

Filtre de Wiener



Analyse en Composantes Principales



École des Ponts
ParisTech

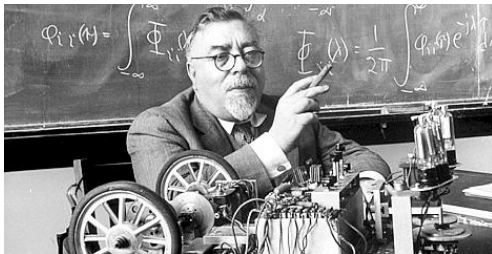
Guillaume Obozinski

LIGM/Ecole des Ponts - ParisTech



Traitement de l'information et vision artificielle
Ecole des Ponts

Filtre de Wiener



Norbert Wiener (1894 - 1964)

Image floutée

$$g(x, y) = f(x, y) * h(x, y) + n(x, y)$$

- ▶ $f(x, y)$ image d'origine
- ▶ $h(x, y)$ noyau de convolution (noyau de floutage)
- ▶ $n(x, y)$ bruit additif
- ▶ $g(x, y)$ image dégradée

But

Estimer le signal d'origine à partir du signal flouté et bruité, en supposant connu le noyau h .

Formalisation

Trouver une estimée du signal \hat{f} tel que

$$\mathbb{E} \left[\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (\hat{f}(x, y) - f(x, y))^2 dx dy \right],$$

où \mathbb{E} est l'espérance par rapport au bruit, soit minimal.

Notations

On utilise les notations classiques pour présenter le filtre de Wiener

Definition	Image	T. de Fourier
image d'origine	$f(x,y)$	$F(u,v)$
noyau	$h(x,y)$	$H(u,v)$
image dégradée	$g(x,y)$	$G(u,v)$
bruit	$n(x,y)$	$N(u,v)$
estimateur	$\hat{f}(x,y)$	$\hat{F}(u,v)$



► \hat{f} ne note pas la transformée de Fourier de f : c'est F .



► Ces transformées sont des fonctions à valeurs complexes !

Approche naïve

Dans le domaine spatial :

$$g(x, y) = f(x, y) * h(x, y) + n(x, y)$$

Dans le domaine spectral :

$$G(u, v) = F(u, v)H(u, v) + N(u, v)$$

Si on ignore le bruit

$$\hat{F}(u, v) = \frac{G(u, v)}{H(u, v)}$$

Problème en présence de bruit

Le noyau est typiquement proche de 0 aux hautes fréquences, alors que le bruit est d'amplitude non nulle

Filtre de Wiener

- ▶ $S_f(u, v) = |F(u, v)|^2$ densité spectrale de puissance du signal
- ▶ $S_n(u, v) = \mathbb{E}[|N(u, v)|^2]$ densité spectrale de puissance du bruit
- ▶ $K(u, v)$ rapport bruit sur signal (inverse du rapport signal sur bruit).

Wiener propose

$$\hat{F}(u, v) = W(u, v) G(u, v)$$

avec

$$W(u, v) = \frac{\overline{H}(u, v)}{|H(u, v)|^2 + K(u, v)}$$

- ▶ **Problème** : $K(u, v)$ est évidemment inconnu.
- ▶ En pratique, on traite $K(u, v) = K$ comme une constante à ajuster.
- ▶ Lorsque $K \rightarrow 0$ on retrouve le cas sans bruit
- ▶ Lorsque $K \gg |H(u, v)|^2$ on atténue les hautes fréquences

Le filtre de Wiener minimise les moindres carrés

Par Plancherel :

$$\mathbb{E} \left[\int_{\mathbb{R}^2} (\hat{f}(x, y) - f(x, y))^2 dx dy \right] = \mathbb{E} \left[\int_{\mathbb{R}^2} |\hat{F}(u, v) - F(u, v)|^2 dudv \right].$$

$$|F(u, v) - \hat{F}(u, v)|^2 = |F(u, v) - W(u, v)(F(u, v)H(u, v) + N(u, v))|^2$$

On n'écrit plus la dépendance en (u, v) . On a alors :

$$\begin{aligned} \mathbb{E}|(1 - WH)F - WN|^2 &= |(1 - WH)F|^2 + \mathbb{E}|WN|^2 \\ &= (1 - 2W\bar{H} + |W|^2|H|^2)|F|^2 + |W|^2\mathbb{E}|N|^2 \end{aligned}$$

$$\frac{1}{2} \frac{\partial(\mathbb{E}|(1 - WH)F - WN|^2)}{\partial W} = (-\bar{H} + W|H|^2)|F|^2 + W\mathbb{E}|N|^2$$

Gradient nul ssi $-\bar{H} + W(|H|^2 + K) = 0$ puisque $K = \frac{\mathbb{E}|N|^2}{|F|^2}$.

Donc

$$W = \frac{\bar{H}}{|H|^2 + K}$$

Analyse en composantes principales (1901)



Karl Pearson (1857 - 1936)

Deux points de vue différents

Le but de l'analyse en composantes principales (ACP) est de trouver une base de l'espace dans lequel une famille de signaux se trouvent et qui soient bien adaptés à ces signaux.

Deux points de vues. Deux problèmes.

- ▶ Point de vue de l'analyse de signaux
- ▶ Point de vue de la synthèse de signaux

Une solution : l'**ACP** !

Matrice de design

Soit une famille de signaux finis $x_i \in \mathbb{R}^d$ pour $i = 1 \dots n$.

Matrice de design

C'est la matrice :

$$X = \begin{bmatrix} \text{---} & x_1^\top & \text{---} \\ \vdots & \vdots & \vdots \\ \text{---} & x_n^\top & \text{---} \end{bmatrix}.$$

On suppose que les signaux sont **centrés**, c'est-à-dire que $\sum_{i=1}^n x_i = 0$.

Si on part de signaux x_i qui ne sont pas centrés on construit la matrice de design X dont les lignes sont les $x_i - \Psi^\top$ avec $\Psi = \frac{1}{n} \sum_{i=1}^n x_i$.

Covariance empirique, corrélation empirique

Pour des signaux centrés :

$$\hat{\Sigma} = \frac{1}{n} X^T X = \frac{1}{n} \sum_{i=1}^n x_i x_i^T$$

Pour des signaux non-centrés :

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (x_i - \Psi)(x_i - \Psi)^T$$

On peut aussi **normalisées** (où réduire) les données en divisant chaque colonne par son écart-type. On obtient alors la matrice de corrélations

$$C = \text{Diag}(\hat{\sigma})^{-1} \hat{\Sigma} \text{Diag}(\hat{\sigma})^{-1} \quad \text{avec} \quad \hat{\sigma}_k^2 = \hat{\Sigma}_{k,k}.$$

$$C_{k,k'} = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i^{(k)} - \Psi^k}{\hat{\sigma}_k} \right) \left(\frac{x_i^{(k')} - \Psi^{k'}}{\hat{\sigma}_{k'}} \right).$$

La normalisation (réduction) est optionnelle...

ACP : Point de vue de l'analyse de signaux

Les signaux vivent dans \mathbb{R}^d on cherche une direction v dans \mathbb{R}^n telle que la variance des signaux projetés selon cette direction soit maximisée. Or

$$\begin{aligned} \text{Var}((v^\top x_i)_{i=1\dots n}) &= \frac{1}{n} \sum_{i=1}^n (v^\top x_i)^2 \\ &= \frac{1}{n} \sum_{i=1}^n v^\top x_i x_i^\top v \\ &= v^\top \left(\frac{1}{n} \sum_{i=1}^n x_i x_i^\top \right) v \\ &= v^\top \hat{\Sigma} v \end{aligned}$$

On cherche donc à résoudre

$$\max_{\|v\|_2=1} v^\top \hat{\Sigma} v$$

Solution : premier vecteur propre de $\hat{\Sigma}$ disons v_1

Déflation

Comment trouver une deuxième direction sur laquelle projeter les données pour maximiser la dispersion des données ?

On fait une déflation

$$\forall i, \quad \tilde{x}_i \leftarrow x_i - v_1(v_1^T x_i)$$

Donc pour la matrice : $\tilde{X} \leftarrow X - Xv_1v_1^T$.

Puis on cherche la direction qui maximise la variance pour

$$\tilde{\Sigma} = \frac{1}{n} \tilde{X}^T \tilde{X}$$

On résout donc $\max_{\|v\|_2} v^T \tilde{\Sigma} v$

De façon équivalente $\max_{\|v\|_2} v^T \hat{\Sigma} v$ tel que $v \perp v_1$.

Solution : Second vecteur propre de $\hat{\Sigma}$ disons v_2 . Etc.

Directions principales et diagonalisation de la covariance

On appelle

- ▶ **directions principales** du nuage de points les vecteurs

$$v_1, v_2, \dots, v_k.$$

- ▶ **composantes principales** :
la projection des données sur les k directions principales.

Les directions principales sont les vecteurs propres de $\hat{\Sigma} = V S^2 V^T$.

Décomposition en valeurs singulières et ACP

La décomposition en valeurs singulières (SVD) d'une matrice $X \in \mathbb{R}^{n \times p}$ telle que $n \leq p$ est une factorisation de la forme $X = USV^T$, avec

- ▶ $U \in \mathbb{R}^{n \times n}$ une base orthogonale de \mathbb{R}^n
- ▶ $S \in \mathbb{R}^{n \times p}$ une matrice diagonale (de forme rectangulaire).
- ▶ $V \in \mathbb{R}^{p \times p}$ une base orthogonale de \mathbb{R}^p

SVD compacte

On utilise plus souvent la SVD réduite : Si r est le rang de X alors $X = USV^T$ avec,

- ▶ $U \in \mathbb{R}^{n \times r}$ dont les colonnes sont orthonormées.
- ▶ $S \in \mathbb{R}^{r \times r}$ une matrice diagonale **carrée**.
- ▶ $V \in \mathbb{R}^{r \times p}$ dont les colonnes sont orthonormées.

Si la diagonale de S est telle que $s_1 > s_2 > \dots > s_r > 0$ et $U_{1k} \geq 0$ pour tout k la SVD réduite est unique.

On a que

- ▶ US^2U^T est une diagonalisation (compacte) de la matrice XX^T
- ▶ VS^2V^T est une diagonalisation (compacte) de la matrice $X^T X$

Décomposition en valeurs singulières et ACP-II

Si $V = [v_1, \dots, v_n]$ et $U = [u_1, \dots, u_n]$ alors

- ▶ s_k est la k ième **valeur singulière** de X .
- ▶ v_k est le k ième **vecteur singulier à droite**
- ▶ u_k est le k ième **vecteur singulier à gauche**

Pseudo-inverse de X de Moore-Penrose

Si $X = USV^T$ est la SVD compacte de X on définit le pseudo inverse par

$$X^\dagger = VS^{-1}U^T.$$

Exercice :

$$X^\dagger X = ? \quad XX^\dagger = ?, \quad XX^\dagger X = ?, \quad X^\dagger XX^\dagger = ?$$

ACP à partir de la SVD

Si $X = USV^T$ alors on a

- ▶ les axes principaux $V_{[k]} = [v_1, \dots, v_k]$
- ▶ les composantes principales sont donc

$$XV_k = USV^T V_{[k]} = U_{[k]}S_{[k]}$$

avec $U_{[k]} = [u_1, \dots, u_k]$ et $S_{[k]} = \text{Diag}(s_1, \dots, s_k)$.

Et symétriquement que peut-on dire de $U_{[k]}$?

ACP des données vs ACP des variables

ACP des données

Nous avons vu jusqu'à présent l'ACP des données :

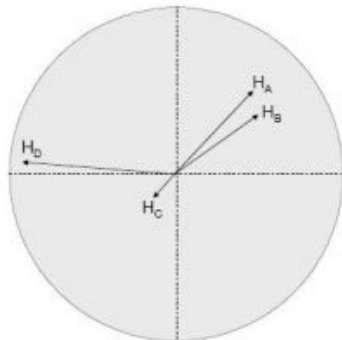
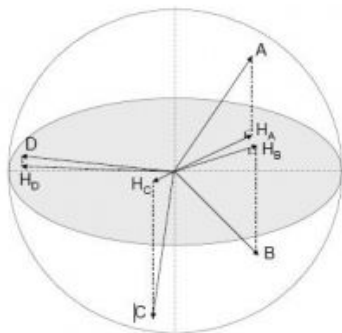
- ▶ Les x_i vivent dans \mathbb{R}^p
- ▶ Les **axes principaux** sont les $v_k \in \mathbb{R}^p$
- ▶ Les **composantes principales** sont les lignes de $U_{[k]}S_{[k]}$

ACP des variables

On centre + normalise la matrice de design + $\times \frac{1}{\sqrt{n}}$. Comme les variables correspondent aux colonnes $x^{(j)}$ de X , on fait une ACP sur X^T dont

- ▶ Les $x^{(j)}$ vivent dans \mathbb{R}^n et $\|x^{(j)}\|_2^2 = 1$ par normalisation.
- ▶ Les **variables principales** sont les $u_k \in \mathbb{R}^n$
- ▶ Les décomposition des variables sur les variables principales sont les lignes de $V_{[k]}S_{[k]}$

Cercle des corrélations



ACP : Point de vue de synthèse de signaux

Etant donnés des signaux on cherche

- ▶ une famille de K **atomes** $\mathbf{A}_k \in \mathbb{R}^p$ tels que $\|\mathbf{A}_k\|_2 = 1$,
- ▶ formant un **dictionnaire** $\tilde{V} \in \mathbb{R}^{p \times K}$
- ▶ et des coefficients de décompositions $\tilde{U} = (\Delta_{ik})_{1 \leq i \leq n, 1 \leq k \leq K}$

de sorte que $x_i \approx \sum_{k=1}^K \Delta_{ik} \mathbf{A}_k$. On demande de minimiser

$$\min_{\tilde{U} \in \mathbb{R}^{n \times K}, \tilde{V} \in \mathbb{R}^{p \times K}} \sum_{i=1}^n \left\| x_i - \sum_{k=1}^K \Delta_{ik} \mathbf{A}_k \right\|_2^2$$

Ce qui est équivalent à

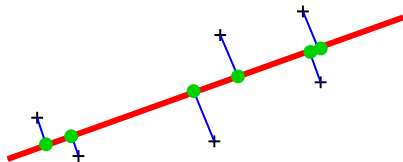
$$\min_{\tilde{U} \in \mathbb{R}^{n \times K}, \tilde{V} \in \mathbb{R}^{p \times K}} \sum_{i=1}^n \left\| X - \tilde{U} \tilde{V}^T \right\|_F^2 \quad \text{t.q.} \quad \forall k, \|\mathbf{A}_k\|_2 = 1.$$

Remarque : Il n'y a pas de contrainte d'orthogonalité entre les v_k .

Solution : $\tilde{V} = V_{[k]}$ et $\tilde{U} = U_{[k]} S_{[k]}$ à permutation des colonnes près et des signes des colonnes de V et U .

Analyse et synthèse

Les problèmes de synthèse et d'analyse sont tous deux résolus par l'ACP



Il existe plusieurs généralisations de l'ACP pour les deux points de vue.

En particulier, si on autorise K à être grand (éventuellement) et qu'on requiert que pour chaque i la plupart des coefficients $(u_{ik})_{1 \leq k \leq K}$ sont nuls. On obtient des dictionnaires qui ont des bonnes propriétés pour le traitement des images et des signaux en général.