX :

$$\phi_{\mathrm{RX}}(Y) \to \begin{cases} \mathcal{H}_0 & \text{si} \quad \max_j D_{\mathrm{M}}(y_j, \mu_0) \leq \eta \\ \mathcal{H}_1 & \text{si} \quad \max_j D_{\mathrm{M}}(y_j, \mu_0) > \eta \end{cases}$$

Un nouveau détecteur d'anomalies spectrales / spatiales

limite de ϕ_{RX} : pixels étudiés de façon indépendante
$\Rightarrow$ néglige la dispersion spatiale des cibles

(i) **Interpolation** de la fonction

$$\forall\, x \in \mathbb{R}^2, \quad I_Y(x) = \sum_{j=1}^{P} D_{\mathrm{M}}(y_j, \mu_0)\psi_j(x)\ ,$$

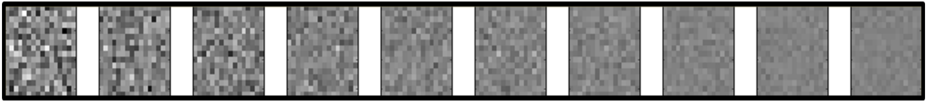
(ii) Etude des **ensembles de niveau** $\eta > 0$ de la fonction I_Y

$$\forall\, Y \in \mathsf{Y}, \quad A_\eta(Y) = \{x \in \mathbb{R}^2,\ I_Y(x) > \eta\}\ ,$$

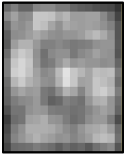
Test de détection $\phi(Y) \to \begin{cases} \mathcal{H}_0 & \text{si} \quad \max\left(\mathrm{Per}(A_\eta(Y))\right) \leq \nu \\ \mathcal{H}_1 & \text{si} \quad \max\left(\mathrm{Per}(A_\eta(Y))\right) > \nu \end{cases}$

Illustration sur une image "*facile*"

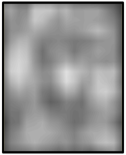
une image multispectrale $Y \in \mathrm{Y}$



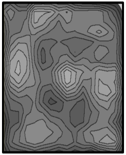
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$



(i) $x \to I_Y(x)$



(ii) $\{A_\eta,\ \eta > 0\}$



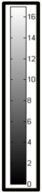
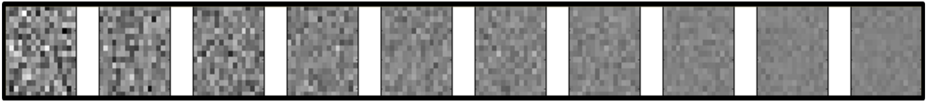
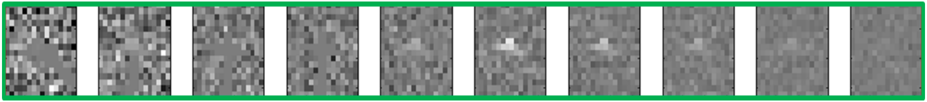


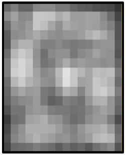
Illustration sur une image "*facile*"

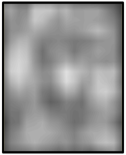
une image multispectrale $Y \in \mathsf{Y}$

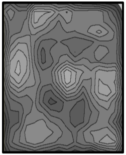




$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$ (i) $x \to I_Y(x)$ (ii) $\{A_\eta,\ \eta > 0\}$







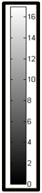
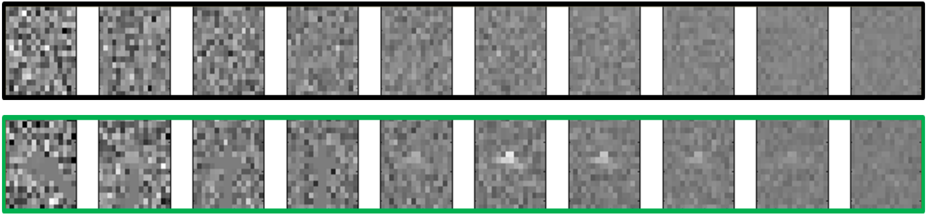
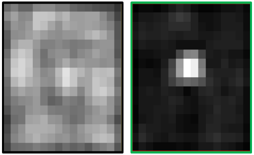


Illustration sur une image "*facile*"

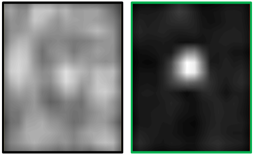
une image multispectrale $Y \in \mathsf{Y}$



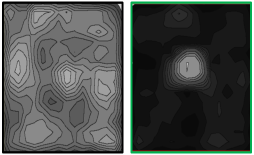
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$



(i) $x \to I_Y(x)$



(ii) $\{A_\eta,\ \eta > 0\}$



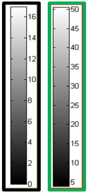
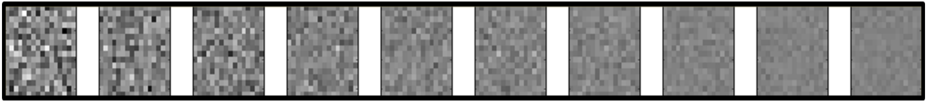
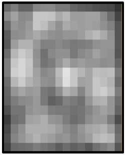


Illustration sur une image "*difficile*"

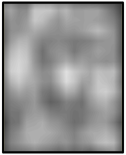
une image multispectrale $Y \in \mathsf{Y}$



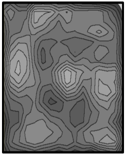
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$



(i) $x \to I_Y(x)$



(ii) $\{A_\eta,\ \eta > 0\}$



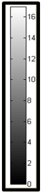
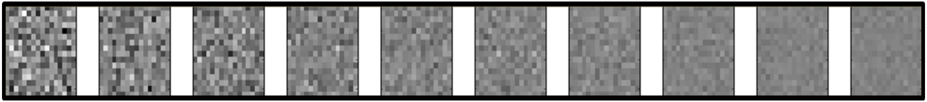
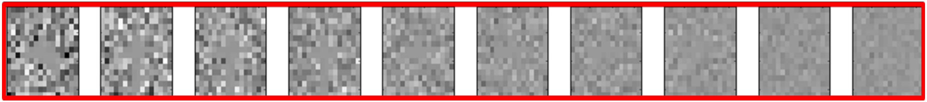


Illustration sur une image "*difficile*"

une image multispectrale $Y \in \mathsf{Y}$

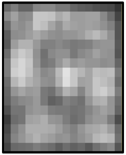


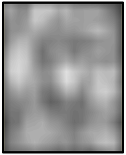


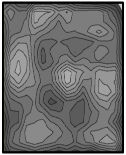
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$

(i) $x \to I_Y(x)$

(ii) $\{A_\eta,\ \eta > 0\}$







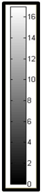
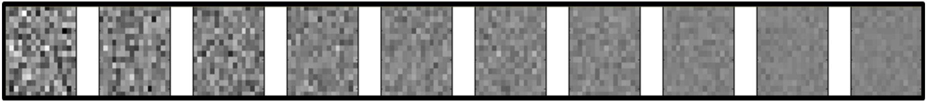
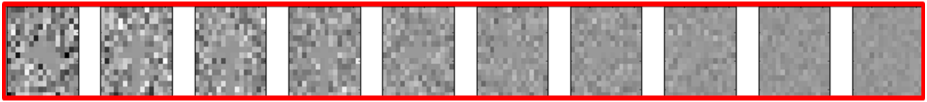


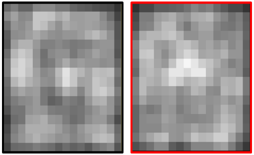
Illustration sur une image "*difficile*"

une image multispectrale $Y \in \mathrm{Y}$

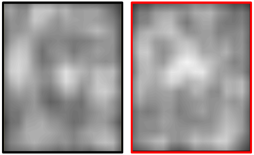




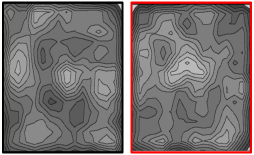
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$

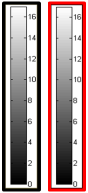


(i) $x \to I_Y(x)$



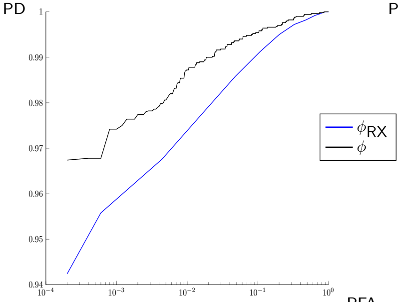
(ii) $\{A_\eta,\ \eta > 0\}$



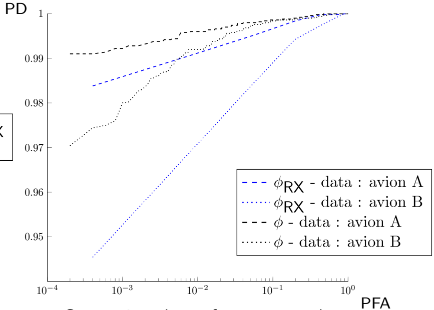


Résultats

Comparaison de la courbe PD/PFA des détecteurs ϕ_{RX} et ϕ



Tous types d'avions mélangés



Comparaison des performances sur deux types d'avions différents

$\Rightarrow$ L'intérêt de ϕ est plus évident pour les avions ayant un *faible* rayonnement IR (donc plus durs à détecter!)

Un détecteur robuste

Le détecteur ϕ permet de :

(i) détecter les SIR avec un faible rayonnement IR...

(ii) ... tout en conservant un taux de fausse alarme acceptable

(iii) donner une information contextuelle
$\Rightarrow$ sur la position (spatiale) de la cible dans l'image
$\Rightarrow$ donner une information sur la géométrie de la cible

mais est limité par :

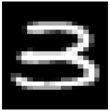
(i) le manque d'expression exacte des seuils correspondant à une P_{FA} donnée

(ii) les performances d'un point de vue calculatoire

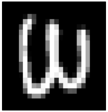
Plan de la présentation

Une première remarque...

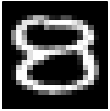
- Analogie avec un problème de reconnaissance de caractère...



Y_1



Y_2

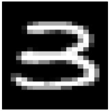


Y_3

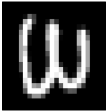
$$d(Y_1, Y_2) \gg d(Y_1, Y_3)$$
d distance sur $\mathbb{R}^P$

Une première remarque...

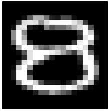
- Analogie avec un problème de reconnaissance de caractère...



Y_1



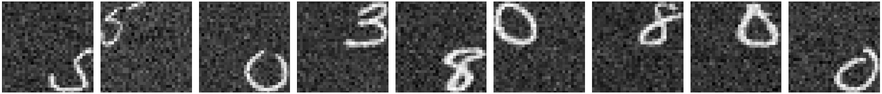
Y_2



Y_3

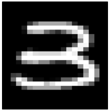
$$d(Y_1, Y_2) \gg d(Y_1, Y_3)$$
d distance sur $\mathbb{R}^P$

- Données variant au niveau du contexte / de l'information

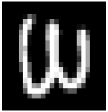


Une première remarque...

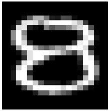
- Analogie avec un problème de reconnaissance de caractère...



Y_1



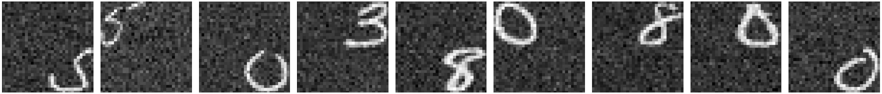
Y_2



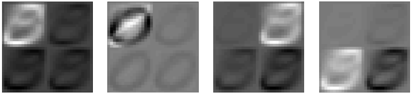
Y_3

$d(Y_1, Y_2) \gg d(Y_1, Y_3)$
d distance sur $\mathbb{R}^P$

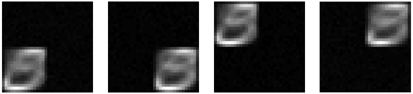
- Données variant au niveau du contexte / de l'information



- Clusters obtenus par deux méthodes classiques de classification :



4 modes principaux obtenus par ACP

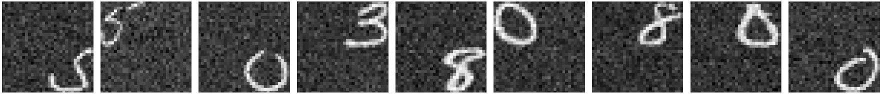


4 modes principaux obtenus par K-means

Un premier constat...

⇒ Nécessiter d'**estimer les déformations** afin de recaler les données entre elles.

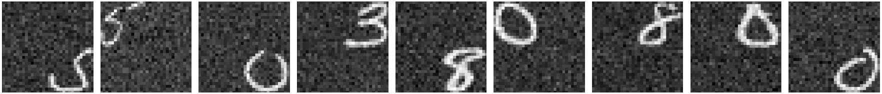
⇒ ***Décontextualiser*** l'information



Un premier constat...

⇒ Nécessiter d'**estimer les déformations** afin de recaler les données entre elles.

⇒ ***Décontextualiser*** l'information



Objectif:

Une approche Bayésienne de la classification

- Observations : v.a. $\{Y_n,\ n \in \mathbb{N}\}$ def. sur $(\mathsf{Y}, \mathcal{Y})$
- Modèle de mélange :

$$\forall A \in \mathcal{Y}, \quad \mathbb{P}\left[Y_n \in A\right] = \sum_{i \in \mathsf{I}} \mathbb{P}_i(A)\, \omega_i\ ,$$

$\mathbb{P}_i$: la distribution de Y_n cond. à $I_n = i$

ω_i : loi *apriori* de $I_n = i$.

- Classification d'une observation $Y_n \in \mathsf{Y}$ par la classe $\hat{Q}_n \in \mathsf{I}$:

$$\hat{Q}_n = \arg\max_{i \in \mathsf{I}} \mathbb{P}\left[I_n = i \mid Y_n\right] = \arg\max_{i \in \mathsf{I}} \mathbb{P}_i(Y_n)\, \omega_i\ ,$$

Problématique sous-jacente :

explorer les distributions conditionnelles $\{\mathbb{P}_i,\ i \in \mathsf{I}\}$

Modèle d'observation : **Modèle à prototype déformable**

Hypothèses :

- Existence de fonctions $\{\mathscr{Y}_n : \mathbb{R}^2 \to \mathbb{R},\ n \in \mathbb{N}\}$ t.q. :

$$\{\mathscr{Y}_n,\ n \in \mathbb{N}\} \xRightarrow[\text{sur grille de pixels}]{\text{discrétisation}} \{Y_n,\ n \in \mathbb{N}\}$$

- $\forall n \in \mathbb{N}$, $\mathscr{Y}_n$ dérive d'une même fonction **déterministe** (**prototype**)

$$\mathscr{T} : \mathbb{R}^2 \to \mathbb{R}$$

  par :

  (i) un processus de déformation aléatoire du plan $D_n : \mathbb{R}^2 \to \mathbb{R}^2$,
  (ii) un processus de bruit additif $\mathscr{W}_n : \mathbb{R}^2 \to \mathbb{R}$

$$\forall u \in \mathbb{R}^2, \quad \mathscr{Y}_n(u) = \mathscr{T} \circ D_n(u) + \mathscr{W}_n(u)\ .$$

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

$$\textbf{sachant que } I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)} \ .$$

- Illustration sur un mélange de 5 classes :

$\mathscr{T}^{(1)}$ $\mathscr{T}^{(2)}$ $\mathscr{T}^{(3)}$ $\mathscr{T}^{(4)}$ $\mathscr{T}^{(5)}$

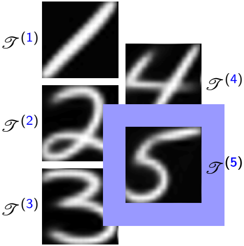
5 prototypes

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

$$\textbf{sachant que } I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)} \ .$$

- Illustration sur un mélange de 5 classes :



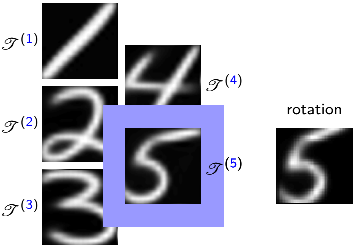
5 prototypes

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i$, $\qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :



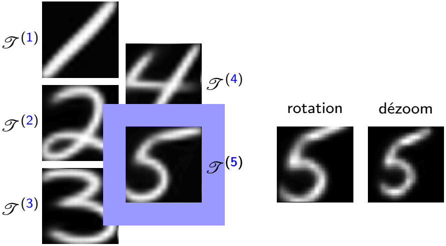
5 prototypes

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  $$\textbf{sachant que } I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)} \ .$$

- Illustration sur un mélange de 5 classes :



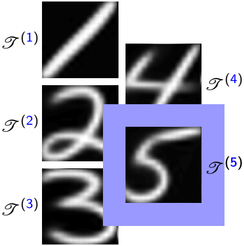
5 prototypes

Adaptation à un modèle de mélange

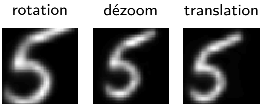
- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i$, $\qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :



5 prototypes

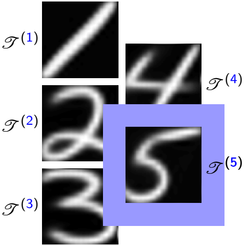


Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

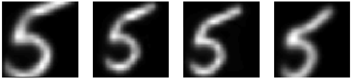
  sachant que $I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :



5 prototypes

rotation | dézoom | translation | deformations locales



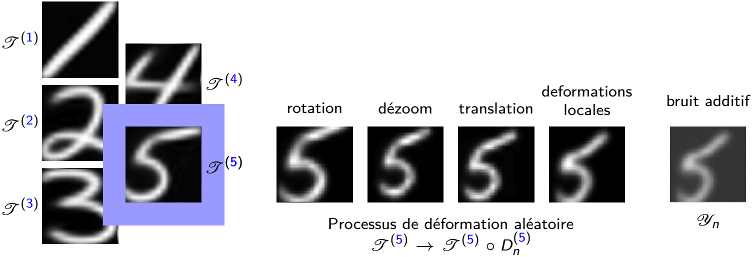
Processus de déformation aléatoire
$\mathscr{T}^{(5)} \to \mathscr{T}^{(5)} \circ D_n^{(5)}$

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i$, $\quad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :



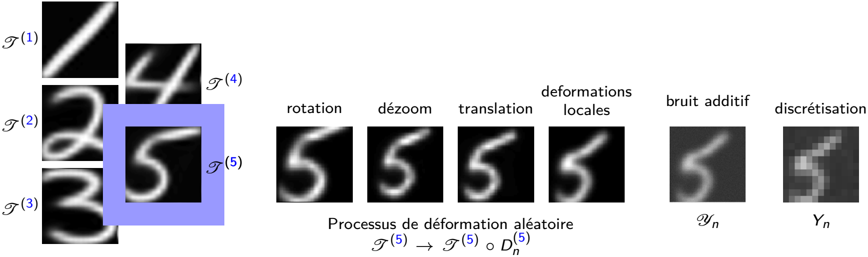
5 prototypes

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :



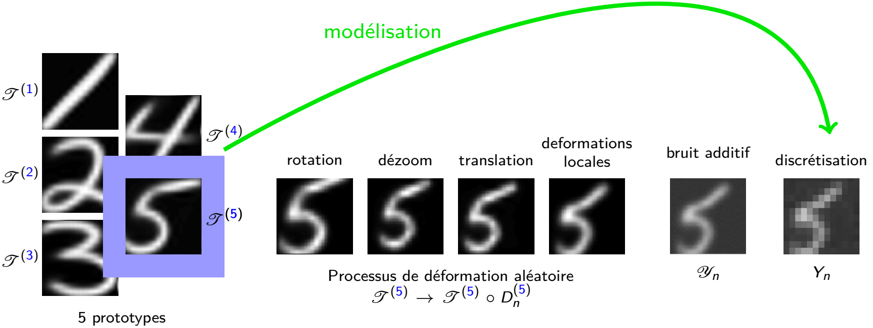
5 prototypes

Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :

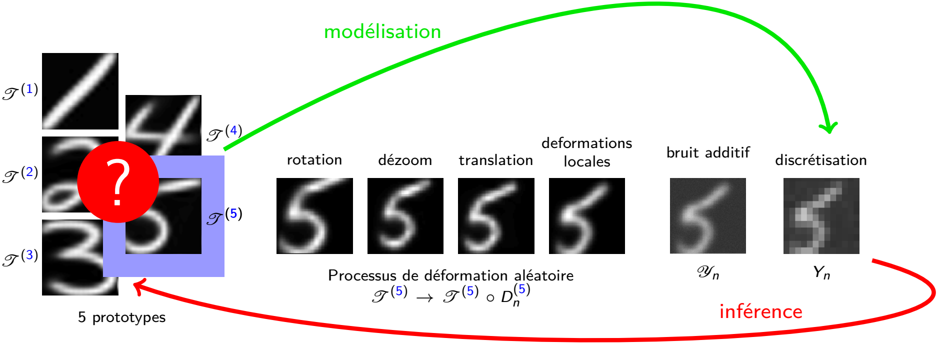


Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :

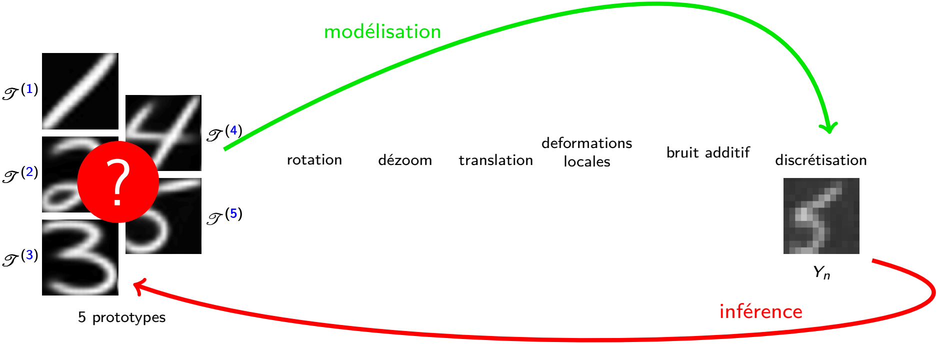


Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)},\ i \in \mathrm{I}\}$

  sachant que $I_n = i, \qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :

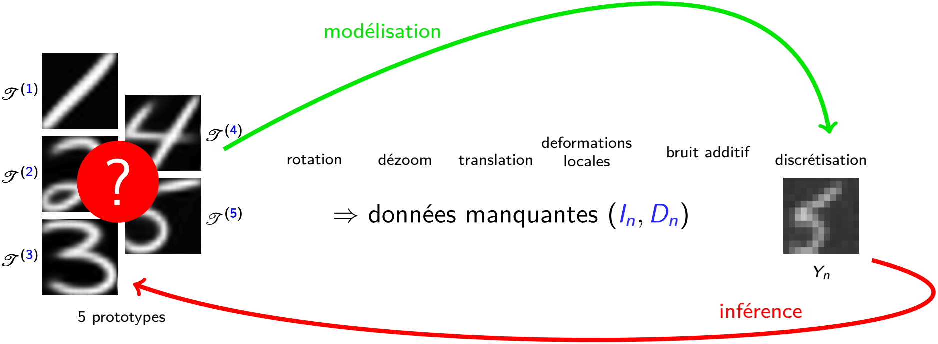


Adaptation à un modèle de mélange

- Existence de **plusieurs** prototypes $\{\mathscr{T}^{(i)}, i \in \mathrm{I}\}$

  sachant que $I_n = i$, $\qquad \mathscr{Y}_n = \mathscr{T}^{(i)} \circ D_n^{(i)} + \mathscr{W}_n^{(i)}$.

- Illustration sur un mélange de 5 classes :



$\Rightarrow$ données manquantes (I_n, D_n)

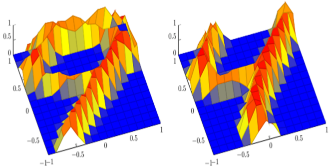
Modèle de déformation dans le cas des SIR

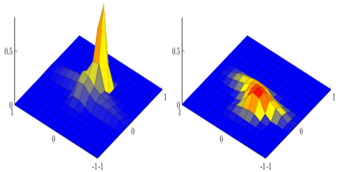
(1) Déformation du plan

$$D = D_\beta \begin{cases} \text{rotation} \\ \text{translation} \\ \text{déformation locale} \end{cases} \qquad \beta \in \mathrm{B} \subseteq \mathbb{R}^d$$

$$\beta \mid i \sim p_{\gamma^{(i)}} \qquad \gamma = (\gamma^{(1)}, \ldots, \gamma^{(C)})$$

(2) Déformation dans l'espace de mesure





$\Rightarrow$ facteur d'ajustement en amplitude $\lambda > 0$ $\quad \lambda \sim \mathrm{Gamma}(a, b)$

Modèle de SIR multispectrales

SIR multispectrale à K bandes $Y_n = (Y_{n,1}, \ldots, Y_{n,K})$ où $\{Y_{n,k} \in \mathrm{Y}\}_{k=1}^{K}$

- **Un** prototype $\mathscr{T}_k$ **par bande spectrale**

$$\mathscr{T}_k \to \mathscr{T}_{\alpha_k} = \sum_{\ell=1}^{m} \alpha_{k,\ell} \phi_\ell \ , \quad \alpha_k \in \mathrm{A} \subseteq \mathbb{R}^m \ ,$$

- **Une** variance de bruit $\{{\sigma_k}^2\}_{k=1}^{K}$ **par bande spectrale**,
- **Même** paramètres i, β, λ **pour toutes les bandes spectrales**,
- Modèle d'observation, **conditionnellement** à i_n, β_n, λ_n :

$$\begin{cases} \mathscr{Y}_n = (\mathscr{Y}_{n,1}, \ldots, \mathscr{Y}_{n,K}) \ , \\ \mathscr{Y}_{n,k} = \lambda_n \mathscr{T}_{\alpha_k(i_n)} \circ D_{\beta_n} + \sigma_k \mathscr{W}_{n,k} \ . \end{cases}$$

Problème initial : explorer les distributions $\mathbb{P}_i$

- Exploration des $\{\mathbb{P}_i,\ i \in \mathsf{I}\} \Longleftrightarrow$ Estimation du vecteur de paramètres $\theta \in \Theta \subset \mathbb{R}^p$

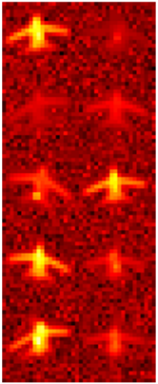
$$\theta = \Big(\underbrace{\Big\{ \alpha_1^{(i)}, \dots, \alpha_K^{(i)}}_{\text{Prototypes}}, \underbrace{\gamma^{(i)}}_{\text{Loi } \textit{apriori}\ \beta}, \underbrace{\omega^{(i)}}_{\text{Loi } \textit{apriori}\ I} \Big\}_{i \in \mathsf{I}}, \underbrace{\sigma_1, \dots, \sigma_K}_{\text{Variance bruit}} \Big) \ ,$$

- Estimation séquentielle de θ : Approximation stochastique de l'EM (partie II)

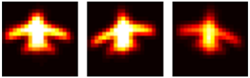
$$\{\hat{\theta}_n,\ n \in \mathbb{N}\} \to \hat{\theta} \ ,$$

Approximation des distributions conditionnelles $\{\mathbb{P}_i,\ i \in \mathsf{I}\}$ par $\{\mathbb{P}_{\hat{\theta}(i)},\ i \in \mathsf{I}\}$

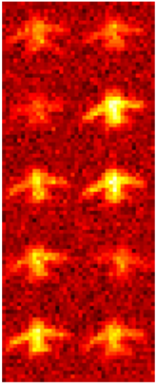
Illustration apprentissage



Exemples de données
$\{Y_n, n \in \mathbb{N}\}$



Prototypes estimés
$\mathscr{T}_{\hat{\theta}(1)}$, $\mathscr{T}_{\hat{\theta}(2)}$ et $\mathscr{T}_{\hat{\theta}(3)}$



Données simulées par le modèle appris $Y \sim \mathbb{P}_{\hat{\theta}}$

Quelques résultats de classification...

En monospectral :

Prédiction \ Réalité	Avion 1	Alphajet
Avion 1	0.912	0.060
Alphajet	0.088	0.940

Prédiction \ Réalité	Avion 1	Alphajet
Avion 1	0.718	0.092
Alphajet	0.282	0.908

semi-supervisé / non-supervisé & σ_2

Quelques résultats de classification...

En monospectral :

Réalité / Prédiction	Avion 1	Alphajet
Avion 1	0.912	0.060
Alphajet	0.088	0.940

Réalité / Prédiction	Avion 1	Alphajet
Avion 1	0.718	0.092
Alphajet	0.282	0.908

semi-supervisé / non-supervisé & σ_2

En multispectral :

Réalité / Prédiction	Avion 1	Alphajet
Avion 1	0.89	0.07
Alphajet	0.11	0.93

non-supervisé & σ_2

Plan de la présentation

Plan de la présentation

Cadre de travail

(i) $\{Y_n,\ n \in \mathbb{N}\}$ observations *i.i.d.* d'une loi $\mathbb{P}_\star$ **inconnue** sur $(\mathsf{Y}, \mathcal{Y})$,

Cadre de travail

(i) $\{Y_n,\ n \in \mathbb{N}\}$ observations *i.i.d.* d'une loi $\mathbb{P}_\star$ **inconnue** sur $(\mathsf{Y}, \mathcal{Y})$,

(ii) Présence d'un **processus latent** $\{X_n \in \mathsf{X},\ n \in \mathbb{N}\}$,

Cadre de travail

(i) $\{Y_n,\ n \in \mathbb{N}\}$ observations *i.i.d.* d'une loi $\mathbb{P}_\star$ **inconnue** sur $(\mathsf{Y}, \mathcal{Y})$,

(ii) Présence d'un **processus latent** $\{X_n \in \mathsf{X},\ n \in \mathbb{N}\}$,

(iii) **Modèle** données complètes $f_\theta : (\mathsf{X}, \mathsf{Y}) \to \mathbb{R}^+$ t.q. $\forall\, A \in \mathcal{Y}$

$$\mathbb{P}[Y_n \in A] = \mathbb{P}_\theta(A) = \int_A \underbrace{\int_{\mathsf{X}} f_\theta(x_n, y_n)\mathrm{d}x_n}_{f_\theta(y_n)\ \text{(incalculable)}} \mathrm{d}y_n \ ,$$

Cadre de travail

(i) $\{Y_n,\ n \in \mathbb{N}\}$ observations *i.i.d.* d'une loi $\mathbb{P}_\star$ **inconnue** sur $(\mathsf{Y}, \mathcal{Y})$,

(ii) Présence d'un **processus latent** $\{X_n \in \mathsf{X},\ n \in \mathbb{N}\}$,

(iii) **Modèle** données complètes $f_\theta : (\mathsf{X}, \mathsf{Y}) \to \mathbb{R}^+$ t.q. $\forall\, A \in \mathcal{Y}$

$$\mathbb{P}[Y_n \in A] = \mathbb{P}_\theta(A) = \int_A \underbrace{\int_{\mathsf{X}} f_\theta(x_n, y_n) \mathrm{d}x_n}_{f_\theta(y_n)\ \text{(incalculable)}} \mathrm{d}y_n \ ,$$

(iv) Modèle **exponentiel**

$$\log f_\theta(x, y) = \psi(\theta) + S(x, y)^{\mathsf{T}} \phi(\theta) \ ,$$

Cadre de travail

(i) $\{Y_n,\ n \in \mathbb{N}\}$ observations *i.i.d.* d'une loi $\mathbb{P}_\star$ **inconnue** sur $(\mathsf{Y}, \mathcal{Y})$,

(ii) Présence d'un **processus latent** $\{X_n \in \mathsf{X},\ n \in \mathbb{N}\}$,

(iii) **Modèle** données complètes $f_\theta : (\mathsf{X}, \mathsf{Y}) \to \mathbb{R}^+$ t.q. $\forall\, A \in \mathcal{Y}$

$$\mathbb{P}[Y_n \in A] = \mathbb{P}_\theta(A) = \int_A \underbrace{\int_{\mathsf{X}} f_\theta(x_n, y_n)\mathrm{d}x_n}_{f_\theta(y_n)\ \text{(incalculable)}} \mathrm{d}y_n \ ,$$

(iv) Modèle **exponentiel**

$$\log f_\theta(x, y) = \psi(\theta) + S(x, y)^{\mathsf{T}} \phi(\theta) \ ,$$

(v) Une **unique** observation est disponible à chaque instant (puis oubliée)

Cadre de travail

(i) $\{Y_n,\ n \in \mathbb{N}\}$ observations *i.i.d.* d'une loi $\mathbb{P}_\star$ **inconnue** sur $(\mathsf{Y}, \mathcal{Y})$,

(ii) Présence d'un **processus latent** $\{X_n \in \mathsf{X},\ n \in \mathbb{N}\}$,

(iii) **Modèle** données complètes $f_\theta : (\mathsf{X}, \mathsf{Y}) \to \mathbb{R}^+$ t.q. $\forall A \in \mathcal{Y}$

$$\mathbb{P}[Y_n \in A] = \mathbb{P}_\theta(A) = \int_A \underbrace{\int_{\mathsf{X}} f_\theta(x_n, y_n)\mathrm{d}x_n}_{f_\theta(y_n)\ \text{(incalculable)}} \mathrm{d}y_n \ ,$$

(iv) Modèle **exponentiel**

$$\log f_\theta(x, y) = \psi(\theta) + S(x, y)^{\mathsf{T}} \phi(\theta) \ ,$$

(v) Une **unique** observation est disponible à chaque instant (puis oubliée)

Problématique : trouver $\theta^\star \in \Theta$ t.q. $\mathbb{P}_{\theta^\star} \approx \mathbb{P}_\star$

Famille d'algorithmes Expectation-Maximization (EM)

- Densité de la distribution des données complètes vue par l'algorithme :

$$\pi_\theta(x, y) = \pi(y) \underbrace{\pi_\theta(x \mid y)}_{\propto f_\theta(x,y)} ,$$

- Séquence d'estimateurs $\{s_k,\ k \in \mathbb{N}\}$

$$s_0 \in \mathsf{S}\,, \quad s_{k+1} = \bar{s} \circ \bar{\theta}(s_k) \quad \text{où} \begin{cases} \bar{s} : \theta \to \mathbb{E}_{\pi_\theta}\left[S(X, Y)\right] \\ \bar{\theta} : s \to \arg\max_{\theta \in \Theta} \psi(\theta) + s^T \phi(\theta) \end{cases}$$

- Si n observations $\{Y_1, \ldots, Y_n\} \in \mathsf{Y}^n$ constamment disponibles

$\Longrightarrow \pi(y) = (1/n) \sum_{k=1}^{n} \delta_{\{y_k\}}(y) \Longrightarrow$ EM "classique" (Dempster et al, 1977)

$\Longrightarrow s_k \to s^\star \in \mathsf{S}^\star_\ell := \{s \in \mathsf{S},\ \nabla_s \ell(\bar{\theta}(s) \mid Y_1, \ldots, Y_n) = 0\}$

Cas séquentiel

Dans ce cas on a $\pi(y) = \mathbb{P}_\star(\mathrm{d}y)$ (qui est inconnue)

- L'algorithme standard n'existe pas car

$$s_k = \int_{\mathsf{Y}} \mathbb{E}_{\bar{\theta}(s_{k-1})}[S(X, Y) \mid Y = y] \, \mathbb{P}_\star(\mathrm{d}y)$$

- L'online EM (Cappé & Moulines, 2009), une **approximation stochastique** : séquence $\{\hat{s}_k,\ k \in \mathbb{N}\}$

$$\hat{s}_k = \hat{s}_{k-1} + \rho_k(\mathbb{E}_{\bar{\theta}(\hat{s}_{k-1})}[S(X_k, Y_k) \mid Y_k] - \hat{s}_{k-1}) \,, \quad Y_k \sim \mathbb{P}_\star$$

- Convergence

$$\hat{s}_k \to \hat{s}^\star \in \mathsf{S}^\star_{\mathsf{KL}} := \{s \in \mathsf{S},\ \nabla_s \mathsf{KL}\left(\mathbb{P}_\star \,\|\, \mathbb{P}_{\bar{\theta}(s)}\right) = 0\} \,,$$

- Cas où $s \to \mathbb{E}_{\bar{\theta}(s)}[S(X_k, Y_k) \mid Y_k]$ non-calculable??

Monte-Carlo online EM (MCoEM)

- MCoEM : séquence $\{\tilde{s}_k,\ k \in \mathbb{N}\}$

$$\text{(i)} \qquad Y_k \sim \mathbb{P}_\star, \qquad X_k^{(r)} \sim \pi_{\bar{\theta}(\tilde{s}_{k-1})}(\,\cdot\,|\,Y_k)\,,$$

$$\text{(ii)} \qquad \tilde{s}_k = \tilde{s}_{k-1} + \rho_k \left(\frac{1}{R} \sum_{r=1}^{R} S(X_k^{(r)}, Y_k) - \tilde{s}_{k-1} \right),$$

$(1/R)\sum_{r=1}^{R} S(X_k^{(r)}, Y_k)$ observation bruitée de s_k

- Convergence dans le cas où $\{X_k^{(1)}, \ldots, X_k^{(R)}\} \overset{i.i.d.}{\sim} \pi_{\bar{\theta}(\tilde{s}_{k-1})}(\,\cdot\,|\,Y_k)$

$$\tilde{s}_k \rightarrow \tilde{s}^\star \in \mathsf{S}^\star_{\mathrm{KL}}\ .$$

Application à un mélange de modèles déformables

- Modèle vectoriel sachant $I_n = i \in \{1, \dots, C\}$, $\beta_n = \beta \in \mathrm{B}$:

$$Y_n = \Phi_\beta \alpha_i + \sigma^2 W_n \,, \qquad W_n \sim \mathcal{N}(0, \mathrm{Id}_{|\mathrm{Y}|})$$

- Loi *aposteriori* :

$$\pi_\theta(i, \beta \mid Y = y) \propto f_\theta(i, \beta, y) = g(y; \Phi_\beta \alpha_i, \sigma^2 Id_{|\mathrm{Y}|})\, p_{\gamma^{(i)}}(\beta)\, \omega_i \ ,$$

- Implémentation MCoEM pour modèles à prototype déformable :

$$(I^{(r)}, \beta^{(r)}) \overset{i.i.d.}{\not\sim} \pi_\theta(\,\cdot \mid Y)$$

$$\Downarrow$$

$$(I^{(r)}, \beta^{(r)}) \sim P_\theta^{(r)}((i^{(0)}, \beta^{(0)}); \,\cdot \mid Y)$$

Simulation des données manquantes par Carlin et Chib

- MCMC usuels (Gibbs, Metropolis) inefficaces sur l'espace $\mathsf{X} = \mathsf{I} \times \mathsf{B}$

Simulation des données manquantes par Carlin et Chib

- MCMC usuels (Gibbs, Metropolis) inefficaces sur l'espace $\mathsf{X} = \mathsf{I} \times \mathsf{B}$

- (Carlin & Chib, 1995) : MCMC sur $\tilde{\mathsf{X}} = (\mathsf{I} \times \underbrace{\mathsf{B} \times \ldots \times \mathsf{B}}_{C})$ visant

$$\forall\, (i, \beta_1, \ldots, \beta_C) \in \tilde{\mathsf{X}}, \qquad \tilde{\pi}_\theta(i, \beta_1, \ldots, \beta_C \,|\, Y) = \pi_\theta(i, \beta_i \mid Y) \prod_{j \neq i} \xi_j(\beta_j) \;,$$

Simulation des données manquantes par Carlin et Chib

- MCMC usuels (Gibbs, Metropolis) inefficaces sur l'espace $\mathsf{X} = \mathsf{I} \times \mathsf{B}$

- (Carlin & Chib, 1995) : MCMC sur $\tilde{\mathsf{X}} = (\mathsf{I} \times \underbrace{\mathsf{B} \times \ldots \times \mathsf{B}}_{C})$ visant

$$\forall\, (i, \beta_1, \ldots, \beta_C) \in \tilde{\mathsf{X}}, \qquad \tilde{\pi}_\theta(i, \beta_1, \ldots, \beta_C \,|\, Y) = \pi_\theta(i, \beta_i \mid Y) \prod_{j \neq i} \xi_j(\beta_j)\ ,$$

- ***Pseudo-priors*** : $\{\xi_i,\ i \in \mathsf{I}\}$ densités sur $(\mathsf{B}, \mathcal{B})$

$$\xi_i \approx \pi_\theta(\,\cdot \mid Y, i)\ ,$$

Simulation des données manquantes par Carlin et Chib

- MCMC usuels (Gibbs, Metropolis) inefficaces sur l'espace $\mathsf{X} = \mathsf{I} \times \mathsf{B}$
- (Carlin & Chib, 1995) : MCMC sur $\tilde{\mathsf{X}} = (\mathsf{I} \times \underbrace{\mathsf{B} \times \ldots \times \mathsf{B}}_{C})$ visant

$$\forall\,(i, \beta_1, \ldots, \beta_C) \in \tilde{\mathsf{X}}, \qquad \tilde{\pi}_\theta(i, \beta_1, \ldots, \beta_C \,|\, Y) = \pi_\theta(i, \beta_i \mid Y) \prod_{j \neq i} \xi_j(\beta_j)\ ,$$

- ***Pseudo-priors*** : $\{\xi_i,\ i \in \mathsf{I}\}$ densités sur $(\mathsf{B}, \mathcal{B})$

$$\xi_i \approx \pi_\theta(\,\cdot \mid Y, i)\ ,$$

- Carlin et Chib est un Gibbs sur $\tilde{\mathsf{X}}$

$$\{I^{(r)}, \beta_1^{(r)}, \ldots, \beta_C^{(r)},\ r \in \mathbb{N}\} \sim \tilde{\pi}_\theta(\,\cdot \mid Y) \Longrightarrow \{I^{(r)}, \beta_{I^{(r)}}^{(r)},\ r \in \mathbb{N}\} \sim \pi_\theta(\,\cdot \mid Y)\,.$$

Illustration sur des images de chiffres manuscrits

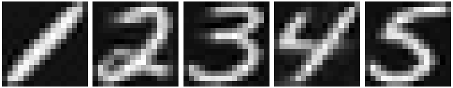
Quatre *implémentations*

	Déformation	prior β	prior θ	nb Data
❶ MCoEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \gamma_i \mathrm{Id}_{\lvert\beta\rvert})$	non	3000
❷ MCoEM	locale + rigide	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \gamma_i \mathrm{Id}_{\lvert\beta\rvert})$	non	3000
❸ MCoEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \Gamma_i)$	oui	3000
❹ SAEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \Gamma_i)$	oui	100

Illustration sur des images de chiffres manuscrits

Quatre *implémentations*

	Déformation	prior β	prior θ	nb Data
❶ MCoEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \gamma_i \mathrm{Id}_{\lvert\beta\rvert})$	non	3000
❷ MCoEM	locale + rigide	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \gamma_i \mathrm{Id}_{\lvert\beta\rvert})$	non	3000
❸ MCoEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \Gamma_i)$	oui	3000
❹ SAEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \Gamma_i)$	oui	100



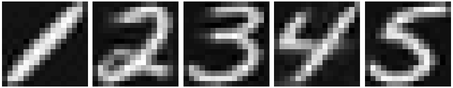
❶ - $\hat{\theta}_{3000}$

❷ - $\hat{\theta}_{3000}$

Illustration sur des images de chiffres manuscrits

Quatre *implémentations*

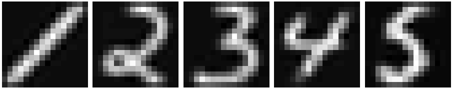
	Déformation	prior β	prior θ	nb Data
❶ MCoEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \gamma_i \mathrm{Id}_{\lvert\beta\rvert})$	non	3000
❷ MCoEM	locale + rigide	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \gamma_i \mathrm{Id}_{\lvert\beta\rvert})$	non	3000
❸ MCoEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \Gamma_i)$	oui	3000
❹ SAEM	locale	$\beta \mid i \sim \mathcal{N}_{\lvert\beta\rvert}(0, \Gamma_i)$	oui	100



❶ - $\hat{\theta}_{3000}$

❸ - $\hat{\theta}_{3000}$

❷ - $\hat{\theta}_{3000}$



❹ - $\hat{\theta}_{100}$

Conclusion

- Avantage du MCoEM
  - + Cadre d'application plus général que l'online EM ,
  - + Rapidité comparé au SAEM,
  - + Complexité indépendante de n,
  - + Évaluer des tendances...

- Limites
  - – Sensibilité Y,
  - – Sensibilité $(I, \beta) \sim \pi_\theta(\cdot \mid Y)$,
  - – Convergence théorique $\hat{s}_k \to \hat{s}^\star \in \mathsf{S}^\star_{\mathsf{KL}}$ à prouver dans le cas non iid...

Plan de la présentation

Définitions

- Soit π une distribution sur $(\mathsf{X}, \mathcal{X})$,

$$\pi f := \int f(x)\pi(\mathrm{d}x)\,, \quad \textbf{(Quantité incalculable)}$$

- MCMC : à partir d'un noyau de Markov $P : (X, \mathcal{X}) \to (0,1)$

$$\widehat{\pi f}_n := \frac{1}{n}\sum_{k=1}^{n} f(X_k) \approx \pi f\,, \quad \text{où} \quad X_k \sim P^k(x_0, \cdot\,)$$

- (Au minimum) P : π-stationnaire *i.e*

$$\pi(A) = \int \pi(\mathrm{d}x)P(x, A)\,, \qquad\qquad (\pi P = \pi)$$

- (Au mieux) P : π-réversible *i.e*

$$\int_A \pi(\mathrm{d}x)P(x, B) = \int_B \pi(\mathrm{d}x)P(x, A)\,.$$

Comparaison

Pour une quantité πf à estimer, un grand nombre de MCMC possibles

- P_0 et P_1 deux noyaux π-réversible
- Trouver une condition sur P_0 et P_1 telle que $\forall f \in \mathrm{F} \subseteq L_2(\pi)$:

$$v(f, P_0) \leq v(f, P_1)\,,$$

où

$$v(f, P) := \lim_{n \to \infty} \frac{1}{n} Var\left(\sum_{k=0}^{n-1} f(X_k)\right) = \lim_{n \to \infty} \sqrt{n} Var\left(\hat{\pi}_n f\right)\ .$$

Comparer des noyaux?

- ordre **hors-diagonal** $P_1 \succeq P_0$:

$$\forall\,(x,A) \in (\mathsf{X},\mathcal{X})\,, \qquad P_1(x,A\setminus\{x\}) \geq P_0(x,A\setminus\{x\})\,,$$

- ordre de **covariance** (d'ordre 1) $P_1 \succcurlyeq_1 P_0$:

$$\forall\, f \in L_2(\pi)\,, \qquad \langle f\,,P_1 f\rangle \leq \langle f\,,P_0 f\rangle\ ,$$

où $\langle f\,,g\rangle = \int f(x)g(x)\pi(\mathrm{d}x)\,.$

Théorèmes (Peskun, 1973) & (Tierney, 1998) :

$$P_1 \succeq P_0 \Longrightarrow P_1 \succcurlyeq_1 P_0 \Longrightarrow v(f\,,P_1) \leq v(f\,,P_0)$$

Que peut-on dire pour les chaînes de Markov inhomogènes?

- Chaînes inhomogènes : $\{X_k,\ k \in \mathbb{N}\}$ évoluent suivant le schéma

$$\pi \sim X_0^{(i)} \xrightarrow{P_i} X_1^{(i)} \xrightarrow{Q_i} X_2^{(i)} \xrightarrow{P_i} X_3^{(i)} \xrightarrow{Q_i} \cdots$$

- A-t-on un résultat *équivalent* à Tierney?
- Et si oui pour quel ensemble de fonctions $f \in \mathrm{F} \subseteq L_2(\pi)$?

Théorème (Maire, Douc & Olsson, 2013) :

$$\left\{ \begin{array}{ll} \forall f \in \mathrm{F} & \\ (P_0,\ P_1)\ \pi\text{-réversibles} & P_1 \succeq P_0 \\ (Q_0,\ Q_1)\ \pi\text{-réversibles} & Q_1 \succeq Q_0 \end{array} \right. \Longrightarrow v(f, P_1, Q_1) \leq v(f, P_0, Q_0)$$

Une nouvelle preuve du Théorème de Tierney

- Tierney : décomposition spectrale des opérateurs auto-adjoints ($\Rightarrow$ restriction aux cas des chaînes homogènes...)
- Avec une hypothèse supplémentaire (qui définit F) pour $i \in \{0, 1\}$

$$\text{(H1)} \quad \sum_{k=1}^{\infty} \left(|\mathrm{Cov}(f(X_0^{(i)}), f(X_k^{(i)}))| + |\mathrm{Cov}(f(X_1^{(i)}), f(X_{k+1}^{(i)}))| \right) < \infty ,$$

on démontre le résultat de Tierney sans recours à la théorie spectrale avec bien sûr

$$\text{(H2)-a} \quad P_1 \text{ et } P_0 \ \pi\text{-réversibles} ,$$

$$\text{(H2)-b} \quad P_1 \succeq P_0 .$$

Squelette de la preuve...

(1) Pour $f \in \mathrm{F}$ (H1) et P π-réversible (H2)

$$v(f, P) = \pi f^2 - \pi^2 f + 2 \sum_{n=0}^{\infty} \underbrace{\mathrm{Cov}(f(X_1), f(X_n))}_{\langle f, P^n f \rangle} ,$$

(2) Définissons

$$\begin{cases} \forall\, \alpha \in (0,1) & P_\alpha = (1-\alpha) P_0 + \alpha P_1 \ , \\ \forall\, \lambda \in (0,1) & w_\lambda(f, P_\alpha) = \sum_{n=1}^{\infty} \lambda^n \left\langle f, P_\alpha^n f \right\rangle \end{cases}$$

(3) et montrons que $\forall \lambda \in (0,1)$, $\alpha \to w_\lambda(f, P_\alpha)$ décroissante sur $(0,1)$.

(4) Concluons par un théorème de convergence dominée $\lambda \to 1$:

$$\left\langle f, P_1^n f \right\rangle \leq \left\langle f, P_0^n f \right\rangle \ .$$

Corollaire pour schémas de Data Augmentation

- Data Augmentation :

$$\forall x \in \mathsf{X},\ x = \begin{bmatrix} z \in \mathsf{Z} & \text{(paramètre d'intérêt)} \\ u \in \mathsf{U} & \text{(variable auxiliaire)} \end{bmatrix}$$

- Problème : estimer πf, $(f \in \mathsf{F})$ où

$$\forall\, x \in \mathsf{X},\ f(x) = f(z)\,,$$

- $\{X_k^{(i)} = (Z_k^{(i)}, U_k^{(i)}),\ k \in \mathbb{N}\}$ de noyau $K^{(i)}$ non π-réversible
- $\{\tilde{X}_k^{(i)},\ k \in \mathbb{N}\}$ de noyaux $(P^{(i)}, Q^{(i)})$ π-réversibles, t.q.

$$\forall\, k \in \mathbb{N},\ \tilde{X}_{2k}^{(i)} = X_k^{(i)}\,,$$

- Notre théorème donne :

$$\begin{cases} P^{(1)} \succeq P^{(0)} \\ Q^{(1)} \succeq Q^{(0)} \end{cases} \Longrightarrow v(f, K^{(1)}) \leq v(f, K^{(0)})\,.$$

Utilité de ce résultat?

- Comparer des MCMC existants en les mettant sous le schéma $P \to Q \to P \to \dots$
  $\Rightarrow$ Possible de comparer des chaînes qui sont *apriori* homogènes et non réversibles
- Créer des MCMC
  $\Rightarrow$ En augmentant le taux de mélange dans la chaîne (par ex. en ajoutant des variables auxiliaires)
- Applications :
  $\Rightarrow$ Méthodes de Data-Augmentation,
  $\Rightarrow$ Algorithmes Pseudo-Marginal,
  $\Rightarrow$ MCMC pour des modèles de mélange (Carlin et Chib),
  $\Rightarrow$ Approximate Bayesian Computation,
  $\Rightarrow$ ...

Plan de la présentation

Détection & classification multispectrale

- Vers un modèle de prototype spectro-spatial...

$$Y_{n,k} = \int_{\nu_k}^{\nu_{k+1}} D \circ \mathscr{T}(u, \nu) \mathrm{d}\nu + \sigma_k^2$$

- Pour une meilleure exploitation de l'information spectrale discriminante

Questions méthodologiques

- Vers une convergence théorique du MCoEM couplé avec un MCMC
- Cas des chaînes de Markov inhomogènes alternant entre n noyaux de transition

Des tests de détection non appropriés

- Test de rapport de vraisemblances (Neyman & Pearson, 1992) :

$$\phi(Y) \to \begin{cases} \mathcal{H}_0 & \text{si} \quad \Lambda(Y) \leq \eta \\ \mathcal{H}_1 & \text{si} \quad \Lambda(Y) > \eta \end{cases} \qquad \text{où} \quad \Lambda(Y) = \frac{\mu_1(Y)}{\mu_0(Y)} .$$

- D'autres tests nécessitent une base d'apprentissage pour estimer μ_1 ce que nous n'avons pas... (*e.g* GLRT)

- Test du Filtre adapté (Manolakis et al., 2003) : hypothèse que μ_1 est Gaussienne (non acceptable dans notre cas...)

Calibrage du test (1/2)

- Contrôle du taux de fausse alarme :

$\Rightarrow$ Comme pour ϕ_{RX} , on souhaite connaître f telle que $(\eta, \nu) = f(p)$, où $p = P_{\mathrm{FA}}(\phi)$

$$
\begin{aligned}
p &= \mathbb{P}\left[\max \mathrm{Per}\{A_\eta(Y)\} > \nu \mid Y \in \mathcal{H}_0\right] , \\
&= \mathbb{P}\Big[\underbrace{\max \mathrm{Per}\{x \in \mathbb{R}^2, \sum_{j=1}^{P} \alpha_j \psi_j(x) > \eta\} > \nu}_{G_{\eta,\nu}(\boldsymbol{\alpha})}\Big] \qquad \text{où} \quad \alpha_j \overset{iid}{\sim} \chi^2_K , \\
&= \mathbb{E}\left[\mathbb{1}_{G_{\eta,\nu}(\boldsymbol{\alpha})}\right] = \int \cdots \int \mathbb{1}_{G_{\eta,\nu}(\boldsymbol{\alpha})} \prod_{j=1}^{P} \chi^2_K(\mathrm{d}\alpha_j) = h(\eta, \nu) .
\end{aligned}
$$

- h n'est pas calculable explicitement, donc $f := h^{-1}$ (si h est inversible) non plus.

Calibrage du test (2/2)

- $p = h(\eta, \nu)$ non calculable
- Approximation de h par la fonction $(\hat{h}_n)_n$ telle que $\hat{h}_n \to h$ (convergence simple presque-sûrement par la LGN)

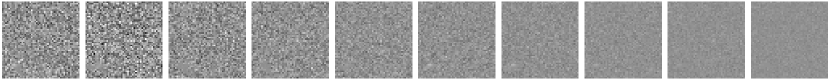
$$\hat{h}_n(\eta, \nu) := n^{-1} \sum_{\ell=1}^{n} \mathbb{1}_{G_{\eta,\nu}(\boldsymbol{\alpha}^{(\ell)})} \ , \quad \boldsymbol{\alpha}^{(\ell)} = (\alpha_1^{(\ell)}, \ldots, \alpha_P^{(\ell)}) \sim \otimes_{\ell=1}^{P} \chi_K^2 \ ,$$

- Pour un taux de fausse alarme $p^* > 0$ fixé, calculer les seuils $\hat{\eta}_n^* > 0$ et $\hat{\nu}_n^* > 0$ définis par :

$$\hat{f}_n(p^*) := (\hat{\eta}_n^*, \hat{\nu}_n^*) := \arg \min_{\eta>0, \nu>0} |\hat{h}_n(\eta, \nu) - p^*| \ .$$

Information spatiale

Localisation d'une cible dans une image de taille 60×60 par ϕ

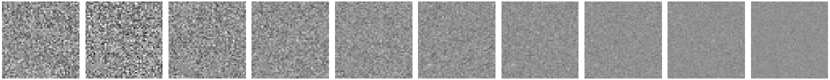


Y

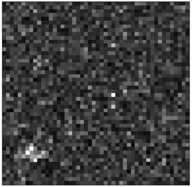
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$

Information spatiale

Localisation d'une cible dans une image de taille 60×60 par ϕ



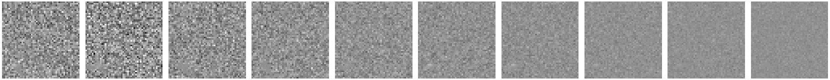
Y



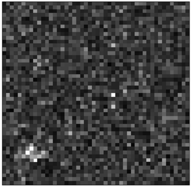
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$

Information spatiale

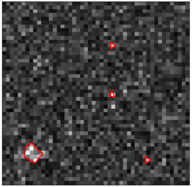
Localisation d'une cible dans une image de taille 60×60 par ϕ



Y



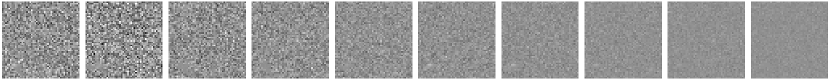
$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$



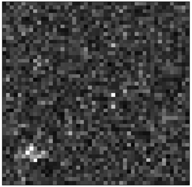
$A_\eta(Y)$

Information spatiale

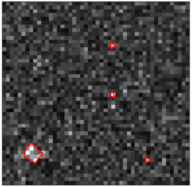
Localisation d'une cible dans une image de taille 60×60 par ϕ



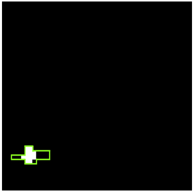
Y



$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$



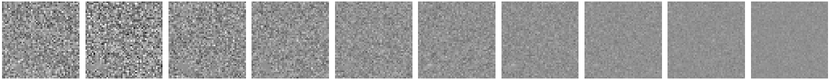
$A_\eta(Y)$



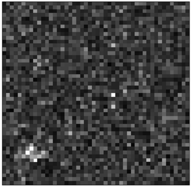
Per $A_\eta(Y) > \nu$

Information spatiale

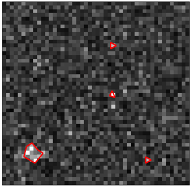
Localisation d'une cible dans une image de taille 60×60 par ϕ



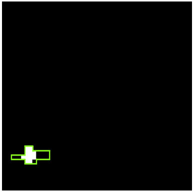
Y



$\{D_{\mathrm{M}}(y_j, \mu_0),\ 1 \leq j \leq P\}$



$A_\eta(Y)$



Per $A_\eta(Y) > \nu$

	Niveau pixel		Niveau image	
	$P_{\mathrm{D}}(\phi)$	$P_{\mathrm{FA}}(\phi)$	$P_{\mathrm{D}}(\phi)$	$P_{\mathrm{FA}}(\phi)$
data mélangé	0.4	0.003	0.976	0.003
data : avion A	0.4	0.004	0.982	0.004
data : avion B	0.5	0.002	0.971	0.005

Approximation Stochastique

Rechercher les racines d'une fonction non connue explicitement

- Soit $h : \mathsf{S} \to \mathsf{S}$

$$\mathsf{S}^\star = \{s \in \mathsf{S}\,, \quad h(s) = 0\}$$

- Approximation stochastique (Robbins & Monro, 1951)

$$\hat{s}_0 \in \mathsf{S}\,, \qquad \hat{s}_{n+1} = \hat{s}_n + \rho_{n+1}\underbrace{\big(h(s_n) + \zeta_n\big)}_{\hat{h}_n}\,,$$

  où

  - $\hat{h}_n$ est une observation bruitée de $h(s_n)$,
  - $\{\rho_n,\ n \in \mathbb{N}\}$ une séquence décroissante de nombres positifs,

$$\begin{cases} \rho_n \to 0 \\ \sum_{n\geq 1} \rho_n = \infty \end{cases}$$

  - $\{\zeta_n,\ n \in \mathbb{N}\}$ un processus aléatoire de bruit.

Convergence Approximation Stochastique

Théorème de convergence (Andrieu, Moulines & Priouret, 2005)

Principales conditions :

- Existence d'une fonction de Lyapunov $w : \mathsf{S} \to \mathsf{S}$ pour h t.q.

$$\forall s \in \mathsf{S}\,, \quad \langle \nabla_s w(s), h(s) \rangle \leq 0 \quad \& \quad \langle \nabla_s w(s), h(s) \rangle = 0 \Leftrightarrow h(s) = 0$$

- Contrôle du bruit

$$\lim_{n \to \infty} \sup_{\ell \geq k} \left| \sum_{k=n}^{\ell} \rho_k \zeta_k \right| = 0$$

Alors

$$s_n \to s^\star \in \mathsf{S}^\star \quad \text{w.p.1}$$

EM stochastique

- Soit $h = \bar{s} \circ \bar{\theta} - Id$

$$\forall\, s \in \mathsf{S}\,, \quad h(s) = 0 \Longleftrightarrow \nabla_s \ell(\bar{\theta}(s) \,|\, Y_1, \ldots, Y_n) = 0\ ,$$

EM stochastique

- Soit $h = \bar{s} \circ \bar{\theta} - Id$

$$\forall\, s \in \mathsf{S}\,, \quad h(s) = 0 \Longleftrightarrow \nabla_s \ell(\bar{\theta}(s) \mid Y_1, \ldots, Y_n) = 0\ ,$$

Hypothèses:

(i) $\bar{s} : \theta \to \sum_{i=1}^{n} \mathbb{E}_\theta \left[S(X_i, Y_i) \mid Y_i \right]$ non calculable

(ii) versions bruitées $\tilde{s}(\theta_k) = \bar{s}(\theta_k) + \zeta_k$ observées

EM stochastique

- Soit $h = \bar{s} \circ \bar{\theta} - Id$

$$\forall\, s \in \mathsf{S}\,, \quad h(s) = 0 \Longleftrightarrow \nabla_s \ell(\bar{\theta}(s) \,|\, Y_1, \ldots, Y_n) = 0\ ,$$

  Hypothèses:
  (i) $\bar{s} : \theta \to \sum_{i=1}^{n} \mathbb{E}_\theta \left[S(X_i, Y_i) \,|\, Y_i \right]$ non calculable
  (ii) versions bruitées $\tilde{s}(\theta_k) = \bar{s}(\theta_k) + \zeta_k$ observées

- Possible de trouver les racines de h à partir de $\tilde{s}$ en remplaçant

$$s_k = \bar{s}(\theta_k) \qquad \text{par} \qquad \hat{s}_k = \hat{s}_{k-1} + \rho_k(\tilde{s}(\theta_{k-1}) - \hat{s}_{k-1})\ ,$$

EM stochastique

- Soit $h = \bar{s} \circ \bar{\theta} - Id$

$$\forall\, s \in \mathsf{S}\,, \quad h(s) = 0 \Longleftrightarrow \nabla_s \ell(\bar{\theta}(s) \,|\, Y_1, \ldots, Y_n) = 0\;,$$

  Hypothèses:
  (i) $\bar{s} : \theta \to \sum_{i=1}^{n} \mathbb{E}_\theta \left[S(X_i, Y_i) \,|\, Y_i \right]$ non calculable
  (ii) versions bruitées $\tilde{s}(\theta_k) = \bar{s}(\theta_k) + \zeta_k$ observées

- Possible de trouver les racines de h à partir de $\tilde{s}$ en remplaçant

$$s_k = \bar{s}(\theta_k) \qquad \text{par} \qquad \hat{s}_k = \hat{s}_{k-1} + \rho_k(\tilde{s}(\theta_{k-1}) - \hat{s}_{k-1})\;,$$

- Preuve de convergence de $\{\hat{s}_k,\ k \in \mathbb{N}\}$ vers $\mathsf{S}^\star := \{s \in \mathsf{S},\ h(s) = 0\}$ (Robbins-Monro)

EM stochastique

- Soit $h = \bar{s} \circ \bar{\theta} - Id$

$$\forall\, s \in \mathsf{S}\,, \quad h(s) = 0 \Longleftrightarrow \nabla_s \ell(\bar{\theta}(s) \,|\, Y_1, \ldots, Y_n) = 0\,,$$

  Hypothèses:
  (i) $\bar{s} : \theta \to \sum_{i=1}^{n} \mathbb{E}_\theta\left[S(X_i, Y_i) \,|\, Y_i\right]$ non calculable
  (ii) versions bruitées $\tilde{s}(\theta_k) = \bar{s}(\theta_k) + \zeta_k$ observées
- Possible de trouver les racines de h à partir de $\tilde{s}$ en remplaçant

$$s_k = \bar{s}(\theta_k) \qquad \text{par} \qquad \hat{s}_k = \hat{s}_{k-1} + \rho_k(\tilde{s}(\theta_{k-1}) - \hat{s}_{k-1})\,,$$

- Preuve de convergence de $\{\hat{s}_k,\, k \in \mathbb{N}\}$ vers $\mathsf{S}^\star := \{s \in \mathsf{S},\, h(s) = 0\}$ (Robbins-Monro)
- $\{\hat{\theta}_k,\, k \in \mathbb{N}\}$ devient **stochastique** $\Rightarrow$ SAEM, SAEM-MCMC...

Reformulation

- Soit π la loi des données utilisées lors de l'inférence

Reformulation

- Soit π la loi des données utilisées lors de l'inférence
- Dans le cas de EM : loi uniforme sur $(Y_1, \ldots, Y_n)$,

$$\pi^{(n)}(\mathrm{d}y) = 1/n \sum_{k=1}^{n} \delta_{Y_k}(\mathrm{d}y) \, ,$$

Reformulation

- Soit π la loi des données utilisées lors de l'inférence
- Dans le cas de EM : loi uniforme sur $(Y_1, \dots, Y_n)$,

$$\pi^{(n)}(\mathrm{d}y) = 1/n \sum_{k=1}^{n} \delta_{Y_k}(\mathrm{d}y) \ ,$$

- Remarque :

$$\bar{s} \circ \bar{\theta}(s) = \frac{1}{n} \sum_{k=1}^{n} \int_{\mathsf{X}} S(x_k, Y_k) \pi_{\bar{\theta}(s)}(\mathrm{d}x_k \mid Y_k) = \iint_{\mathsf{YX}} S(x, y) \pi^{(n)}(\mathrm{d}x, \mathrm{d}y) \ ,$$

où $\pi^{(n)}(\mathrm{d}x, \mathrm{d}y) = \pi^{(n)}(\mathrm{d}y) \pi_{\bar{\theta}(s)}(\mathrm{d}x \mid y)$

Reformulation

- Soit π la loi des données utilisées lors de l'inférence
- Dans le cas de EM : loi uniforme sur $(Y_1, \ldots, Y_n)$,

$$\pi^{(n)}(\mathrm{d}y) = 1/n \sum_{k=1}^{n} \delta_{Y_k}(\mathrm{d}y) \; ,$$

- Remarque :

$$\bar{s} \circ \bar{\theta}(s) = \frac{1}{n} \sum_{k=1}^{n} \int_{\mathsf{X}} S(x_k, Y_k) \pi_{\bar{\theta}(s)}(\mathrm{d}x_k \mid Y_k) = \iint\limits_{\mathsf{YX}} S(x, y) \pi^{(n)}(\mathrm{d}x, \mathrm{d}y) \; ,$$

où $\pi^{(n)}(\mathrm{d}x, \mathrm{d}y) = \pi^{(n)}(\mathrm{d}y) \pi_{\bar{\theta}(s)}(\mathrm{d}x \mid y)$

- Réécriture de h :

$$h(s) = (1/n) \sum_{k=1}^{n} \mathbb{E}_{\bar{\theta}(s)}[S(X_k, Y_k) \mid Y_k] - s = \mathbb{E}^{(n)}_{\mathrm{inf}, \bar{\theta}(s)}[S(X, Y)] - s \; .$$

Equivalence en apprentissage séquentiel

- Dans le cas séquentiel à l'itération n, une seule observation $Y_n \sim \mathbb{P}_\star$

$$\pi(\mathrm{d}y) = \mathbb{P}_\star(\mathrm{d}y) \ ,$$

Equivalence en apprentissage séquentiel

- Dans le cas séquentiel à l'itération n, une seule observation $Y_n \sim \mathbb{P}_\star$

$$\pi(\mathrm{d}y) = \mathbb{P}_\star(\mathrm{d}y) \ ,$$

- h incalculable car $\mathbb{P}_\star$ inconnue

$$h(s) = \int_Y \int_X S(x,y)\pi_{\bar{\theta}(s)}(\mathrm{d}x \,|\, y)\mathbb{P}_\star(\mathrm{d}y) - s \ ,$$

Equivalence en apprentissage séquentiel

- Dans le cas séquentiel à l'itération n, une seule observation $Y_n \sim \mathbb{P}_\star$

$$\pi(\mathrm{d}y) = \mathbb{P}_\star(\mathrm{d}y) \, ,$$

- h incalculable car $\mathbb{P}_\star$ inconnue

$$h(s) = \int_Y \int_X S(x,y) \pi_{\bar{\theta}(s)}(\mathrm{d}x \,|\, y) \mathbb{P}_\star(\mathrm{d}y) - s \, ,$$

- online EM : approximation stochastique de h par

$$\tilde{h}(s) = \int_X S(x, Y_n) \pi_{\bar{\theta}(s)}(\mathrm{d}x \,|\, Y_n) - s \, ,$$

Double approximation stochastique : MCoEM

- **Deux** intégrales incalculables

$$h(s) = \int_Y \int_X S(x, y) \pi_{\bar{\theta}(s)}(\mathrm{d}x \mid y) \mathbb{P}_\star(\mathrm{d}y) - s \, ,$$

Double approximation stochastique : MCoEM

- **Deux** intégrales incalculables

$$h(s) = \int_Y \int_X S(x, y)\pi_{\bar{\theta}(s)}(\mathrm{d}x \mid y)\mathbb{P}_\star(\mathrm{d}y) - s \,,$$

- MCoEM : *Double* approximation stochastique de h par

$$\check{h}(s) = \frac{1}{R}\sum_{r=1}^{R} S(X_n^{(r)}, Y_n) - s \,, \quad Y_n \sim \mathbb{P}_\star,\ X_n^{(r)} \sim \pi_{\bar{\theta}(s)}(\,\cdot \mid Y_n) \,,$$

Double approximation stochastique : MCoEM

- **Deux** intégrales incalculables

$$h(s) = \int_Y \int_X S(x, y)\pi_{\bar{\theta}(s)}(\mathrm{d}x \mid y)\mathbb{P}_\star(\mathrm{d}y) - s \, ,$$

- MCoEM : *Double* approximation stochastique de h par

$$\check{h}(s) = \frac{1}{R}\sum_{r=1}^{R} S(X_n^{(r)}, Y_n) - s \, , \quad Y_n \sim \mathbb{P}_\star, \, X_n^{(r)} \sim \pi_{\bar{\theta}(s)}(\, \cdot \mid Y_n) \, ,$$

- Convergence dans le cas où $\{X_n^{(1)}, \ldots, X_n^{(R)}\} \overset{i.i.d.}{\sim} \pi_{\hat{\theta}_{n-1}}(\, \cdot \mid Y_n)$

$$\hat{\theta}_n \to \theta^\star \in \Theta_{\mathrm{KL}}^\star := \{\theta \in \Theta, \, \nabla_\theta \mathrm{KL}\left(\mathbb{P}_\star \,\|\, \mathbb{P}_\theta\right) = 0\} \, ,$$

Comparaison bloc / en ligne

$$\boxed{h(s) = \mathbb{E}_\pi \mathbb{E}_{\bar{\theta}(s)}[S(X, Y) \,|\, Y] - s}$$

	En bloc		En ligne	
distribution obs.	$\pi^{(n)}(\mathrm{d}y) = n^{-1} \sum_{k=1}^{n} \delta_{Y_k}(\mathrm{d}y)$		$\pi(\mathrm{d}y) = \mathbb{P}_\star(\mathrm{d}y)$	
propriété	$h(s) = 0 \Leftrightarrow \nabla_\theta \ell(\theta \,\vert\, Y_1, \ldots, Y_n) = 0$		$h(s) = 0 \Leftrightarrow \nabla_\theta \mathsf{KL}\left(\mathbb{P}_\star \,\Vert\, \mathbb{P}_\theta\right) = 0$	
convergence	$\{\theta_k, k \in \mathbb{N}\} \to \nabla_\theta \ell(\theta \,\vert\, Y_1, \ldots, Y_n) = 0$		$\{\theta_k, k \in \mathbb{N}\} \to \nabla_\theta \mathsf{KL}\left(\mathbb{P}_\star \,\Vert\, \mathbb{P}_\theta\right) = 0$	
dynamique	$\ell(\theta_{k+1} \,\vert\, Y_1, \ldots, Y_n) \leq \ell(\theta_k \,\vert\, Y_1, \ldots, Y_n)$		$\mathsf{KL}\left(\mathbb{P}_\star \,\Vert\, \mathbb{P}_{\theta_{k+1}}\right) \leq \mathsf{KL}\left(\mathbb{P}_\star \,\Vert\, \mathbb{P}_{\theta_k}\right)$	
(argument)	(inégalité Jensen)		(existence fonction de Lyapunov)	
algorithmes	EM	SAEM	online EM	MCoEM
$\mathbb{E}_{\mathrm{inf}}[G(Y)]$	exacte	exacte	$\approx G(Y_n)$	$\approx G(Y_n)$
$\mathbb{E}_\theta[S(X, Y) \,\vert\, Y]$	exacte	$\approx \frac{1}{R} \sum_{k=1}^{R} S(X^{(r)}, Y)$	exacte	$\approx \frac{1}{R} \sum_{k=1}^{R} S(X_n^{(r)}, Y_n)$
Approx. sto.	non	oui	oui	oui

Simulation Pseudo-priors

- Heuristique : $\{X_r^{(i)}, r \in \mathbb{N}\}$ visant $\pi_{\hat{\theta}}(\cdot \mid Y, i)$
- $\varrho_i = \mathcal{N}(\hat{\mu}^{(i)}, \hat{\Sigma}^{(i)})$

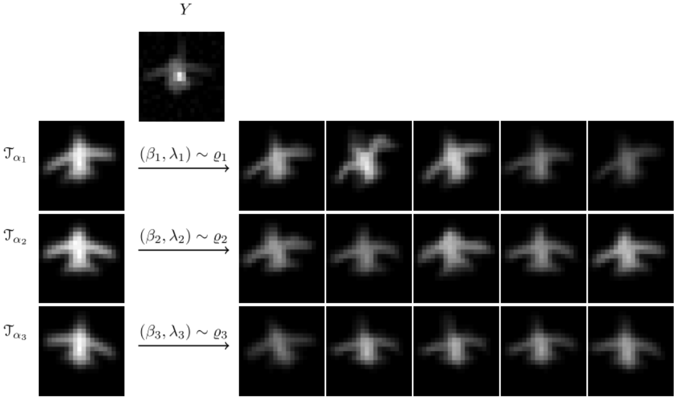
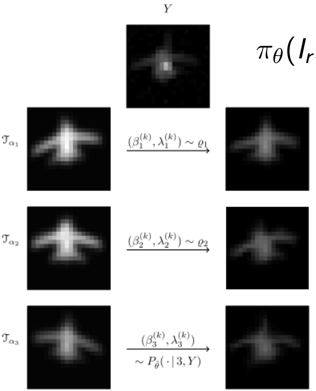


Illustration Carlin & Chib

$$\pi_\theta(I_{r+1} = i | X_{r+1}^{(i)}, Y)$$

$I_r = 3 \longrightarrow$



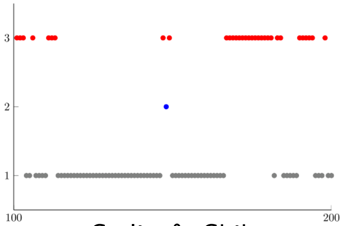
0.08

0.05

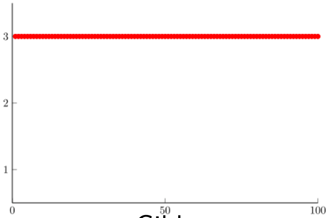
0.87

$\longrightarrow I_{r+1} = 3$

Gibbs / Carlin & Chib



Carlin & Chib



Gibbs

Illustration échantillonnage

En régime stationnaire

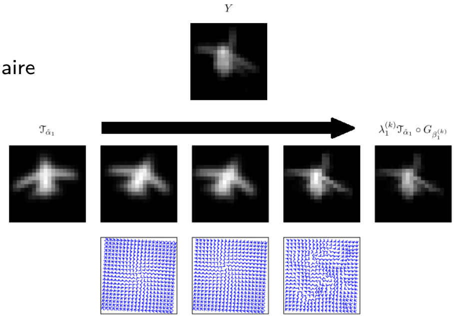
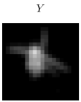


Illustration échantillonnage

En régime stationnaire



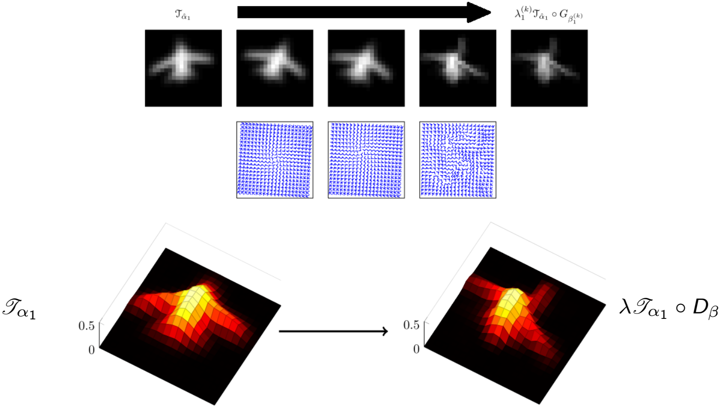
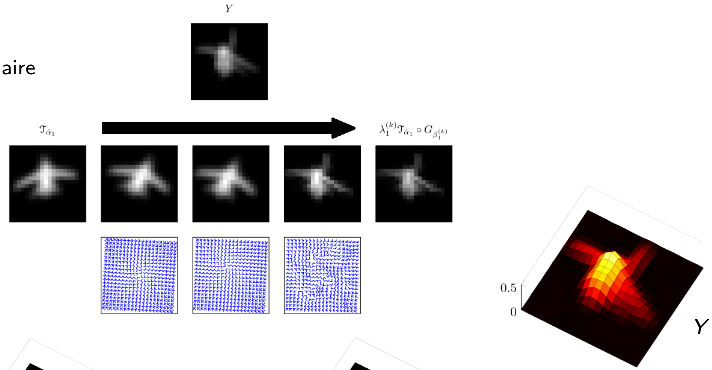


Illustration échantillonnage

En régime stationnaire



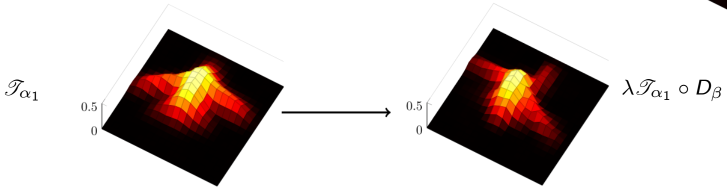
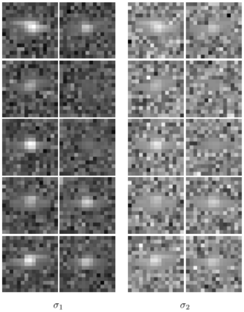
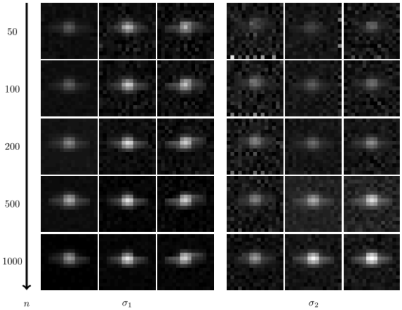


Illustration apprentissage (monospectral & semi-supervisé)

Estimation des prototypes pour deux niveaux de bruit et $C = 3$ classes



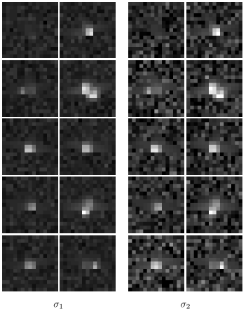
Exemples d'observations traitées



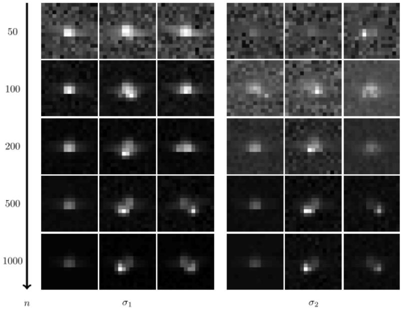
Evolution des estimateurs

Illustration apprentissage (monospectral & semi-supervisé)

Estimation des prototypes pour deux niveaux de bruit et $C = 3$ classes



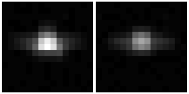
Exemples d'observations traitées



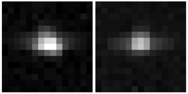
Evolution des estimateurs

Illustration apprentissage (non-supervisé)

En monospectral :
($C = 2$)



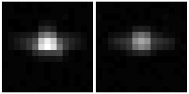
σ_1



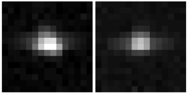
σ_2

Illustration apprentissage (non-supervisé)

En monospectral :
$(C = 2)$



σ_1



σ_2

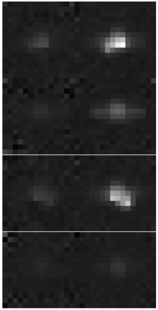
En multispectral :
$(C = 4)$

$\{\mathcal{T}_{\hat{\alpha}_{k,1}}\}_{k=1}^{2}$

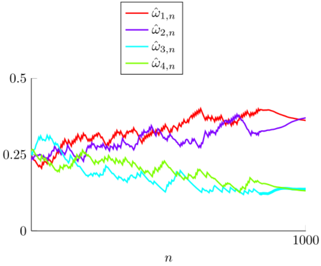
$\{\mathcal{T}_{\hat{\alpha}_{k,2}}\}_{k=1}^{2}$

$\{\mathcal{T}_{\hat{\alpha}_{k,3}}\}_{k=1}^{2}$

$\{\mathcal{T}_{\hat{\alpha}_{k,4}}\}_{k=1}^{2}$



σ_2



The mixture Model problem

Rewriting the Gibbs and the Carlin and Chib samplers:

$$P_{\mathtt{G}} \begin{cases} I_n^{(\mathtt{G})} \sim \pi(\,\cdot\,|\,x) \\ X_n^{(\mathtt{G})} \sim \delta_x(\,\cdot\,) \end{cases} \qquad Q_{\mathtt{G}} \begin{cases} I_{n+1}^{(\mathtt{G})} \sim \delta_{i'}(\,\cdot\,) \\ X_{n+1}^{(\mathtt{G})} \sim \pi(\,\cdot\,|\,i') \end{cases}$$

$$Z_{2n+1}^{(\mathtt{G})} = \begin{pmatrix} I_n^{(\mathtt{G})} = i' \\ X_n^{(\mathtt{G})} = x \end{pmatrix}$$

$$Z_{2n}^{(i)} = \begin{pmatrix} I_n^{(i)} = i \\ X_n^{(i)} = x \end{pmatrix} \qquad Z_{2n+2}^{(i)} = \begin{pmatrix} I_{k+1}^{(i)} = i' \\ X_{k+1}^{(i)} = x' \end{pmatrix}$$

$$Z_{2n+1}^{(\mathtt{CC})} = \begin{pmatrix} I_n^{(\mathtt{CC})} = i' \\ X_n^{(\mathtt{CC})} = u_{i'} \end{pmatrix}$$

$$P_{\mathtt{CC}} \begin{cases} \forall\, i \neq j \;\; U_i \sim \zeta_i(\,\cdot\,) \\ U_j \sim \delta_x(\,\cdot\,) \\ I_n^{(\mathtt{CC})} \sim \pi(\,\cdot\,|\,u_1, \dots, u_C) \end{cases} \qquad Q_{\mathtt{CC}} \begin{cases} I_{n+1}^{(\mathtt{CC})} \sim \delta_{i'}(\,\cdot\,) \\ X_{n+1}^{(\mathtt{CC})} \sim \pi(\,\cdot\,|\,i') \end{cases}$$

(i) $P_{\mathtt{CC}}$ is π-reversible, (ii) $Q_{\mathtt{CC}} = Q_{\mathtt{G}}$, (iii) $P_{\mathtt{CC}} \overset{?}{\succeq} P_{\mathtt{G}} \Rightarrow v(f, \mathtt{CC}) \leq v(f, \mathtt{G})$.

Pseudo-Marginal Algorithms

Pseudo-Marginal (Andrieu & Robert, 2009): no exact expression of the target distribution π *e.g*

$$\pi(x) = \int \pi(x, \mathrm{d}u)$$

for all $(x, x') \in \mathsf{X}^2$, $\pi(x)/\pi(x')$ intractable

$\Rightarrow$ Idea: simulate a Markov chain targeting

$$\tilde{\pi}(\mathrm{d}x, \mathrm{d}u) = \underbrace{\pi(\mathrm{d}x) w_u(x)}_{\hat{\pi}_u(\mathrm{d}x),\text{calculable}} \underbrace{R(x, \mathrm{d}u)}_{\text{samplable}}$$

(note that $\int \tilde{\pi}(x, \mathrm{d}u) = \pi(x)$)
For example, use Importance Sampling estimate:

$$\hat{\pi}_u(x) = \frac{1}{n} \sum_{k=1}^{n} \frac{\pi(x, u^{(k)})}{R(x, u^{(k)})}, \quad U^{(k)} \overset{i.i.d}{\sim} R(x, \cdot)\ .$$

Monte Carlo within Metropolis (MCWM)

A Markov chain $\{X_n,\ n \in \mathbb{N}\}$ on $(\mathsf{X}, \mathcal{X})$: given $X_n = x$, X_{n+1} is obtained as follows

(i) propose $X' \sim K(x, \cdot\,)$

(ii) simulate aux. var. for both states X_n and X':
$U \sim R(x, \cdot\,), U' \sim R(x', \cdot\,)$

(iii) accept $X_{n+1} = x'$ w.p

$$\hat{\alpha}(x, x', u, u') = 1 \wedge \frac{\hat{\pi}_{u'}(x')K(x', x)}{\hat{\pi}_u(x)K(x, x')}$$

MCWM is not π-reversible but targets an approximate of $\pi(x)$...
$\Rightarrow$ noisy algorithm!

Grouped-Independence Metropolis Hastings (GIMH)

A Markov chain $\{(X_n, U_n),\ n \in \mathbb{N}\}$ targeting $\tilde{\pi}$ such that given $(X_n, U_n) = (x, u)$, (X_{n+1}, U_{n+1}) is obtained as follows

(i) propose $X' \sim K(x, \cdot\,)$

(ii) simulate aux. var. for the state X': $U' \sim R(x', \cdot\,)$

(iii) accept $(X_{n+1}, U_{n+1}) = (x', u')$ w.p

$$\hat{\alpha}((x, u), (x', u')) = 1 \wedge \frac{\hat{\pi}_{u'}(x')K(x', x)}{\hat{\pi}_u(x)K(x, x')}$$

GIMH is **Metropolis-Hastings** algorithm $\Rightarrow$ $\tilde{\pi}$-reversible.

Remark

- MCWM & GIMH cannot be properly compared with Tierney's Theorem
- They may be rewritten artificially as:

$$\text{MCWM:} \quad \left(X_n^{(M)} \right) \xrightarrow{P_M} \begin{pmatrix} X_n^{(M)} \\ U \end{pmatrix} \xrightarrow{Q} \left(X_{n+1}^{(M)} \right) \xrightarrow{P_M} \cdots$$

$$\text{GIMH:} \quad \begin{pmatrix} X_n^{(G)} \\ U_n^{(G)} \end{pmatrix} \xrightarrow{P_G} \begin{pmatrix} X_n^{(G)} \\ U_n^{(G)} \end{pmatrix} \xrightarrow{Q} \begin{pmatrix} X_{n+1}^{(G)} \\ U_{n+1}^{(G)} \end{pmatrix} \xrightarrow{P_G} \cdots$$

A Random-Refreshment Pseudo Marginal algorithm

A Markov chain $\{(X_n, U_n),\ n \in \mathbb{N}\}$ targeting $\tilde{\pi}$ such that given $(X_n, U_n) = (x, u)$, (X_{n+1}, U_{n+1}) is obtained as follows:

(i) (a) propose a new aux. var. for state X; $\tilde{U} \sim R(x, \cdot\,)$
(b) refresh the aux. var. U_n by $\tilde{U}$ with a certain probability $\omega_{u,\tilde{u}}$

(ii) propose $X' \sim K(x, \cdot\,)$

(iii) simulate aux. var. for the state X': $U' \sim R(x', \cdot\,)$

(iv) accept $(X_{n+1}, U_{n+1}) = (x', u')$ w.p

$$\hat{\alpha}((x, u), (x', u')) = 1 \wedge \frac{\hat{\pi}_{u'}(x')K(x', x)}{\hat{\pi}_{u}(x)K(x, x')}$$

Comparing GIMH & Random Refreshment

GIMH: $$\begin{pmatrix} X_n^{(G)} \\ U_n^{(G)} \end{pmatrix} \xrightarrow{P_G} \begin{pmatrix} X_n^{(G)} \\ U_n^{(G)} \end{pmatrix} \xrightarrow{Q} \begin{pmatrix} X_{n+1}^{(G)} \\ U_{n+1}^{(G)} \end{pmatrix} \xrightarrow{P_G} \cdots$$

Random Refreshment: $$\begin{pmatrix} X_n^{(R)} \\ U_n^{(R)} \end{pmatrix} \xrightarrow{P_R} \begin{pmatrix} X_n^{(R)} \\ \tilde{U} \end{pmatrix} \xrightarrow{Q} \begin{pmatrix} X_{n+1}^{(R)} \\ U_{n+1}^{(R)} \end{pmatrix} \xrightarrow{P_G} \cdots$$

Our Theorem holds and show that for any $f \in \mathsf{F}$

$$v(f, R) \leq v(f, G) \; .$$