

Au sujet d'un contraste non symétrique

Eric Moreau⁽¹⁾ et Nadège Thirion⁽²⁾

⁽¹⁾MS-GESSY, ISITV, Université de Toulon et du Var
av. G. Pompidou, BP56, 83162 La Valette du Var Cedex
e-mail : moreau@isitiv.univ-tln.fr

⁽²⁾LESI, ESPEO, Université d'Orléans
12 rue de Blois, BP6744, 45067 Orléans Cedex 2
e-mail : thirion@lesiv.univ-orleans.fr

RÉSUMÉ

On considère le problème de la séparation de sources basée sur l'optimisation de critères. Nous proposons une définition des fonctions de contraste plus générale afin de pouvoir considérer des fonctions non symétriques et nous donnons deux nouveaux contrastes. Dans le cas de deux sources, nous déterminons le coefficient de dissymétrie optimal en minimisant l'erreur quadratique moyenne d'estimation. Nous illustrons les résultats au moyen de simulations numériques qui témoignent de l'intérêt qu'il peut y avoir à considérer un contraste non symétrique.

ABSTRACT

We consider the problem of sources separation and particularly criteria based approaches. A generalized definition of contrast function is given in order to consider non symmetrical functions. Two generalized contrasts involving high-order cumulants are proposed. In the case of two sources, we derive the optimal non symmetrical coefficient by minimizing the mean square error. Finally computer simulations are presented in order to illustrate the results and to show the interest in considering a non symmetrical contrast.

1 Introduction

Nous considérons le problème de la *séparation de sources* [4]. Les sorties d'un système réalisant un mélange de signaux aléatoires inconnus appelés sources, sont observées. Le mélange étant lui aussi supposé inconnu, on désire retrouver le plus fidèlement possible chacun des signaux d'entrée (ou sources) grâce aux seules observations des sorties. Il s'agit donc de déterminer ou d'estimer un inverse du système mélangeant. Ce problème est souvent qualifié d'autodidacte, de non-supervisé ou d'aveugle dans la mesure où la séparation est réalisée avec pour seule information celle apportée par les observations. Pour cela il est nécessaire de faire l'hypothèse importante suivante : *les signaux sources sont statistiquement indépendants*.

Parmi les nombreuses approches récemment proposées dans la littérature, nous nous intéressons plus particulièrement à l'optimisation de critères prenant en compte des statistiques d'ordre supérieur à deux, e.g. [1]-[7]. Dans ce domaine les fonctions de contraste constituent des critères de séparation en ce sens que leur maximisation résout le problème de la séparation de sources. La définition initialement proposée impose aux contrastes d'être des fonctions symétriques et invariantes par changement d'échelle. Leur maximisation est une condition nécessaire et suffisante de séparation.

Dans cette communication nous considérons "simplement" des conditions suffisantes, notre principal but étant de montrer quel avantage on peut en retirer. Auparavant, nous proposons

une généralisation de la notion de contraste afin de pouvoir inclure les nouvelles fonctions non symétriques. Notons que des critères non symétriques ont été proposés dans [1] avec une approche différente.

2 Notations et hypothèses

Le modèle de mélange est :

$$\mathbf{x}(n) = \mathbf{G} \mathbf{a}(n) \quad (1)$$

où $\mathbf{x}(n)$ est le vecteur $(N, 1)$ des observations, $\mathbf{a}(n)$ le vecteur $(N, 1)$ des sources, n le temps discret et \mathbf{G} une matrice carrée (N, N) *invertible* appelée matrice de mélange. Le but étant de restituer les sources, il s'agit d'estimer une matrice de séparation \mathbf{H} agissant sur les observations \mathbf{x} :

$$\mathbf{y}(n) = \mathbf{H} \mathbf{x}(n) \quad (2)$$

de tel sorte que les composantes du vecteur $(N, 1)$ $\mathbf{y}(n)$ correspondent aux différentes sources $a_i(n)$, $i \in \{1, \dots, N\}$. On définit la matrice \mathbf{S} du système global par :

$$\mathbf{S} \triangleq \mathbf{H} \mathbf{G}, \quad (3)$$

et ainsi grâce à (1) et (2), on a :

$$\mathbf{y}(n) = \mathbf{S} \mathbf{a}(n). \quad (4)$$

Les signaux sources étant inobservables, il existe deux indéterminations quant à leurs restitutions. En général, la puissance

de même que l'ordre des sources ne peuvent pas être restitués. Elles seront considérées comme séparées lorsque :

$$\mathbf{S} = \mathbf{D} \mathbf{P} \quad (5)$$

où \mathbf{D} est une matrice diagonale inversible quelconque et \mathbf{P} une matrice de permutation elle aussi quelconque.

Nous faisons les hypothèses suivantes :

A1a. Les sources $a_i(n)$, $i \in \{1, \dots, N\}$, sont centrées, de puissance 1 et mutuellement indépendantes d'un point de vue statistique.

A1b. $\mathbf{a}(n)$ est un processus aléatoire stationnaire au moins à l'ordre considéré, i.e. $\forall i \in \{1, \dots, N\}$, le cumulatif $\text{Cum}(\underbrace{a_i(n), \dots, a_i(n)}_{p \times}, \underbrace{a_i^*(n), \dots, a_i^*(n)}_{q \times})$ est indépendant de

n . Il est noté $\mathbf{C}_{p,q} a_i$.

A1c. Etant donné deux entiers p and q , les cumulants des sources sont supposés satisfaire l'une des deux relations suivantes :

$$\begin{aligned} \text{c1. } & |\mathbf{C}_{p,q} a_1| \geq \dots \geq |\mathbf{C}_{p,q} a_N| > 0; \\ \text{c2. } & |\mathbf{C}_{p,q} a_1| \geq \dots \geq |\mathbf{C}_{p,q} a_{N-1}| > |\mathbf{C}_{p,q} a_N| = 0. \end{aligned}$$

En particulier, cela signifie qu'au plus un des cumulants $\mathbf{C}_{p,q} a_i$, $i \in \{1, \dots, N\}$ soit nul.

A2. \mathbf{S} est une matrice orthonormale, i.e. $\mathbf{S} \mathbf{S}^H = \mathbf{I}$ où \mathbf{I} est la matrice identité.

Il est à remarquer que les hypothèses A1c et A2 sont faites sans perte de généralité. En effet la première peut être adoptée car l'ordre des sources ne pouvant pas être restitué, nous pouvons les ranger dans un ordre quelconque et la seconde grâce à un blanchiment préalable des observations.

L'ensemble des vecteurs aléatoires satisfaisant A1a à A1c est noté \mathcal{A} . L'ensemble des matrices satisfaisant A2 est noté \mathcal{U} . Le sous ensemble de \mathcal{U} des matrices satisfaisant la relation (5) est noté \mathcal{P} et le sous ensemble de \mathcal{P} des matrices diagonales est noté \mathcal{D} . Finalement l'ensemble des vecteurs aléatoires $\mathbf{y}(n)$ déduit de (4) où $\mathbf{a}(n) \in \mathcal{A}$ et $\mathbf{S} \in \mathcal{U}$ est noté \mathcal{Y} .

3 Nouveaux contrastes

Rappelons d'abord la définition proposée dans [2].

Définition 1 Un contraste sur \mathcal{Y} est une fonction multivariable \mathcal{F} de l'ensemble \mathcal{Y} sur \mathbb{R} satisfaisant les trois propriétés suivantes :

- *1. $\forall \mathbf{y} \in \mathcal{Y}, \forall \mathbf{S} \in \mathcal{P}, \mathcal{F}(\mathbf{S} \mathbf{y}) = \mathcal{F}(\mathbf{y})$;
- *2. $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{U}, \mathcal{F}(\mathbf{S} \mathbf{a}) \leq \mathcal{F}(\mathbf{a})$;
- *3. $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{U}, \mathcal{F}(\mathbf{S} \mathbf{a}) = \mathcal{F}(\mathbf{a}) \Leftrightarrow \mathbf{S} \in \mathcal{P}$.

De tels contrastes sont des fonctions symétriques et invariants par changement d'échelle (*1) devant être maximisées (*2) afin de résoudre le problème de la séparation (*3). De nombreux contrastes définis de cette façon ont été proposés, e.g. [7]. Afin de pouvoir considérer des fonctions non symétriques nous proposons la définition suivante

Définition 2 Un contraste sur $(\mathcal{Y}, \mathcal{P}_d)$ est une fonction multivariable \mathcal{F} de l'ensemble \mathcal{Y} sur \mathbb{R} satisfaisant $\forall \mathbf{a} \in \mathcal{A}$ et $\forall \mathbf{S} \in \mathcal{P}$, les deux propriétés suivantes :

- R1. $\forall \mathbf{y} \in \mathcal{Y}, \forall \mathbf{D} \in \mathcal{D}, \mathcal{F}(\mathbf{D} \mathbf{y}) = \mathcal{F}(\mathbf{y})$;
- R2. $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{P}, \mathcal{F}(\mathbf{S} \mathbf{a}) \leq \mathcal{F}(\mathbf{a})$;
- R3. $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{P}, \exists \mathcal{P}_d \subset \mathcal{P}, \mathcal{P}_d \neq \emptyset / \mathcal{F}(\mathbf{S} \mathbf{a}) = \mathcal{F}(\mathbf{a}) \Leftrightarrow \mathbf{S} \in \mathcal{P}_d$.

Ainsi défini, un contraste n'est plus, a priori, imposé symétrique. Notons que tous les maxima globaux doivent maintenant former un sous ensemble (non vide) de l'ensemble des matrices de séparation. La maximisation d'un tel contraste n'est alors plus, en général, une condition nécessaire et suffisante de séparation mais une condition suffisante. Notons finalement que les contrastes définis au sens de la définition 1, le sont aussi (bien évidemment) au sens de la définition 2. La réciproque n'étant, en général, pas vrai comme le montrent les résultats qui suivent.

Soit la fonction :

$$\mathcal{F}_{p,q}^{\gamma}(\mathbf{y}) = \sum_{i=1}^N \gamma_i |\mathbf{C}_{p,q} y_i| \quad (6)$$

où les $\gamma_i, i \in \{1, \dots, N\}$, appartiennent à \mathbb{R}_+^* . Nous proposons le résultat suivant :

Proposition 1 Si $p + q \geq 3$ et

$$\gamma_1 \geq \dots \geq \gamma_N > 0, \quad (7)$$

alors la fonction $\mathcal{F}_{p,q}^{\gamma}(\mathbf{y})$ est un contraste au sens de la définition 2.

En raison du manque de place, la preuve est reportée dans un papier à venir [8]. Considérons maintenant le cas particulier de sources possédant un cumulatif d'ordre (p, p) de signe ε_p identique. Nous avons alors le résultat suivant :

Proposition 2 Avec (7), si $p \geq 2$ et $\forall i, \text{sgn}(\mathbf{C}_{p,p} a_i) = \varepsilon_p$ alors la fonction :

$$\mathcal{F}_{p,p}^{\gamma}(\mathbf{y}) = \varepsilon_p \sum_{i=1}^N \gamma_i \mathbf{C}_{p,p} y_i$$

est un contraste au sens de la définition 2.

4 Analyse

L'étape préalable de blanchiment des observations étant supposée réalisée, il s'agit d'estimer une matrice orthonormale \mathbf{H} de telle façon qu'un des contrastes précédents soit maximum. A cette fin, \mathbf{H} est paramétré au moyen de rotations planes :

$$\mathbf{H} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

Pour simplifier, nous allons considérer le cas de 2 sources réelles et i.i.d. et le contraste $\mathcal{F}_{2,2}^{\gamma}$. Ce dernier peut s'écrire :

$$\mathcal{F}_{2,2}^{\gamma}(\mathbf{y}) = \gamma_1 \mathcal{F}(\mathbf{y}) + C$$

où C est une constante et

$$\mathcal{F}(\mathbf{y}) = \varepsilon_2 \sum_{i=1}^N \frac{\gamma_i}{\gamma_1} \mathbf{E}[y_i^4].$$

Dans la maximisation de $\mathcal{F}'_{2,2}$ les constantes γ_1 et C ne jouant aucun rôle, nous ne considérons donc que \mathcal{F} . Pour $N = 2$, la version empirique de $\mathcal{F}(\mathbf{y})$ s'écrit :

$$\hat{\mathcal{F}}(\mathbf{y}) = \varepsilon_2 (\hat{\mathbf{E}}[y_1^4] + \delta \hat{\mathbf{E}}[y_2^4])$$

où $\delta = \gamma_2/\gamma_1$ et

$$\hat{\mathbf{E}}[y_i^4] = \frac{1}{N} \sum_{k=1}^N y_i^4(k).$$

Un maxima de $\hat{\mathcal{F}}$ doit satisfaire la condition

$$\hat{\mathcal{F}}'(\theta) = \frac{\partial \hat{\mathcal{F}}}{\partial \theta} = 0. \quad (8)$$

Supposons maintenant θ proche de la valeur $\tilde{\theta}$ (de séparation) correspondant à $y_1 = a_1$ et $y_2 = a_2$, en développant $\hat{\mathcal{F}}'$ au premier ordre on obtient :

$$\hat{\mathcal{F}}'(\theta) \approx \hat{\mathcal{F}}'(\tilde{\theta}) + (\theta - \tilde{\theta}) \hat{\mathcal{F}}''(\tilde{\theta}) \quad (9)$$

où $\hat{\mathcal{F}}''(\tilde{\theta}) = \partial \hat{\mathcal{F}}' / \partial \theta |_{\theta=\tilde{\theta}}$. En reportant (9) dans (8), on a

$$\theta - \tilde{\theta} \approx - \frac{\hat{\mathcal{F}}'(\tilde{\theta})}{\hat{\mathcal{F}}''(\tilde{\theta})}. \quad (10)$$

Après quelques calculs, il vient :

$$\hat{\mathcal{F}}''(\tilde{\theta}) = 4\varepsilon_2 (3\hat{\mathbf{E}}[a_1^2 a_2^2] - \hat{\mathbf{E}}[a_1^4] + \delta (3\hat{\mathbf{E}}[a_1^2 a_2^2] - \hat{\mathbf{E}}[a_2^4])) \quad (11)$$

que l'on va supposer égal à sa vraie valeur :

$$\hat{\mathcal{F}}''(\tilde{\theta}) \approx 4(|\mathbf{C}_{2,2} a_1| + \delta |\mathbf{C}_{2,2} a_2|). \quad (12)$$

Pour le terme $\hat{\mathcal{F}}'(\tilde{\theta})$ on a

$$\hat{\mathcal{F}}'(\tilde{\theta}) = \frac{4\varepsilon_2}{N} \sum_{k=1}^N a_1(k) a_2(k) (a_1^2(k) - \delta a_2^2(k)) \quad (13)$$

dont on peut remarquer que son espérance mathématique est nulle, i.e. l'estimée de $\tilde{\theta}$ est non biaisée. Evaluons la variance de l'erreur, i.e. $\mathbf{E}[(\theta - \tilde{\theta})^2]$, on obtient aisément

$$\mathbf{E}[(\theta - \tilde{\theta})^2] \approx \frac{1}{N} \frac{\mathbf{E}[a_1^6] + \delta^2 \mathbf{E}[a_2^6] - 2\delta \mathbf{E}[a_1^4] \mathbf{E}[a_2^4]}{(|\mathbf{C}_{2,2} a_1| + \delta |\mathbf{C}_{2,2} a_2|)^2}. \quad (14)$$

Cette variance dépend à la fois des statistiques des sources et d'un paramètre "libre" δ . On peut maintenant se demander si il existe une valeur de δ rendant minimum $\mathbf{E}[(\theta - \tilde{\theta})^2]$. La valeur de δ notée δ_o annulant la dérivée $\partial \mathbf{E}[(\theta - \tilde{\theta})^2] / \partial \delta$ vaut

$$\delta_o = \frac{\mathbf{E}[a_1^6] |\mathbf{C}_{2,2} a_2| + \mathbf{E}[a_1^4] \mathbf{E}[a_2^4] |\mathbf{C}_{2,2} a_1|}{\mathbf{E}[a_2^6] |\mathbf{C}_{2,2} a_1| + \mathbf{E}[a_1^4] \mathbf{E}[a_2^4] |\mathbf{C}_{2,2} a_2|}. \quad (15)$$

La dérivée seconde en δ_o étant positive, il s'agit bien d'un minimum.

En fonction des statistiques des sources, si $\delta_o \leq 1$ (cf. (7)), on pourra alors l'utiliser afin d'obtenir les meilleures performances dans la mesure où $\mathbf{E}[(\theta - \tilde{\theta})^2]$ sera minimum.

Prenons un premier exemple, si les deux sources possèdent les mêmes statistiques alors $\delta_o = 1$ et donc dans ce cas le

contraste symétrique est celui qui donne les meilleures performances. Maintenant supposons la seconde source Gaussienne ($\mathbf{E}[a_2^4] = 3$ et $\mathbf{E}[a_2^6] = 15$) alors $\delta_o = \mathbf{E}[a_1^4]/5$. En prenant une première source binaire, on obtient $\delta_o = 1/5 = 0.2$. Dans ce dernier cas le contraste non symétrique est le "meilleur". Pour finir, notons qu'il peut exister des cas pour lesquels $\delta_o > 1$ (e.g. lorsque $\mathbf{E}[a_1^4] > 5$ dans le cas précédent) et donc, parfois, on ne peut pas construire de contraste "optimal" de cette manière.

5 Simulations

Afin d'illustrer les résultats précédents, nous présentons quelques simulations informatiques dans le cas de deux sources ($N = 2$). Nous considérons un algorithme de gradient afin de maximiser le contraste $\mathcal{F}(\mathbf{y})$. La convergence de l'algorithme est illustrée par un indice défini grâce à la matrice globale \mathbf{S} ainsi

$$\text{ind}_\alpha(\mathbf{S}) \triangleq \frac{1}{2} \left[\sum_i \left(\sum_j \frac{|s_{ij}|^\alpha}{\max_\ell |s_{i\ell}|^\alpha} - 1 \right) + \sum_j \left(\sum_i \frac{|s_{ij}|^\alpha}{\max_\ell |s_{\ell j}|^\alpha} - 1 \right) \right] \quad (16)$$

où $\alpha \geq 1$. Cet indice positif est en effet nul si \mathbf{S} satisfait la relation (5). La matrice de mélange est considérée orthonormale de telle manière que l'étape de blanchiment n'est pas nécessaire :

$$\mathbf{G} = \begin{pmatrix} \cos \theta_1 & \sin \theta_1 \\ -\sin \theta_1 & \cos \theta_1 \end{pmatrix}, \quad \theta_1 = 20 \frac{\pi}{180}. \quad (17)$$

Nous utiliserons l'indice $\text{ind}_2(\cdot)$. La première source est prise binaire tandis que la seconde est Gaussienne. Ainsi le coefficient optimal est $\delta = 0.2$. Sur la première figure sont affichées 50 évaluations de l'angle estimé en fonction du nombre d'itérations, avec un contraste classique ($\delta = 1$) et le contraste "optimal" non symétrique. Le gain en variance d'erreur est clair. Ceci est confirmé par les figures 2 et 3 où sont représentés la moyenne et l'écart type de l'indice ind_2 en fonction des itérations.

En conclusion, il semble intéressant de tirer partie de la non symétrie d'un contraste afin d'optimiser les performances d'estimation de la matrice de séparation. Sous réserve, bien entendu, de disposer d'une connaissance a priori sur certains moments des sources.

Références

- [1] J.F. Cardoso, S. Bose and B. Friedlander, "On Optimal Source Separation Based on Second and Fourth Order Cumulants", in Proc. *IEEE SP Workshop on SSAP*, Corfu, Greece, pp 198-201, June 1996.
- [2] P. Comon, "Independent Component Analysis, a New Concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.

- [3] N. Delfosse and P. Loubaton, "Adaptive Blind Separation of Independent Sources : a Deflation Approach", *Signal Processing*, Vol. 45, pp 59-83, 1995.
- [4] C. Jutten and J. Herault, "Blind Separation of Sources, Part I : An Adaptative Algorithm Based on Neuromimetic Architecture", *Signal Processing*, Vol. 24, pp 1-10, 1991.
- [5] E. Moreau and O. Macchi, "High Order Contrasts for Self-Adaptive Source Separation", *International Journal of Adaptive Control and Signal Processing*, Vol. 10, No. 1, pp 19-46, Jan. 1996.
- [6] E. Moreau, "Criteria for Complex Sources Separation", in Proc. *EUSIPCO'96, European Signal Processing Conference*, Trieste, Italy, Vol. II, pp 931-934, September 1996.
- [7] E. Moreau and J.-C. Pesquet, "Independence / Decorrelation Measures with Applications to Optimized Orthonormal Representations", in Proc. *ICASSP'97, Int. Conf. on Acoustics, Speech and Signal Processing*, Munich, Germany, Vol. 5, pp 3425-3428, April 1997.
- [8] E. Moreau and N. Thirion, "Non Symmetrical Contrasts for Sources Separation", en préparation.

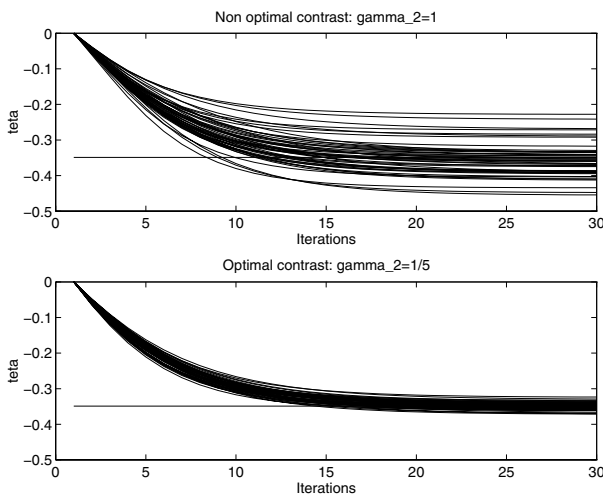


FIG. 1 — 50 évaluations de l'angle estimé en fonction du nombre d'itérations, avec un contraste classique et le contraste "optimal" non symétrique dans le cas du mélange d'une source binaire avec une source gaussienne.

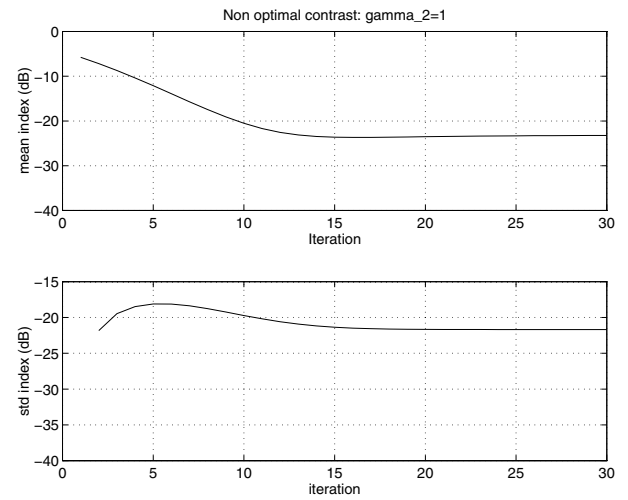


FIG. 2 — Moyenne et écart type de l'indice ind_2 avec le contraste "classique" ($\delta = 1$).

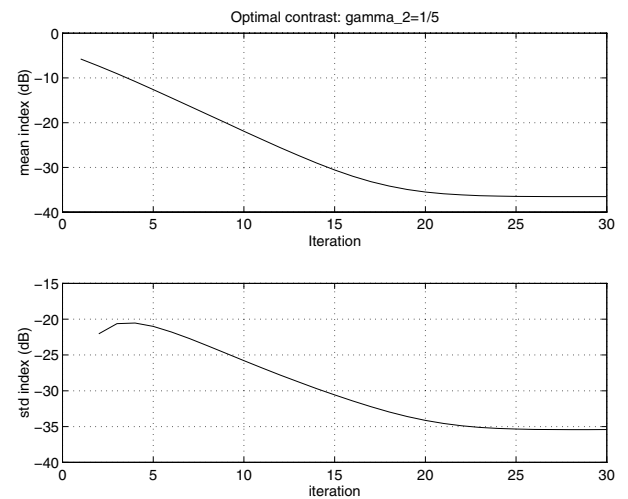


FIG. 3 — Moyenne et écart type de l'indice ind_2 avec le contraste "optimal" ($\delta = 0.2$).