

Ecole Internationale des Sciences du Traitement de l'Information

Ingénieurs 2<sup>ème</sup> année : Mathématiques et Informatique

## *Compressive Sensing*

### **Cours 2 : Reconstruction d'une solution parcimonieuse**

Nisrine Fortin

2015-2016



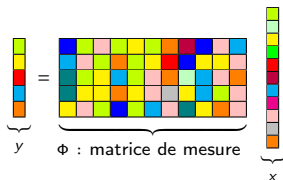
- 1 Matrices de mesure
  - Condition d'unicité
  - Nombre minimal de mesures
  - Restricted Isometry Property
  - Cohérence
  - Cohérence mutuelle
  - Exemples de matrices de mesure
  
- 2 Algorithmes de reconstruction
  - Relaxation convexe
  - Poursuite gloutonne
  
- 3 Références

# Matrices de mesure

## Matrices de mesure

On représente le processus de mesure (acquisition) par la matrice  $\Phi$  :

$$y = \Phi x$$



- La matrice de mesure intervient dans la phase d'acquisition du Compressive Sensing mais elle doit être conçue pour **assurer la reconstruction du signal parcimonieux**.
- La matrice de mesure doit donc répondre à des critères d'existence et d'unicité du signal reconstruit et doit remplir certaines conditions.

# Matrices de mesure

## 1- Unicité de la solution parcimonieuse



## Condition d'unicité

**Proposition 1.** Soient  $\alpha \in \mathbb{R}^N$  et  $s = \|\alpha\|_0$ . On a l'équivalence entre les assertions suivantes :

- 1  $\alpha$  est l'unique solution du problème de minimisation  $\ell_0$ .
- 2  $(\alpha + \ker(A)) \cap \Sigma_s = \{\alpha\}$ .

## Condition d'unicité : Commentaires

- Pour pouvoir reconstruire le signal de façon unique, l'espace  $\ker(A)$  doit être éloigné de l'ensemble  $\Sigma_s$ .
- Géométriquement, on souhaite que  $\ker A$  intersecte  $\Sigma_s$  seulement en 0 et qu'il soit proche des éléments les moins sparses de  $\mathbb{R}^N$ .

# Matrices de mesure

## 2- Nombre minimal de mesures

## Nombre minimal de mesures

- La matrice de mesure intervient dans un système hautement sous-déterminé : le nombre de lignes  $M$  est très inférieur au nombre de colonnes  $N$ .
- $N$  étant la taille du signal (fixée), l'utilisateur intervient dans la phase d'acquisition avec un nombre de mesures  $M$  qui intervient directement dans la recherche de la solution  $s$ -sparse.

*Quel est le nombre minimal de mesures  $M$  nécessaires pour reconstruire tout vecteur de  $s$ -sparse à partir de  $M$  mesures ?*

## Nombre minimal de mesures

**Proposition 2.** Les assertions suivantes sont équivalentes :

- 1  $\forall x \in \Sigma_s, \forall t \in \Sigma_s, x \neq t \Rightarrow \mathbf{A}x \neq \mathbf{A}t.$
- 2  $\ker(\mathbf{A}) \cap \Sigma_{2s} = \{0\}$  (donc  $M \geq 2s$ ).
- 3 Toutes les sous-matrices de  $\mathbf{A}$  d'ordre  $M \times 2s$  sont injectives (donc  $M \geq 2s$ ).

## Unité et nombre minimal de mesures

Une conclusion des deux propositions nous permettra d'émettre une hypothèse pour la reconstruction exacte à partir du problème de minimisation ( $\ell_0$ ) de la solution  $s$ -sparse avec un nombre minimal de mesures.

**Proposition 3.** Les assertions suivantes sont équivalentes :

- ①  $\ker(\mathbf{A}) \cap \Sigma_{2s} = \{0\}$ .
- ② La procédure de minimisation  $\ell_0$  permet de reconstruire exactement tout vecteur de  $\Sigma_s$ .

## Commentaires et conclusion

- *2s est le nombre nécessaire et suffisant pour pouvoir reconstruire exactement tout vecteur s-sparse.*
- *La procédure de minimisation ( $\ell_0$ ) fournit dans ce cas une procédure de reconstruction exacte quand la matrice de mesure  $\mathbf{A}$  vérifie la condition*

$$\ker(\mathbf{A}) \cap \Sigma_{2s} = \{0\}$$

- *De telles matrices existent par exemple : la matrice de Vandermonde.*
- **Inconvénient** : *La procédure de minimisation ( $\ell_0$ ) est optimale d'un point de vue théorique mais elle ne peut pas être utilisée en pratique car c'est une procédure NP-difficile (combinatoire).*

# Matrices de mesure

## 3- Restricted Isometry Property

## Restricted Isometry Property

Une matrice  $A$  satisfait le restricted isometry property RIP d'ordre  $s$  si il existe une constante  $0 < \delta_s < 1$  telle que

$$\forall \alpha \in \Sigma_s, \quad (1 - \delta_s) \|\alpha\|_2^2 \leq \|A\alpha\|_2^2 \leq (1 + \delta_s) \|\alpha\|_2^2$$

→ La plus petite constante vérifiant le RIP d'ordre  $s$  est appelée Restricted Isometry Constant (RIC).

*Le RIP permet de vérifier si la matrice de mesure  $A$  est proche d'une isométrie, c'est-à-dire si elle préserve la distance entre deux vecteurs de mesure.*

## Restricted Isometry Property

- Si la matrice de mesure satisfait le RIP d'ordre  $2s$ , alors la distance entre deux vecteurs de mesure  $y_1 = A\alpha_1$  et  $y_2 = A\alpha_2$  est proportionnelle à la distance entre  $\alpha_1$  et  $\alpha_2$ .
- Le RIP est une propriété importante qui permet de garantir la reconstruction du signal.

# Matrices de mesure

## 3- Cohérence

# Cohérence

La cohérence  $\mu(A)$  d'une matrice  $A$  est égale à la plus grande valeur absolue du produit scalaire entre deux vecteurs colonnes distincts de  $A$  :

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$$

où  $a_i$  et  $a_j$  sont les colonnes de la matrice  $\mathbf{A}$ .

*La cohérence permet de mesurer la corrélation maximale entre les différents vecteurs colonnes d'une matrice.*

## Cohérence

- Lorsque  $M \ll N$ ,  $\frac{1}{\sqrt{M}} \leq \mu(\mathbf{A}) \leq 1$ .
- Une matrice  $\mathbf{A}$  est dite **incohérente** lorsque la valeur de  $\mu(\mathbf{A})$  tend vers la limite de *Welch*  $\frac{1}{\sqrt{M}}$ .

## Unicité et la cohérence

- Si la relation suivante est satisfait

$$s < \frac{1}{2} \left( 1 + \frac{1}{\mu(\mathbf{A})} \right)$$

alors pour tout vecteur de mesure  $y \in \mathbb{R}^M$ , il existe au plus un signal  $\alpha \in \mathbb{R}^N$  tel que  $y = \mathbf{A}\alpha$

- La matrice de mesure  $\mathbf{A}$  doit avoir une faible cohérence : une faible valeur de la cohérence  $\mu$  augmente la borne supérieure de  $s$ .
- Pour minimiser la valeur de  $\mu$ , il faut que les vecteurs colonnes de  $\mathbf{A}$  soient les plus orthogonaux possibles.

# Matrices de mesure

## 4- Cohérence mutuelle

## Cohérence mutuelle

La notion de cohérence peut être étendue pour une paire de bases. La cohérence mutuelle entre  $\Phi$  et  $\Psi$  est définie comme suit

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{i \neq j} \frac{|\langle \Phi_i, \Psi_j \rangle|}{\|\Phi_i\|_2 \|\Psi_j\|_2}$$

où  $\Phi_i$  et  $\Psi_j$  représentent respectivement les vecteurs lignes de  $\Phi$  et les vecteurs colonnes de  $\Psi$ .

*La cohérence mutuelle mesure la corrélation maximale entre les vecteurs lignes de  $\Phi$  et les vecteurs colonnes de  $\Psi$ .*

La plage de valeurs de la cohérence mutuelle est

$$1 \leq \mu(\Phi, \Psi) \leq \sqrt{N}.$$

# Matrices de mesure

## 4- Exemples de matrices de mesure

## Exemples de matrices de mesure

- Les matrices de mesure construites aléatoirement : les conditions vues précédemment peuvent être remplies avec une forte probabilité.
- En particulier, lorsque les éléments de la matrice de mesure  $\mathbf{A}$  de type  $M \times N$  sont générés à partir :
  - ① d'un processus gaussien identique et Indépendamment distribué (i.i.d) avec une moyenne nulle et une variance  $\frac{1}{M} : \mathcal{N}\left(0, \frac{1}{M}\right)$ ;
  - ② d'un processus aléatoire équiprobable prenant des valeurs  $\pm \frac{1}{\sqrt{M}}$ .

# Algorithmes de reconstruction

## Algorithmes de reconstruction

- En tenant compte de l'hypothèse que  $\alpha$  est parcimonieux dans le domaine  $\Psi$ , l'étape 1 consiste à résoudre le problème de minimisation suivant qu'on appellera la minimisation ( $\ell_0$ ) :

$$\min_{\tilde{\alpha}} \|\alpha\|_0 \text{ sous la contrainte } \mathbf{A}\tilde{\alpha} = y \quad (1)$$

- En tenant compte du bruit additif et de la parcimonie, l'équation (1) devient :

$$\min_{\tilde{\alpha}} \|\alpha\|_0 \text{ sous la contrainte } \|y - \mathbf{A}\tilde{\alpha}\|_2 \leq \|\epsilon\|_2 \quad (2)$$

## Algorithmes de reconstruction

- La résolution des équations (1) et (2) nécessite une recherche exhaustive de la solution la plus parcimonieuse  $\tilde{\alpha}$  et est très complexe à mettre en œuvre.
- Plusieurs méthodes ont été proposées dans la littérature pour contourner ce problème. Principalement, elles peuvent être catégorisées en trois groupes :
  - 1 Relaxation convexe, par exemple la poursuite de base "Basis Pursuit (BP)" et la méthode LASSO.
  - 2 Poursuite gloutonne, par exemple la poursuite adaptative ou "Matching Pursuit (MP)" et son extension appelée poursuite adaptative orthogonale ou OMP.
  - 3 Inférence bayésienne, comme le "sparse bayesian learning".

# Relaxation convexe

## 1- Méthode du Basis Pursuit

## Relaxation convexe : Basis Pursuit

- Pour contourner la complexité liée à la norme  $\ell_0$ , la relaxation convexe, Basis Pursuit, remplace la norme  $\ell_0$  par une norme  $\ell_1$  et utilise des solveurs convexes pour le résoudre :

$$\min_{\tilde{\alpha}} \|\tilde{\alpha}\|_1 \quad \text{tel que} \quad \mathbf{A}\tilde{\alpha} = y \quad (3)$$

- L'approche "Poursuite de base" Basis Pursuit utilise des techniques de programmation linéaire pour résoudre le problème de reconstruction (3).
- En tenant compte du bruit additif, le problème formulé par l'équation (3) devient

$$\min_{\tilde{\alpha}} \|\tilde{\alpha}\|_1 \quad \text{tel que} \quad \|y - \mathbf{A}\tilde{\alpha}\|_2 \leq \epsilon \quad (4)$$

Cette méthode est appelée "Basis Pursuit with Inequality Constraints (BPIC)".

## Relaxation convexe : Basis Pursuit

- La procédure de Basis Pursuit est très populaire en Compressive Sensing.
- Le problème (3) peut être réécrit comme un problème de programmation linéaire : fonction objectif linéaire sous des contraintes linéaires.
- On considère le problème de programmation linéaire suivant :

$$(LP) \quad \begin{cases} \min_{(x_j^+, x_j^-)} \sum_{j=1}^N (x_j^+ + x_j^-) \\ \text{s.c. } [A| -A] \begin{bmatrix} x^+ \\ x^- \end{bmatrix} = y \\ \begin{bmatrix} x^+ \\ x^- \end{bmatrix} \geq 0 \end{cases}$$

**Notation :**

$$x_j^+ = \max(0, x_j), \quad x_j^- = \max(0, -x_j), \quad x_j = x_j^+ - x_j^- \quad \text{et} \quad |x_j| = x_j^+ + x_j^-.$$

## Relaxation convexe : Basis Pursuit

**Proposition 4 :** Les deux problèmes (BP) et (LP) sont équivalents.

- 1 Si  $\hat{\alpha}$  est une solution du problème (BP) alors  $(\hat{\alpha}^+, \hat{\alpha}^-)$  est une solution du problème (LP).
- 2 Si  $(\hat{\alpha}^+, \hat{\alpha}^-)$  est une solution du problème (LP) alors  $\hat{\alpha}$  est une solution du problème (BP).

## Relaxation convexe : Basis Pursuit et nombre de mesures

- Pour reconstruire tout vecteur  $\alpha$ ,  $s$ -parcimonieux à partir du vecteur de mesure  $y$  grâce au **Basis Pursuit**, on doit nécessairement avoir le nombre de mesures  $M$  qui vérifie

$$M \geq s \log \left( \frac{eN}{s} \right)$$

à une constante près.

- Lorsque  $\delta_{2s}(A) \leq \sqrt{2} - 1$  alors le problème (BP) admet une solution unique.
- D'un point de vue computationnel, le résultat obtenu par le BP est très important par rapport au résultat obtenu par la procédure de minimisation  $\ell_0$  nécessitant  $M \geq 2s$ .

# Relaxation convexe

## 2- Méthode LASSO

## Relaxation convexe : LASSO

- La méthode du Lasso (*Least Absolute Shrinkage and Selection Operator*) a été proposée par Tibshirani en 1996 dans une rétrospective publiée en 2011, Tibshirani cite la méthode dont il s'est inspiré pour construire le Lasso.
- Elle consiste à estimer le vecteur de paramètres de la solution du problème de reconstruction par
  - ① minimisation du critère quadratique des moindres carrés sous une contrainte sur la somme des valeurs absolues des coefficients

$$\min_{\tilde{\alpha}} \|y - \mathbf{A}\tilde{\alpha}\|_2 \quad \text{s.c.} \quad \|\tilde{\alpha}\|_1 \leq t \quad (\text{LASSO})$$

- ② ou de façon équivalente, par minimisation du critère quadratique pénalisé par la norme  $\ell_1$  des coefficients.

$$\min_{\tilde{\alpha}} \frac{1}{2} \|y - \mathbf{A}\tilde{\alpha}\|_2^2 + \lambda \|\tilde{\alpha}\|_1 \quad (\text{LASSO})$$

## Relaxation convexe : LASSO

- L'introduction d'une pénalisation réduit la variabilité de l'estimation, améliorant ainsi la précision de prédiction.
- La pénalisation de type  $\ell_1$  rétrécit certains coefficients, alors que les autres sont annulés exactement, aboutissant ainsi à des modèles parcimonieux.
- Le paramètre  $\lambda$  contrôle la régularisation : plus ce paramètre augmente plus les composantes  $\tilde{\alpha}_i$  de  $\tilde{\alpha}$  issu du LASSO tendent vers 0 et pour un  $\lambda$  suffisamment grand, certains sont exactement égaux à zéro.

## Relaxation convexe : LASSO

- Problème standard des moindres carrés :

$$\min_{x \in \mathbb{R}^N} \|y - \mathbf{A}x\|_2^2 \implies x^* = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T y$$

- Ridge regression :

$$\min_{x \in \mathbb{R}^N} \|y - \mathbf{A}x\|_2^2 + \lambda \|x\|_2 \implies x^* = (\mathbf{A}^T \mathbf{A} + I_N)^{-1} \mathbf{A}^T y$$

- Le problème LASSO :

$$\min_{\tilde{\alpha}} \frac{1}{2} \|y - \mathbf{A}x\|_2^2 + \lambda \|x\|_1 \quad (\text{LASSO})$$

$\|x\|_1 = \sum_{j=1}^N |x_j|$  et  $|\cdot|$  n'est pas dérivable. Pour y remédier :  $x_j = x_j^+ - x_j^-$  et  $|x_j| = x_j^+ + x_j^-$ . On peut reformuler le problème du LASSO avec des variables positives :

$$\min_{(x^+, x^-)} \frac{1}{2} \|y - \mathbf{A}x^+ + \mathbf{A}x^-\|_2^2 + \lambda e^T (x^+ + x^-).$$

## Relaxation convexe : LASSO et nombre de mesures

- Pour reconstruire tout vecteur  $\alpha$ ,  $s$ -parcimonieux à partir du vecteur de mesure  $y$  grâce au **LASSO**, on doit nécessairement avoir le nombre de mesures  $M$  qui vérifie

$$M \geq s \log \left( \frac{eN}{s} \right)$$

à une constante près.

- Lorsque  $\delta_{2s}(A) \leq \sqrt{2} - 1$  alors le problème LASSO admet une solution unique.
- D'un point de vue computationnel, le résultat obtenu par le LASSO est très important par rapport au résultat obtenu par la procédure de minimisation  $\ell_0$  nécessitant  $M \geq 2s$  mais le Lasso marche bien lorsque la matrice définie positive  $\mathbf{A}^T \mathbf{A}$  est inversible.

# Poursuite gloutonne

## 1- Méthode Matching Pursuit

## Algorithmes de reconstruction : Poursuite adaptative

- La méthode de la poursuite adaptative, "Matching Pursuit" et son extension : la poursuite adaptative orthogonale, "Orthogonal Matching Pursuit" consistent à
  - trouver itérativement dans la matrice de mesure  $\Phi$  le vecteur le plus corrélé avec le signal,
  - d'ôter la contribution de ce vecteur au signal
  - et de recommencer le processus jusqu'à obtention d'un nombre d'éléments égal à la dimension du signal.

## Algorithmes de reconstruction : Poursuite adaptative

- Ce sont des méthodes itératives et généralement faciles à mettre en œuvre.
- A chaque itération, elles sélectionnent une ou plusieurs colonnes de la matrice  $A$  en fonction de sa corrélation avec le vecteur de mesure  $y$ .
- Puis, elles calculent une approximation du signal et mettent à jour le résiduel qui sera utilisé dans la prochaine itération.
- Les algorithmes gloutons se distinguent notamment par la méthode de sélection des colonnes de la matrice  $A$  et sur la façon dont le résiduel est mis à jour.
- La poursuite adaptative (MP) est la version la plus simple des algorithmes gloutons.

## Algorithmes de reconstruction : Poursuite adaptative

- ① **Initialiation** : La MP commence par initialiser le résiduel  $r$  avec le vecteur de mesure  $y$ .

Elle initialise aussi l'approximation du signal  $\tilde{\alpha}$  par un vecteur nul :

$$\mathbf{r}_0 \leftarrow \mathbf{y} \text{ et } \tilde{\alpha} \leftarrow \mathbf{0}$$

- ② **A chaque itération  $k$**  : la MP sélectionne une colonne de la matrice  $\mathbf{A}$  ayant une corrélation maximale avec le résiduel :

$$\lambda_k = \operatorname{argmax}_i \left| \langle r_{k-1}, a_i \rangle \right|$$

où :

- $\lambda_k$  est l'indice de la colonne sélectionnée.
- $r_{k-1}$  est le résiduel de l'itération précédente.
- $a_i, i \in \llbracket 1, N \rrbracket$  représente les colonnes de la matrice  $\mathbf{A}$ .

## Algorithmes de reconstruction : Poursuite adaptative OM

- ③ **Mise à jour** : La MP calcule une nouvelle approximation du signal et met à jour le résiduel :

$$\tilde{\alpha}_k = \tilde{\alpha}_{k-1} + \langle r_{k-1}, a_{\lambda_k} \rangle a_{\lambda_k} \quad (5)$$

$$r_k = r_{k-1} - \langle r_{k-1}, a_{\lambda_k} \rangle a_{\lambda_k} \quad (6)$$

- $\tilde{\alpha}_{k-1}$  représente l'approximation du signal obtenue durant l'itération précédente.
- $a_{\lambda_k}$  est la colonne de la matrice  $\mathbf{A}$  sélectionnée.

Le résiduel peut aussi être exprimé en fonction de l'approximation du signal  $\tilde{\alpha}_k$  et du vecteur de mesure  $y$  :

$$r_k = y - \tilde{\alpha}_k. \quad (7)$$

## Algorithme de poursuite adaptative : critères d'arrêt

- 1 Arrêt après un nombre fini d'itérations.
- 2 Arrêt lorsque l'amplitude du résiduel est inférieure à un seuil prédéfini.
  - L'inconvénient de cette méthode est que la matrice de mesure  $\mathbf{A}$  n'est pas toujours orthogonale.
  - Dans ce cas, une colonne de la matrice  $\mathbf{A}$  peut être sélectionnée à plusieurs reprises pendant la phase de sélection.
  - La MP converge exponentiellement.

# Poursuite gloutonne

## 2- Méthode Orthogonal Matching Pursuit

## Algorithmes de reconstruction : Poursuite adaptative orthogonale OMP

- 1 La phase d'initialisation et de sélection sont les mêmes que pour la poursuite adaptative OM.
- 2 La méthode de la poursuite adaptative orthogonale OMP améliore la convergence de l'OM : à l'itération  $k$ , l'OMP sélectionne le résidu orthogonal avec les colonnes de la matrice  $A$ .
- 3 Une colonne de la matrice  $A$  est choisie au maximum une seule fois pendant la phase de sélection.

## Références

- ① An Introduction To Compressive Sampling, Emmanuel J. Candès and Michael B. Wakin, EEE SIGNAL PROCESSING MAGAZINE, MARCH 2008.
- ② Compressive sampling, Emamnuel J. Candès, Proceedings of the International Congress of Mathematicians, Madrid, Spain, 2006, European Mathematical Society.
- ③ Thèse : Représentations parcimonieuses pour les signaux multivariés - Quentin Barthélemy
- ④ Une petite introduction aux représentations parcimonieuses de signaux et d'images - Jérôme Landré 01-07-2008
- ⑤ Allen Y. Yang. Compressed sensing meets machine learning - classification of mixture subspace models via sparse representation. Mini Lectures in Image Processing, TRUST Center Seminar, University of California, Berkeley, USA, 2008.
- ⑥ Analyse Multi-Résolution et Ondelettes (AMO) Représentations des signaux 1D et 2D, Nicolas Thome Laboratoire d'Informatique de Paris 6 (LIP6) Université Pierre et Marie Curie (UPMC) Master 2 Informatique - Spécialité IMA
- ⑦ Compressive Sensing Richard G. Baraniuk IEE SIGNAL PROCESSING MAGAZINE [118] JULY 2007
- ⑧ R.G. Baraniuk. Compressive Sensing [lecture notes]. IEEE Signal Processing Magazine,
- ⑨ [DE03] David L. Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via  $l_1$  minimization. Proceedings of the National Academy of Sciences,
- ⑩ Approche d'éterministe de l'acquisition comprimée et la reconstruction des signaux issus de capteurs intelligents distribués THESE présentée et soutenue publiquement le 9 novembre 2015 pour l'obtention du Doctorat de l'Université de Lorraine (mention systèmes électroniques) par Andrianiaina Ravelomanantsoa
- ⑪ A User's Guide to Compressed Sensing for Communications Systems Kazunori HAYASHI, Masaaki NAGAHARA, and Toshiyuki TANAHA, IEICE TRANS COMMUN, VOL.E96-B, NO.3 MARCH 2013

## Références

- 12 D.L Donoho, "Compressed sensing", IEEE Tran Inf. Theory, vol52, no.4, pp. 1289-1306 April 2006.
- 13 E.J. Candes and T.Tao, "Decoding by linear programming" ; IEEE Trans. Inf. Theory, Vol.51, no12, pp.4203-4215, Dec.2005
- 14 Regression Shrinkage and Selection via the Lasso, Robert Tibshirami, journal of the Royal Statistical Society. Series B (Methodological), Volume 58, Issue 1 (1996), 267-288
- 15 Compressed Sensing with Coherent and Redundant Dictionaries, Emmanuel J. Candes, Yonina C. Eldar, Deanna Needell, Paige Randall, Applied and Computational Harmonic Analysis, Volume 31, Issue 1, July 2011, Pages 59-73
- 16 Introduction au Compressed Sensing, notions de complexité algorithmique et relaxation convexe, Guillaume Lécué, CNRS, CREST, ENSAE.