

**SURECHANTILLONNAGE, VITESSE ET VARIANCE RESIDUELLE  
D'UN FILTRE ADAPTATIF POUR SIGNAL DE DONNEES**

Sylvie Marcos et Odile Macchi

Laboratoire des Signaux et Systèmes-ESE Plateau du Moulon  
91190 Gif-sur-Yvette

**RESUME.** On considère l'influence de la fréquence d'échantillonnage des signaux sur les performances d'un filtre adaptatif utilisé en identification, lorsque l'entrée est un signal de données. Un exemple pratique est celui de l'annulation d'écho, et on parle alors du suréchantillonnage par rapport au rythme des données. Le suréchantillonnage augmente considérablement la dispersion de la matrice de covariance du signal d'entrée dégradant les performances du filtre adaptatif en termes de vitesse de convergence. Au contraire, on montre que la variance résiduelle n'est pas dégradée. On distingue deux cas suivant que la fréquence d'adaptation est supérieure ou inférieure à la fréquence de Shannon. Alors que les performances demeurent optimales dans le premier cas, elles se dégradent notablement dans le second cas. On propose alors une solution pour diminuer la dispersion de la matrice de covariance.

**SUMMARY** In adaptive filtering used in identification problems with an input data signal, the influence of the sampling rate is analysed. Echo cancellation is a practical example. It is the so-called "oversampling rate" relative to the data rate. Oversampling increases strongly the eigenvalue spread of the input signal covariance matrix, deteriorating the convergence speed of the adaptive filter. At the opposite it is found that the residual variance is not deteriorated. Two cases are considered corresponding to an adaptation rate that does or does not satisfy the Shannon theorem. In the first case performances remain optimal, whereas in the second case they are deteriorated a lot. Then we propose a solution to decrease the eigenvalue spread of the covariance matrix.

**I. INTRODUCTION**

Le contexte considéré est celui de l'identification d'un système analogique par un système numérique adaptatif. Ce problème est très général et donne lieu à de nombreuses applications ; c'est le cas par exemple de l'annulation d'écho où il s'agit de reproduire le canal d'écho à l'aide d'un filtre adaptatif (FA) non récursif.

Soit  $c(t)$  la réponse impulsionnelle du système inconnu et  $a(t)$  le signal à son entrée. Le signal de référence dont on dispose pour identifier  $c(t)$  est, en l'absence de bruit, le signal de sortie

$$\sigma'(t) = a(t) * c(t). \tag{1}$$

Le problème d'identification se pose alors comme l'estimation adaptative de  $\sigma'(t)$  à partir d'un traitement numérique linéaire du signal connu  $a(t)$  échantillonné.

On s'intéresse ici à l'influence de la fréquence d'échantillonnage sur l'algorithme du gradient stochastique à pas constant. Soit  $[-w, w]$  la bande passante du système inconnu et de  $a(t)$  ; on peut écrire

$$\sigma'(t) = \theta \sum_n a(n\theta) c(t-n\theta) = \theta \sum_n c(n\theta) a(t-n\theta) \tag{2}$$

pour une période d'échantillonnage  $\theta$  vérifiant le théorème de Shannon :

$$\theta < 1/2w. \tag{3}$$

Dans la suite le filtrage sera toujours effectué à la fréquence fixe  $1/\theta$  ; c'est la fréquence d'adaptation  $1/T$  du FA qui sera variable.

Le signal d'erreur à la sortie s'écrit à l'instant  $jT$

$$e_j = \sigma'_j - \sigma_{j+n_j} \tag{4}$$

où d'après (2)

$$\sigma'_j = \sigma'(jT) = C^T A_j. \tag{5}$$

Dans (4), (5),  $A_j$  désigne le vecteur formé par les échantillons  $a(jT-n\theta)$  ;  $\sigma_j$  est l'estimée de  $\sigma'_j$  à la sortie du FA lequel est désigné par  $\hat{C}_j$

$$\sigma_j = \hat{C}_j^T A_j, \tag{6}$$

et  $n_j$  constitue du bruit que l'on supposera blanc, centré et indépendant de  $A_j$  et de  $\hat{C}_j$ .

Les vecteurs  $\hat{C}_j$  optimaux réalisant le minimum de l'erreur quadratique moyenne (eqm) vérifient

$$R(C - \hat{C}) = 0 \tag{7}$$

où

$$R = E[A_j * A_j^T]. \tag{8}$$

Lorsque le processus à l'entrée du FA est stationnaire la matrice  $R$  est de Toeplitz. La variété des vecteurs optimaux  $\hat{C}$  dépend de la régularité ou de la singularité de la matrice  $R$ . Même lorsque (7) admet d'autres solutions, nous appellerons  $C$  le vecteur optimal.

L'algorithme remet à jour les coefficients de  $\hat{C}_j$  à chaque itération  $jT$  en fonction de l'observation  $\sigma'_j + n_j$  et du vecteur  $A_j$ , selon

$$\hat{C}_{j+1} = \hat{C}_j + \mu e_j A_j^* \tag{9}$$

Il réalise une estimation de l'algorithme du gradient déterministe qui minimise l'eqm à la sortie du filtre. D'après (4), cette dernière vaut

$$E_j = E[|e_j|^2] = e_{\min} + I_j \tag{10}$$

où

$$e_{\min} = E[|n_j|^2], \tag{11}$$



est le bruit de sortie, le second terme étant l'erreur résiduelle due à la non optimalité du FA :

$$r_j = E[|\sigma_j - \sigma_j^f|^2]. \quad (12)$$

Il est connu [1] que l'erreur résiduelle ou fluctuation de l'algorithme se décompose en deux termes

$$r_j = r_j^t + r_j^f. \quad (13)$$

1°) Le premier terme (transitoire  $r_j^t$ ) correspond au cas sans bruit ( $n_j=0$  dans (4)). Il décroît exponentiellement vers zéro à la condition que R soit inversible. Son temps de convergence  $\tau$  est caractéristique de la vitesse de convergence de l'algorithme. Il est connu [2][3] que  $\tau$  est inversement proportionnel au pas d'incrémentation  $\mu$  de l'algorithme mais aussi à la plus petite valeur propre de la matrice de covariance du signal d'entrée. Du fait de la contrainte sur  $\mu$  pour que l'algorithme (9) converge, on a

$$\tau > \rho/2 \quad (14)$$

où  $\rho = \lambda_{\max}/\lambda_{\min}$  est la dispersion (rapport de la plus grande à la plus petite des valeurs propres) de la matrice de covariance R.

2°) Le second terme de (13)  $r_j^f$  est la variance des fluctuations de l'algorithme autour de positions optimales telles que  $\sigma_j = \sigma_j^f$  ; il est connu [4] que

$$e_{\min} \mu N \sigma_a^2 / 2\rho \leq \lim_j r_j^f \leq e_{\min} \mu N \sigma_a^2 \rho / 2 \quad (15)$$

où  $\sigma_a^2$  est la puissance des échantillons d'entrée du FA, N est le nombre de prises. Remarquons que les résultats (14) et (15) ont été établis pour une matrice R inversible.

On voit donc que les performances du FA, mesurées en termes de vitesse de convergence  $\tau$  et d'erreur résiduelle  $r_j$  dépendent fortement des valeurs propres de la matrice de covariance R des données d'entrée.

Dans la suite on va donc analyser séparément les points 1°) et 2°) mentionnés ci-dessus. Pour mieux comprendre le phénomène on se place dans le cas où le signal  $a(t)$  est obtenu par filtrage de bruit blanc. En transmission de données cela correspond à des données indépendantes mises en forme par filtrage avant émission.

## II. INFLUENCE DE LA FREQUENCE D'ECHANTILLONNAGE SUR LA VITESSE DE CONVERGENCE

### II.1. Cas du simple échantillonnage (rappels)

En annulation d'écho pour transmission de données, le signal  $a(t)$  est un signal de données

$$a(t) = \sum_i a_i \alpha(t - iT_b) \quad (16)$$

où les  $\{a_i\}$  sont des données blanches et centrées générées avec la période baud  $T_b$ ,  $\alpha(t)$  est la réponse impulsionnelle du filtre émission. La méthode classique consiste à échantillonner l'entrée de l'annuleur d'écho à  $T = T_b$  ; le vecteur  $A_j$  de (5), (6) est donc ( $N = 2M + 1$ )

$$A_j^T = [a_{j+M}, a_{j+M-1}, \dots, a_j, \dots, a_{j-M+1}, a_{j-M}] ; \quad (17)$$

la matrice de covariance R définie en (8) a une dispersion  $\rho = 1$  puisque  $R = \sigma_a^2 I$ . La vitesse de convergence est

constante pour un pas d'incrémentation  $\mu$  constant sur l'algorithme d'adaptation ; elle suit la loi [2]  $\tau = 1/4\sigma_a^2 \mu$ . C'est ce que montre la fig.1, où  $P/r_j$  en dB est porté en ordonnée en fonction du nombre j d'itérations ; P est la puissance initiale de l'écho. Le quotient  $r_j/P$  représente l'atténuation du résidu d'écho par rapport à l'écho initial.

Lorsqu'on échantillonne l'entrée à une fréquence supérieure à  $1/T_b$ , on parle alors de suréchantillonnage.

### II.2. Adaptation à une fréquence de Shannon

Pour plus de simplicité, on choisit  $T = \theta = T_b/m$ . Le vecteur des échantillons de  $a(t)$  de (16), contenant  $a(kT)$  au centre prend alors m formes  $A'_k$  selon que  $k = mj + p$  et  $p = 0, \dots, m-1$  :

$$A'_k{}^T = [\alpha_p^T A_{j+M}, \alpha_{p+1}^T A_{j+M}, \dots, \alpha_{m-1}^T A_{j+M}, \\ \alpha_0^T A_{j+M-1}, \dots, \alpha_{m-1}^T A_{j-M}, \alpha_0^T A_{j-M-1}, \alpha_1^T A_{j-M-1}, \\ \dots, \alpha_{p-1}^T A_{j-M-1}] \quad (18)$$

soient  $mN$  composantes, où  $\alpha_p$  représente le vecteur formé des échantillons de  $\alpha(t)$  espacés de  $T_b$  mais pris aux phases  $pT$  ; le vecteur des données blanches est celui de (17). Remarquons que  $\alpha_p^T A_j$  est un scalaire ; on note K la longueur du filtre de mise en forme égale à la longueur de  $\alpha_p$ . Dans la suite on prend  $K = N = 2M + 1$ . D'autre part le FA nécessite m fois plus de coefficients que dans le cas où on échantillonne à  $T_b$ . C'est pourquoi on parle de suréchantillonnage.

Il existe alors m matrices de covariance pour les entrées  $A'_k$ , soient

$$R_p = E[A'^*{}_{mj+p} A'^T_{mj+p}], \quad p = 0, \dots, m-1. \quad (19)$$

Il est facile de voir que les matrices  $R_p$  ne sont pas de Toeplitz ; de plus elles alternent au cours du temps. Donc à chaque itération on passe de  $R_p$  à  $R_{p+1}$  pour écrire l'équation (7) d'optimalité. Le problème se ramène alors à identifier C par  $\hat{C}$  avec des références entrée/sortie non stationnaires ( $\{a_j\}, \{\sigma_j\}$ ). Il vient

$$R_p = \sigma_a^2 H_p^* H_p^T \quad (20)$$

où  $H_p^* H_p^T$  est une matrice carrée d'ordre  $mN$  ; elle n'est pas inversible lorsque  $K \leq (m-1)N$ , en particulier pour  $m \geq 2$  dans notre cas.

Le filtre d'estimation  $\hat{C}$  optimal doit donc vérifier les systèmes d'équations

$$R_p (C - \hat{C}) = 0, \quad p = 0, \dots, m-1. \quad (21)$$

Pour chacun de ces systèmes  $\hat{C} = C$  est une solution. Puisque les  $R_p$  sont non inversibles, chaque système (21) admet aussi des solutions non nulles. Comme les  $R_p$  n'ont pas le même noyau, de telles solutions peuvent être différentes suivant la valeur p de la phase d'échantillonnage. Pour appliquer les résultats (14), (15) on se place dans le cas limite d'une matrice inversible mais à très forte dispersion. La fig.2 montre que la vitesse de convergence est ralentie du fait de cette grande dispersion.

**condition des simulations** : Dans les simulations le canal d'écho  $c(t)$  est un canal de Nyquist en cosinus surélevé dont la largeur de bande est  $[-1400, 1400 \text{ Hz}]$  correspondant à  $T_b = 1/2400\text{s}$  et à un coefficient d'arondi de  $1/6$ . Les symboles sont complexes et correspondent à un diagramme à 16 points. L'anneur d'écho a 21 coefficients baud soit encore 42 coefficients lorsqu'on échantillonne à  $\theta = T = T_b/2$ . Cette fréquence 4800 Hz, qui correspond au double échantillonnage, vérifie le théorème de Shannon. Les simulations ont été faites sans bruit  $n_j$  dans l'erreur de pilotage  $e_j$  de (4). Le critère pris en compte est  $P/r_j$  où  $P$  est la puissance initiale de l'écho.

Par l'intermédiaire du temps de convergence  $\tau$ , apparaît l'influence de la fréquence d'échantillonnage sur la dispersion de la matrice de covariance. Par ailleurs, il existe plusieurs modes de convergence correspondant aux différentes valeurs propres  $\lambda_n$ ,  $n=1, \dots, (m-1)N$  de  $R$ , suivant la relation  $\tau_n = 1/4\lambda_n\mu$  [2]. Les valeurs  $1/\tau_n$  s'interprètent comme autant de pentes de convergence. Au fur et à mesure que la convergence s'améliore, les directions propres associées aux petites valeurs propres apparaissent dans le vecteur  $\hat{C}_j$ ; on comprend que la vitesse de convergence devient gouvernée par les valeurs propres les plus faibles et devient très petite. Par ailleurs, connaissant les coefficients  $\alpha_p$  du filtre de mise en forme, il est facile de calculer les valeurs propres des matrices  $H_p^* H_p^T$  de (20). On vérifie ainsi l'existence de valeurs propres nulles, la valeur propre maximale étant égale à  $m$  (par normalisation de l'émission).

Pourtant il faut remarquer (fig.2) que les résultats restent bons puisqu'on observe 100 dB d'atténuation en 12000 itérations.

### II.3 Adaptation à une fréquence inférieure à la fréquence de Shannon

Dans le cas de figure simulé en II.2, supposons que la sortie du filtre soit calculée avec la période  $\theta$  alors que les coefficients de  $\hat{C}_j$  ne sont adaptés qu'une fois sur deux ( $T = 2\theta = T_b$ ). Le rythme d'adaptation  $1/T = 2400 \text{ Hz}$  est donc inférieur au rythme de Shannon au moins égal à 2800 Hz. L'algorithme ne résoud qu'un seul des systèmes d'équations (21) par exemple

$$R_0(C - \hat{C}) = 0, \quad (p=0) \tag{22}$$

ce qui n'annule pas  $R_1(C - \hat{C})$ . C'est ce que l'on constate sur les simulations de la fig.3. Le résidu  $r_j$  est calculé pour les deux phases  $t = jT_b$  ( $p=0$ , phase adaptée) et  $t = jT_b + \theta$  ( $p=1$ , phase non adaptée). On voit bien que la solution de  $R_0(C - \hat{C}) = 0$  atteinte sur l'une des phases ne convient pas à annuler  $R_1(C - \hat{C})$ . Par ailleurs, pour la phase où l'on adapte, le résultat est plus mauvais qu'en double échantillonnage (voir fig.2). La vitesse est aussi lente puisqu'on obtient la même pente limite correspondant à une dispersion de  $R_0$  très grande, mais la valeur propre minimale impose sa vitesse plus tôt qu'en double échantillonnage. Cette fois l'atténuation n'est plus que de 60 dB. En conclusion il est préférable d'adapter le FA à la fréquence de Shannon.

Pourtant, remarquons que l'on peut envisager de travailler à un rythme inférieur à la fréquence de Shannon (ici en adaptant une fois sur deux) à condition d'ajouter du bruit blanc sur le signal  $a(t)$ . Dans ce cas les matrices  $R_0$  et  $R_1$  deviennent inversibles;  $R_0(C - \hat{C})$  donne la solution unique  $\hat{C} = C$  et les dispersions de  $R_0$  et  $R_1$  diminuent, ce qui favorise la convergence (cf. fig.4 où la puissance du bruit ajouté est de 10 dB inférieure à celle du signal  $a(t)$ ).

### III. ERREUR DE FLUCTUATION EN PRESENCE DE BRUIT

Dans (4) le bruit  $n_j$  de puissance  $e_{\min}$  est constitué par les données lointaines. Il est indépendant des signaux  $a_j$  et  $\sigma'_j$ . L'erreur de fluctuation  $r_k^f$  par rapport à l'optimalité ou bruit de l'algorithme suit à l'infini la relation (15). Dans le cas du simple échantillonnage (cf. II.1) où la matrice a une dispersion  $\rho = 1$ , on obtient directement la limite  $r_k^f = e_{\min} \mu N \sigma_a^2 / 2$ , ce que l'on vérifie aisément sur la fig.5 où  $N = 21$ ,  $\sigma_a^2 = 10$  en observant les niveaux de saturation de  $e_{\min} / r_k$ . Notons que cette quantité correspond au rapport signal sur bruit pour la détection des signaux lointains et que le temps est mesuré en temps baud.

Dans le cas du suréchantillonnage, où  $\rho$  est grand, la relation (15) n'indique plus rien sur l'erreur de fluctuation limite. On peut cependant reprendre le calcul de  $r_k^f$ . A partir de (9), (4), (5) et (6), on calcule en posant  $V_k = \hat{C}_k - C$

$$E[|V_{k+1}|^2] = E[|V_k|^2] + \mu^2 E[|V_k^T A_k^* A_k^T|^2] - 2\mu E[|V_k^T A_k|^2] + \mu^2 e_{\min} E[|A_k|^2]. \tag{24}$$

Il n'est pas déraisonnable de supposer que  $E[|V_k^T A_k^* A_k^T|^2] = E[|A_k|^2] E[|V_k^T A_k|^2]$ . On montre aisément d'après (18) que  $\text{Trace}(R_p) = N \sigma_a^2 \sum_p |\alpha_p|^2 = N m \sigma_a^2$ . D'après (24), en supposant  $\mu$  suffisamment petit il vient

$$\lim_k r_k = e_{\min} \mu N m \sigma_a^2 / 2. \tag{25}$$

Ainsi le résidu (25) est indépendant de la phase d'échantillonnage  $p$ . Contrairement à ce que laissait croire (15) la variance n'est pas affectée par la dispersion de la matrice de covariance; la finesse de convergence n'est donc pas dégradée. Par l'intermédiaire du produit  $m\mu$  seule la normalisation de  $\mu$  change par rapport au simple échantillonnage. Sur la fig.6 apparaît  $e_{\min} / r_k$  pour différentes valeurs de  $\mu$ . Le résidu est proportionnel à  $\mu$  en régime permanent.

Ainsi pour augmenter la finesse de convergence il faut diminuer  $\mu$ . Mais ceci est aux dépens de la vitesse de convergence qui est déjà lente d'après l'étude de II. Au lieu de cela il est préférable de diminuer  $e_{\min}$ ; par exemple en annulation d'écho on fera fonctionner d'abord le FA sans émettre de signal lointain, ce qui en diminuant  $e_{\min}$  permet dans (25) de garder le même  $\mu$ . Lorsque la finesse de convergence désirée est atteinte, on peut ajouter le signal distant tout en diminuant  $\mu$  de façon à garder cette précision de convergence.



**IV. CONCLUSION**

L'analyse faite ici dans le cas de l'annulation d'un écho de données, générées à une période  $T_b$ , a permis d'explicitier l'influence de la fréquence d'échantillonnage et de la fréquence d'adaptation sur un système d'identification adaptative. L'identification numérique est faite à une fréquence de Shannon  $1/\theta > 1/T_b$ . Mais on peut adapter le FA à une fréquence plus basse  $1/T < 1/\theta$ . Le suréchantillonnage rend les processus d'entrée/sortie du FA non stationnaires ; il existe autant de matrices de covariance que de phases d'échantillonnage  $T_b/\theta$ . Les performances optimales ne sont atteintes que si le FA est adapté à une fréquence de Shannon. Cependant la convergence reste très lente du fait d'une dispersion accrue de la covariance. Cet inconvénient est facile à éliminer au cours d'une phase d'apprentissage. Il suffit d'ajouter du bruit blanc dans la référence d'entrée. Enfin, il est satisfaisant de constater que l'erreur de fluctuation n'est pas dégradée par le suréchantillonnage.

**REFERENCES**

- [1] O. Macchi, Advances in adaptive filtering, Digital Communications, Edited by E. Biglieri and G. Prati, North Holland 1986, pp 41-57
- [2] B. Widrow, S.D. Stearns, Adaptive signal processing, Prentice-Hall, 1985
- [3] C.F.N. Cowan, P.M. Grant, Adaptive filters, Prentice-Hall, 1985
- [4] O. Macchi, Le filtrage adaptatif en télécommunications, Ann. des télécom., tome 36, n°11-12, Nov-Dec 1981, pp 615-625

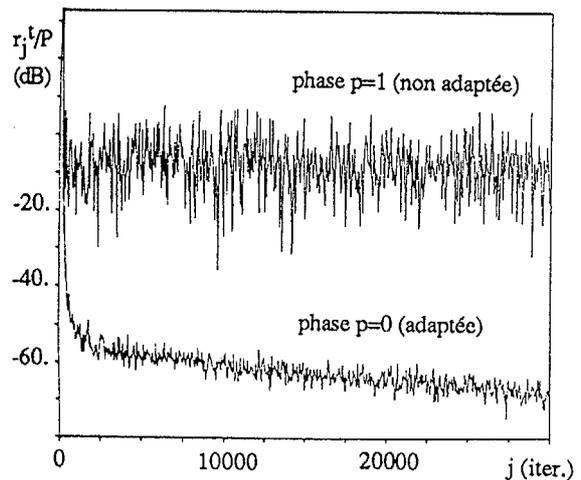


Figure 3 :  $r_j^t/P$ , cas du double échantillonnage, adaptation à une fréquence inférieure à Shannon.  $\mu=2^{-9}$

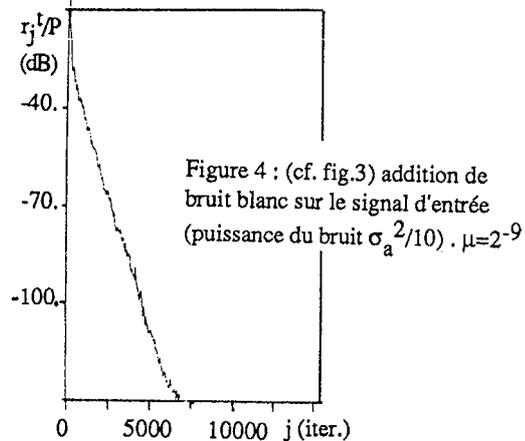


Figure 4 : (cf. fig.3) addition de bruit blanc sur le signal d'entrée (puissance du bruit  $\sigma_a^2/10$ ).  $\mu=2^{-9}$

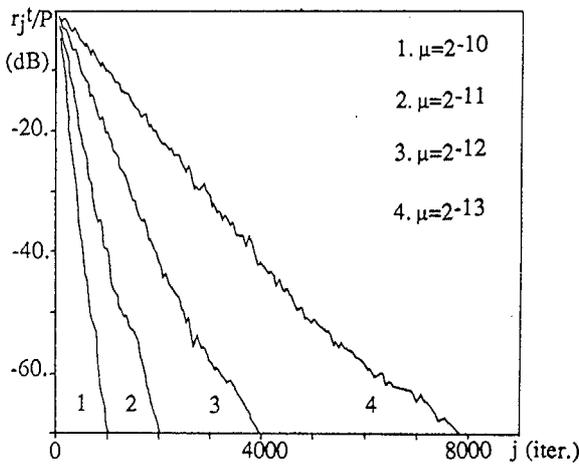


Figure 1 :  $r_j^t/P$  en fonction de  $\mu$ , cas du simple échantillonnage  $\rho=1$

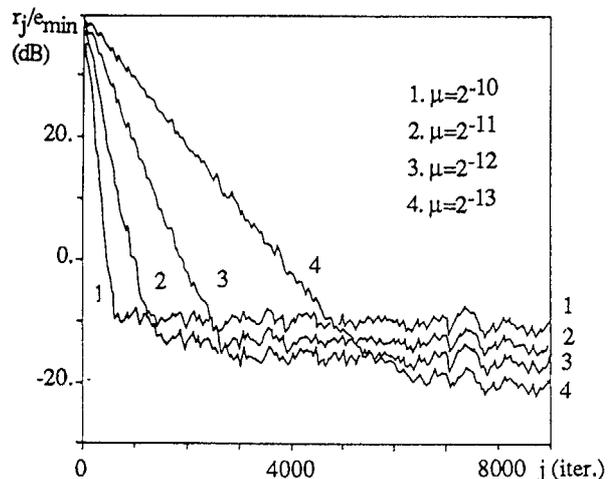


Figure 5 :  $r_j/e_{min}$  en fonction de  $\mu$ , cas du simple échantillonnage.  $P/e_{min}=40$  dB

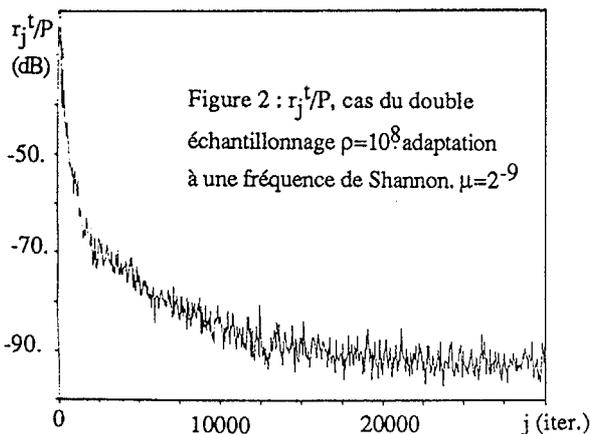


Figure 2 :  $r_j^t/P$ , cas du double échantillonnage  $\rho=10^8$  adaptation à une fréquence de Shannon.  $\mu=2^{-9}$

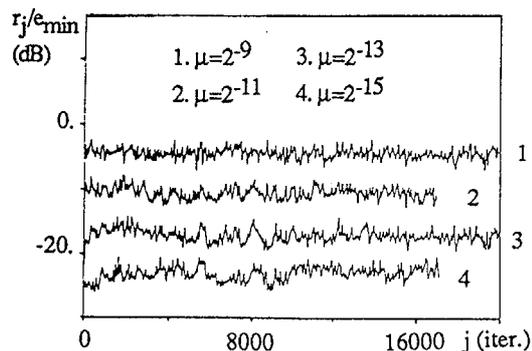


Figure 6 :  $r_j/e_{min}$  en fonction de  $\mu$ , cas du double échantillonnage.  $P/e_{min}=40$  dB