

Décompositions parcimonieuses : approches Bayésiennes et application à la compression d'image

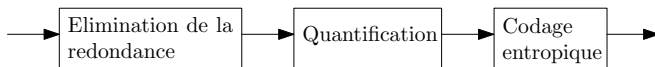
Angélique Drémeau

Co-direction : Jean-Jacques Fuchs, Christine Guillemot

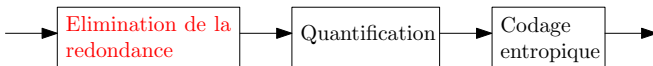
INRIA Rennes - Bretagne Atlantique

19 novembre 2010

Compression d'image



Compression d'image

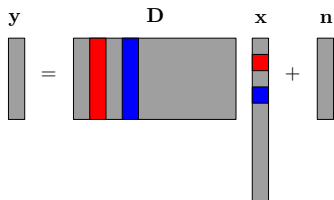


Transformation



Prédiction

Décompositions parcimonieuses



- ▶ **Problème** On cherche \mathbf{x} parcimonieux.
- ▶ **Plusieurs formulations**

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 \text{ soumis à } \|\mathbf{x}\|_0 \leq L,$$

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_0,$$

λ multiplicateur Lagrangien, $\|\cdot\|_0$ pseudo-norme ℓ_0 .

Axes de recherche

Prédiction et parcimonie

Nouvel algorithme basé sur un mélange de décompositions parcimonieuses

Transformations adaptatives

Nouvel algorithme d'apprentissage de bases

Algorithmes Bayésiens

Nouveaux algorithmes basés sur un modèle Bernoulli-Gaussien

Axes de recherche

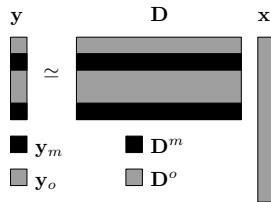
Prédiction et parcimonie

Nouvel algorithme basé sur un mélange de décompositions parcimonieuses

Transformations adaptatives

Algorithmes Bayésiens

Prédiction basée sur des décompositions parcimonieuses

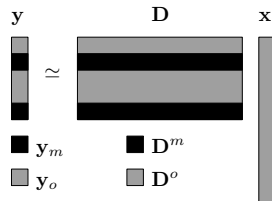
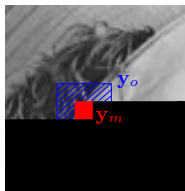


Le vecteur \mathbf{y} a une décomposition parcimonieuse dans un dictionnaire donné.

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y}_o - \mathbf{D}^o \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_0,$$

$$\mathbf{y}_m^* = \mathbf{D}^m \mathbf{x}^*.$$

Prédiction basée sur des décompositions parcimonieuses



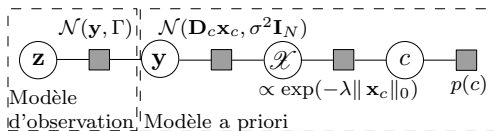
Le vecteur \mathbf{y} a une décomposition parcimonieuse dans un dictionnaire donné.

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y}_o - \mathbf{D}^o \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_0,$$

$$\mathbf{y}_m^* = \mathbf{D}^m \mathbf{x}^*.$$

Contexte probabiliste

Modèle considéré



Problème d'estimation MMSE approché

$$\mathbf{y}_m^* = \underset{\hat{\mathbf{y}}_m}{\operatorname{argmin}} E_{\mathbf{y}_m | \mathcal{X}^*, \mathbf{z}} [\|\hat{\mathbf{y}}_m - \mathbf{y}_m\|_2^2]$$

Description de l'algorithme

1. Calcul des décompositions parcimonieuses dans chaque dictionnaire

$$\mathbf{x}_c^* = \underset{\mathbf{x}_c}{\operatorname{argmin}} \frac{1}{2(\sigma_o^2 + \sigma^2)} \|\mathbf{z}_o - \mathbf{D}_c^o \mathbf{x}_c\|_2^2 + \lambda \|\mathbf{x}_c\|_0,$$

2. Estimation du signal à prédire

$$\hat{\mathbf{y}}_m^* = \sum_c p(c | \mathcal{X}^*, \mathbf{z}) \mathbf{D}_c^m \mathbf{x}_c^*.$$

Résultats en prédiction



Méthode standard

$$L_{max} = 56$$

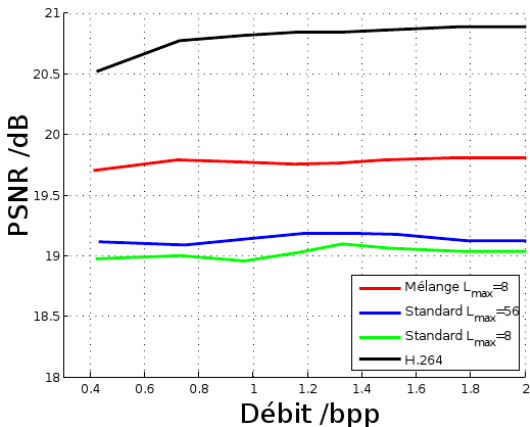


Méthode proposée

$$L_{max} = 8$$

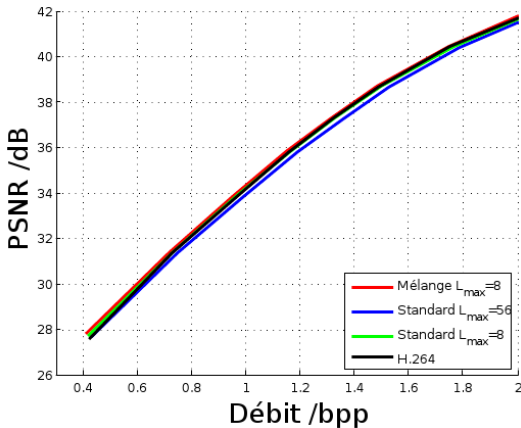
Performances débit-distorsion de prédiction : image Cameraman

Minimisation de l'erreur de prédiction



Performances débit-distorsion finales : image Cameraman

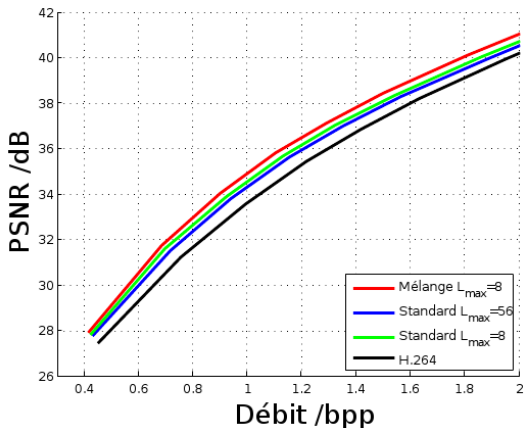
Minimisation de l'erreur de prédiction



	dB	%
Mél. $L_{max}=8$	0,15	-1,60
Std. $L_{max}=56$	-0,21	2,08
Std. $L_{max}=8$	-0,01	-0,03

Performances débit-distorsion finales : image Barbara

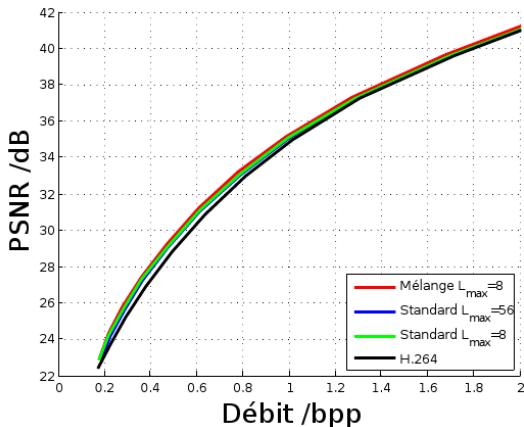
Minimisation de l'erreur de prédiction



	dB	%
Mél. $L_{max}=8$	1,05	-9,93
Std. $L_{max}=56$	0,54	-5,13
Std. $L_{max}=8$	0,75	-7,14

Performances débit-distorsion finales : image Barbara

Optimisation débit-distorsion



	dB	%
Mél. $L_{max}=8$	0,53	-5,77
Std. $L_{max}=56$	0,26	-2,84
Std. $L_{max}=8$	0,37	-4,06

Axes de recherche

Prédiction et parcimonie

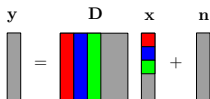
Transformations adaptatives

Nouvel algorithme d'apprentissage de bases

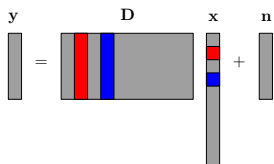
Algorithmes Bayésiens

Base orthogonale vs dictionnaire redondant

Transformation linéaire



Décomposition parcimonieuse



A nombre de coefficients non nuls fixés,

- ▶ la **distorsion** diminue avec la redondance du dictionnaire,
- ▶ le **débit** augmente avec la redondance du dictionnaire.

Structuration du dictionnaire

Ensemble de bases orthonormées

$$y = \begin{bmatrix} D_1 & D_2 \end{bmatrix} x + n$$

$$y = \begin{bmatrix} D_1 & D_2 \end{bmatrix} x + n$$

$$y = \begin{bmatrix} D_1 & D_2 \end{bmatrix} x + n$$

Structuration du dictionnaire

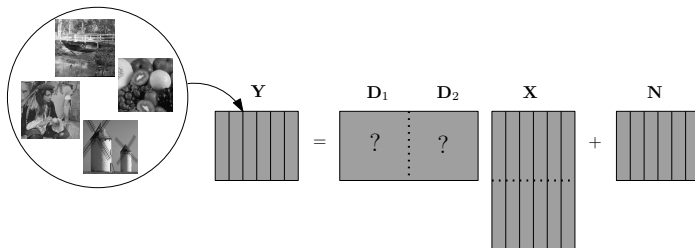
Ensemble de bases orthonormées

$$y = \begin{bmatrix} D_1 & D_2 \end{bmatrix} x + n$$

$$y = \begin{bmatrix} D_1 & D_2 \end{bmatrix} x + n$$

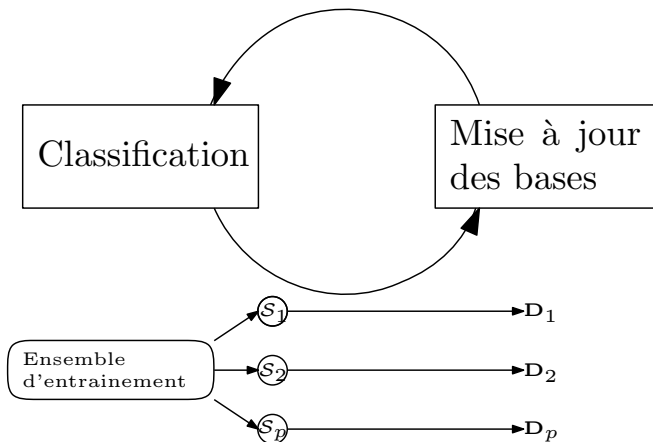
~~$$y = \begin{bmatrix} D_1 & D_2 \end{bmatrix} x + n$$~~

Apprentissage de bases favorisant la parcimonie des décompositions



Algorithme de Sezer

[Sezer2008]



Algorithme de Sezer : formulation

K signaux d'entraînement (indiqués par j),
 P bases orthonormées (indiquées par i).

1. Classification

$$\mathcal{S}_i = \{j \in \{1, \dots, K\} \mid c_j = i\},$$

$$\text{où } c_j = \underset{i \in \{1, \dots, P\}}{\operatorname{argmin}} \left\{ \|\mathbf{y}_j - \mathbf{D}_i \mathbf{x}_{ji}\|_2^2 + \lambda' \|\mathbf{x}_{ji}\|_0 \right\}.$$

2. Mise à jour des bases et des décompositions parcimonieuses

$$\mathbf{x}_{ji} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \|\mathbf{y}_j - \mathbf{D}_i \mathbf{x}\|_2^2 + \lambda' \|\mathbf{x}\|_0 \right\},$$

$$\mathbf{D}_i = \underset{\mathbf{D}}{\operatorname{argmin}} \left\{ \sum_{j \in \mathcal{S}_i} \|\mathbf{y}_j - \mathbf{D} \mathbf{x}_{ji}\|_2^2 \right\} \text{ soumis à } \mathbf{D}^T \mathbf{D} = \mathbf{I}_N.$$

Contexte probabiliste

Modèle considéré

$$\begin{aligned} p(\mathbf{y}_j | \mathbf{x}_j, \mathbf{D}, c_j = i) &= \mathcal{N}(\mathbf{D}_i \mathbf{x}_{ji}, \sigma^2 \mathbf{I}_N), \\ p(\mathbf{x}_j | c_j = i) &\propto \exp\{-\lambda \|\mathbf{x}_{ji}\|_0\}. \end{aligned}$$

Algorithme de Sezer : problème d'estimation **MAP**

$$(\mathbf{D}^*, \mathbf{X}^*, \mathbf{c}^*) = \underset{(\mathbf{D}, \mathbf{X}, \mathbf{c})}{\operatorname{argmax}} \sum_{j=1}^K \log p(\mathbf{y}_j, \mathbf{x}_j, \mathbf{D}, c_j).$$

Approche alternative

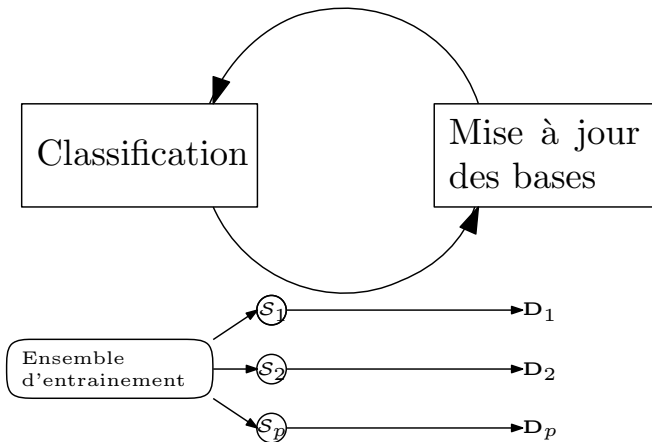
Problème d'estimation **MAP marginalisé**

$$(\mathbf{D}^*, \mathbf{X}^*) = \operatorname{argmax}_{(\mathbf{D}, \mathbf{X})} \sum_{j=1}^K \log p(\mathbf{y}_j, \mathbf{x}_j, \mathbf{D}),$$

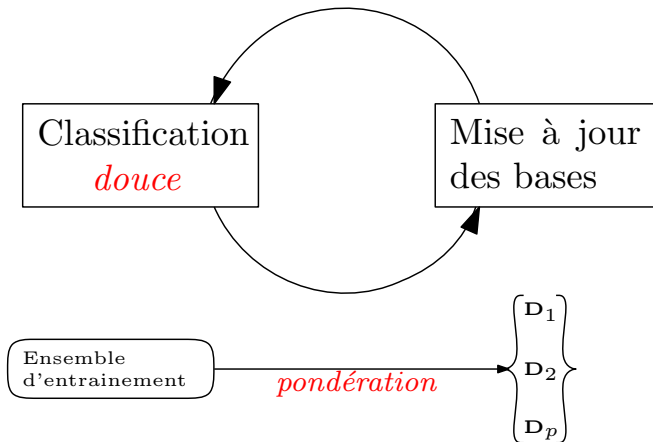
$$\text{où } p(\mathbf{y}_j, \mathbf{x}_j, \mathbf{D}) = \sum_{c_j=1}^P p(\mathbf{y}_j, \mathbf{x}_j, \mathbf{D}, c_j).$$

↪ Résolu par un algorithme EM.

Description de l'algorithme



Description de l'algorithme



Description de l'algorithme : formulation

K signaux d'entraînement (indiqués par j),
 P bases orthonormées (indiqués par i).

1. E-step

$$w_{ji} \propto \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y}_j - \mathbf{D}_i \mathbf{x}_{ji}\|_2^2 - \lambda \|\mathbf{x}_{ji}\|_0\right) p(c_j).$$

2. M-step

$$\mathbf{x}_{ji} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ \|\mathbf{y}_j - \mathbf{D}_i \mathbf{x}\|_2^2 + \lambda' \|\mathbf{x}\|_0 \right\},$$

$$\mathbf{D}_i = \underset{\mathbf{D}}{\operatorname{argmin}} \left\{ \sum_{j=1}^K w_{ji} \|\mathbf{y}_j - \mathbf{D} \mathbf{x}_{ji}\|_2^2 \right\} \quad \text{soumis à } \mathbf{D}^T \mathbf{D} = \mathbf{I}_N,$$

Estimation de la variance de bruit

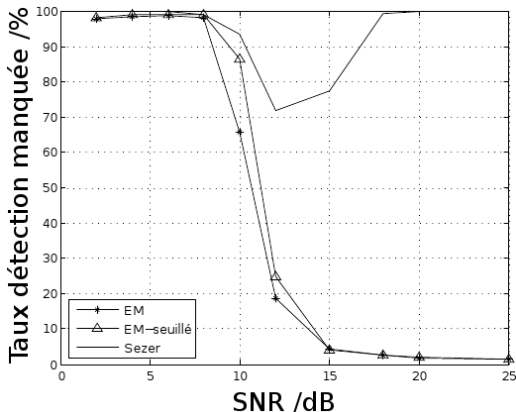
Variance de bruit comme variable inconnue du problème MAP

$$(\mathbf{D}^*, \mathbf{X}^*, (\sigma^2)^*) = \underset{(\mathbf{D}, \mathbf{X}, \sigma^2)}{\operatorname{argmax}} \sum_{j=1}^K \log p(\mathbf{y}_j, \mathbf{x}_j, \mathbf{D}).$$

Ajout de l'estimation de la variance de bruit dans le M-step

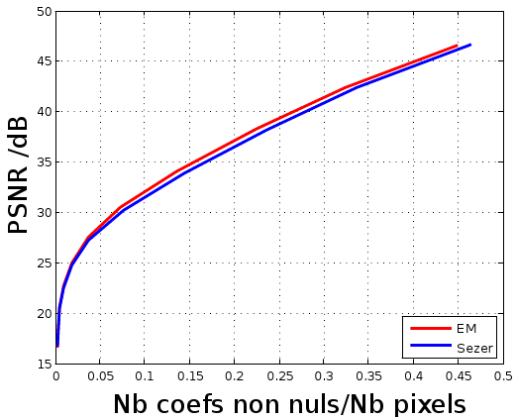
$$\sigma^2 = \frac{1}{NK} \sum_{j=1}^K \sum_{i=1}^P w_{ji} \|\mathbf{y}_j - \mathbf{D}_i \mathbf{x}_{ji}\|_2^2.$$

Performances en reconstruction



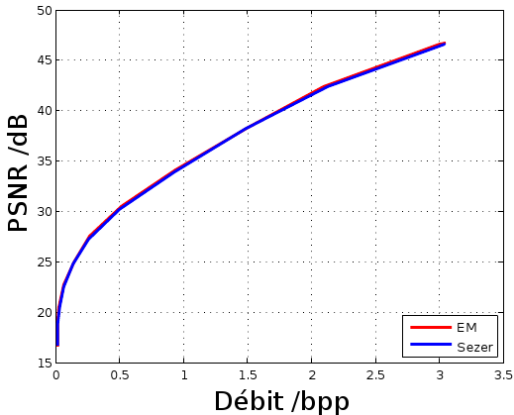
- ▶ 500 signaux générés,
- ▶ de dimension 16,
- ▶ avec 4 coefficients non nuls,
- ▶ $\text{SNR} = 10 \log \frac{\sigma_x^2}{\sigma^2}$

Performances en compression : parcimonie-distorsion



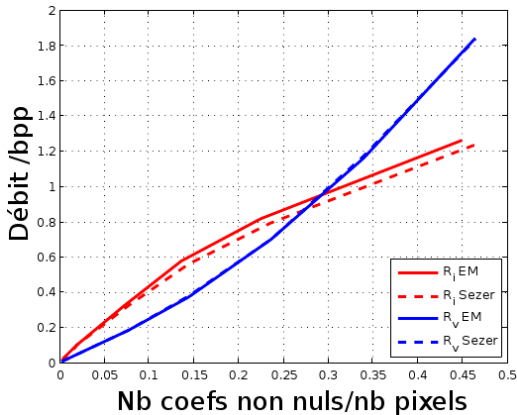
- ▶ Image Cameraman,
- ▶ Blocs 16×16 pixels.

Performances en compression : débit-distorsion



- ▶ Image Cameraman,
- ▶ Blocs 16×16 pixels.

Comparaison des coûts de codage



- ▶ Image Cameraman,
- ▶ Blocs 16×16 pixels.

Axes de recherche

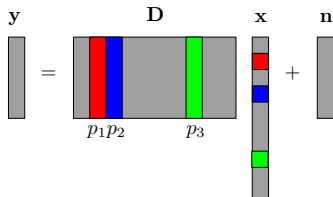
Prédiction et parcimonie

Transformations adaptatives

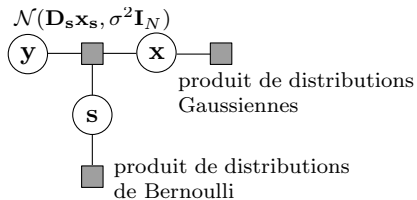
Algorithmes Bayésiens

Nouveaux algorithmes basés sur un modèle Bernoulli-Gaussien

Contexte probabiliste



Modèle considéré



Recherche de décompositions parcimonieuses

Problème d'estimation MAP

$$(\hat{\mathbf{x}}, \hat{\mathbf{s}}) = \underset{\mathbf{x}, \mathbf{s}}{\operatorname{argmax}} \log p(\mathbf{y}, \mathbf{x}, \mathbf{s}).$$

On montre que, pour certains paramètres, $\mathbf{x}^* = \hat{\mathbf{x}}$ avec

$$\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_0.$$

Exemple : algorithme StOMP

[Donoho2006]

1. Mise à jour du support

$$\hat{s}_i = \begin{cases} 1 & \text{si } \langle \mathbf{r}, \mathbf{d}_i \rangle^2 > T, \\ \hat{s}_i & \text{sinon.} \end{cases}$$

2. Calcul des coefficients du vecteur \mathbf{x} correspondant

$$\hat{\mathbf{x}}_{\hat{s}} = (\mathbf{D}_{\hat{s}}^T \mathbf{D}_{\hat{s}})^{-1} \mathbf{D}_{\hat{s}}^T \mathbf{y}.$$

Bayesian StOMP

1. Redéfinition du support

$$\hat{s}_i = \begin{cases} 1 & \text{si } \langle \mathbf{r} + \hat{\mathbf{x}}_i \mathbf{d}_i, \mathbf{d}_i \rangle^2 > T_i, \\ 0 & \text{sinon,} \end{cases}$$

avec
$$T_i = 2 \sigma^2 \frac{\sigma^2 + \sigma_x^2}{\sigma_x^2} \log \frac{1 - p_i}{p_i}.$$

2. Calcul des coefficients du vecteur \mathbf{x} correspondant

$$\hat{\mathbf{x}}_{\hat{s}} = \left(\mathbf{D}_{\hat{s}}^T \mathbf{D}_{\hat{s}} + \frac{\sigma^2}{\sigma_x^2} \mathbf{I}_{\|\hat{s}\|_0} \right)^{-1} \mathbf{D}_{\hat{s}}^T \mathbf{y}.$$

Estimation des paramètres

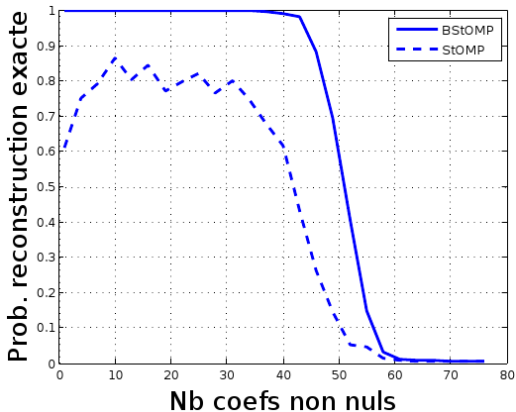
Variance de bruit

$$\begin{aligned}\sigma^2 &= \underset{\sigma^2}{\operatorname{argmax}} \log p(\mathbf{y}, \hat{\mathbf{x}}, \hat{\mathbf{s}}), \\ &= \frac{\|\mathbf{r}\|_2^2}{N}.\end{aligned}$$

Seuil dépend alors de l'itération

$$T_i = 2 \frac{\|\mathbf{r}\|_2^2}{N} \frac{\sigma_x^2 + N^{-1} \|\mathbf{r}\|_2^2}{\sigma_x^2} \log \frac{1 - p_i}{p_i}.$$

Evaluation des performances



- ▶ 500 signaux générés,
- ▶ de dimension 128,
- ▶ Dictionnaire de 256 atomes.

Conclusions

Codage par prédiction : nouvel algorithme basé sur un mélange de décompositions parcimonieuses,

- ▶ Résultats encourageants.

Codage par transformation : nouvel algorithme d'apprentissage de bases basé sur l'algorithme EM,

- ▶ Bonnes performances en reconstruction vis à vis de l'algorithme de Sezer,
- ▶ Amélioration de la parcimonie des décompositions,
- ▶ Augmentation du coût de codage de la combinaison des atomes utilisés.

Recherche de décompositions parcimonieuses : nouveaux algorithmes basés sur un modèle Bernoulli-Gaussien,

- ▶ Bon comportement général des algorithmes Bayésiens par rapport à leurs homologues standard.

Perspectives (1/2)

Prédiction basée sur un mélange de décompositions parcimonieuses

- ▶ Intégration de l'algorithme dans un codeur vidéo,
- ▶ Application au débruitage d'images, en association avec un apprentissage de bases.

Perspectives (2/2)

Algorithmes de **recherche** de décompositions parcimonieuses Bayésiens

- ▶ Evaluation de la capacité des algorithmes à retrouver le support d'une décomposition parcimonieuse,
- ▶ Prise en compte d'une certaine hiérarchisation des atomes (paramètres de Bernoulli différents, suivant une certaine loi, par des probabilités conditionnelles...).

Apprentissage de bases orthonormées encourageant la parcimonie des décompositions

- ▶ Optimisation du nombre de bases,
- ▶ Conception d'un algorithme d'apprentissage de dictionnaires basé sur un modèle Bernoulli-Gaussien.