ISSPA 2005, 8th International Symposium on Signal Processing and its Applications
Sydney (Australia), 28-31 August 2005, pp. 37-40.

SPURIOUS ENTROPY MINIMA FOR MULTIMODAL SOURCE SEPARATION

Dinh-Tuan Pham

Laboratoire de Modélisation et Calcul, CNRS
BP 53, 38041 Grenoble, France
E-mail: Dinh-Tuan.Pham@imag.fr

Fédréric Vrins, Michel Verleysen

UCL Machine Learning Group
Université catholique de Louvain
1348 Louvain-la-Neuve, Belgium
E-mail: {vrins,verleysen}@dice.ucl.ac.be

ABSTRACT

This paper presents two approaches for showing that spurious minima of the entropy may exist in the blind source separation context. The first one is based on the calculation of first and second derivative of the output entropy and The second one is based on entropy approximator for multimodal variable having small overlap between the modes. It is shown that spurious entropy minima arise when the source distribution becomes more and more multimodal.

1. INTRODUCTION

It is known that minimizing globally the entropy cost function under the whitening constraint leads to recover the source with the lowest entropy [6]. Nevertheless, previous works have suggested that local entropy minima exist, and correspond to spurious solutions of the blind source separation (BSS) problem (see [5] and references therein). However, such works have been based on numerical simulation and Parzen density estimation, and thus do not constitute an absolute proof. In this communication we shall provide a specific class of source distribution for which we prove rigorously (in the case of two sources and two sensors), that it does exist spurious local minima of the entropy contrast.

We shall consider the simplest and most widely used BSS model, which assumes noise-free mixtures with the same number of sensors and sources: $\mathbf{X} = \mathbf{AS}$ where $\mathbf{X} = [X_1 \quad \cdots \quad X_K]^{\mathrm{T}}$ is the observation vector, $\mathbf{S} = [S_1 \quad \cdots \quad S_K]^{\mathrm{T}}$ is the source vector and $\mathbf{A}$ is an invertible matrix; $^{\mathrm{T}}$ denotes the transposition operator. For simplicity, we shall focus on the deflation approach which extracts the sources one by one by maximizing a non Gaussianity index. A popular and statistically efficient non Gaussianity index of a random variable Y is the negentropy $J(Y)$ [8]. We consider the problem of maximizing the negentropy of $\mathbf{b}^{\mathrm{T}}\mathbf{X}$ with respect to the vector $\mathbf{b}$. Remind that the negentropy of Y is defined as $J(Y) = \frac{1}{2}\{\log[2\pi \mathrm{var}(Y)] + 1\} - H(Y)$ where $\mathrm{var}(\cdot)$ denotes the variance and $H(Y)$ denotes the entropy of Y. For a random variable Y with density p_Y, the entropy is given by [2]: $H(Y) = -\int p_Y(y) \log p_Y(y)\, dy$. Since $J(\cdot)$ is scale invariant, we may normalize $\mathbf{b}$ such that $\mathbf{b}^{\mathrm{T}}\mathbf{X}$ has a given variance. Further, we may without loss of generality assume that the sources have the same variance since one can divide any source by a constant and multiply the corresponding columns of $\mathbf{A}$ by the same constant. Thus, setting $\mathbf{w} = \mathbf{A}^{\mathrm{T}}\mathbf{b}$, we are led to consider the problem of maximizing $J(\mathbf{w}^{\mathrm{T}}\mathbf{S})$, which is the same as minimizing $H(\mathbf{w}^{\mathrm{T}}\mathbf{S})$ under the $\|\mathbf{w}\| = 1$ constraint.

We focus on the $K = 2$ case. Vector $\mathbf{w}$ can then be parameterized as $[\cos\theta \quad \sin\theta]^{\mathrm{T}}$. Let $Z_\theta = \mathbf{w}^{\mathrm{T}}\mathbf{S} = \cos\theta S_1 + \sin\theta S_2$, the problem is to minimize $H(Z_\theta)$ with respect to θ; this function is periodic with period π, hence one may restrict oneself to $\theta \in [0, \pi)$. In this paper, we focus on the possible existence of entropy local minima that do not correspond to a source but well to a mixture, that is the output entropy admits local minima for θ which are not integer multiples of $\pi/2$. The existence of such local minima will be theoretically proven when the source distribution is *multimodal enough*, as observed in [5].

Two complementary approaches are presented. In Section II, the first and second derivatives of the function $H(Z_\theta)$ are analyzed for *specific* source probability distribution function (pdf). By contrast, the approach derived in Section III relies on an *entropy approximation* when the source pdf is strongly multimodal; it is a formalization of the idea first presented in [5]. It yields more general results than the first approach, but is valid only when the overlap between the modes becomes *small enough*.

2. THE DERIVATIVE APPROACH

This approach relies on an expansion up to second order of the entropy of a random variable slightly contaminated with another variable, which has been established in [3]. Specifically, let Y be some random variable with density p_Y and δY a small (arbitrary) random increment, one has

$$H(Y + \delta Y) \approx H(Y) + \mathrm{E}[\psi_Y(Y)\delta Y] + \frac{1}{2}\{\mathrm{E}[\mathrm{var}(\delta Y|Y)\psi_Y'(Y)] - [\mathrm{E}(\delta Y|Y)]'^2\} \ ,$$

ISSPA 2005, 8th International Symposium on Signal Processing and its Applications
Sydney (Australia), 28-31 August 2005, pp. 37-40.

up to second order in δY, where $\psi_Y = (-\log p_Y)'$ is the score function of Y, $'$ denotes the derivative. As usual, $\mathrm{E}(\cdot)$ and $\mathrm{E}(\cdot|Y)$ denote the expectation and conditional expectation given Y and $\mathrm{var}(\delta Y|Y) = \mathrm{E}[(\delta Y)^2|Y] - \mathrm{E}(\delta Y|Y)]^2$ is the conditional variance of δY given Y. We apply the above result to $Y = Z_\theta$ and δY being the increment δZ_θ of Z_θ induced by an increment $\delta\theta$ of θ. Elementary calculation yields

$$\delta Z_\theta = Z_\theta^\perp \delta\theta - Z_\theta(\delta\theta)^2 + o((\delta\theta)^2)$$

where $Z_\theta^\perp = -\sin\theta S_1 + \cos\theta S_2$. Therefore, noting that $\mathrm{E}[\psi_{Z_\theta}(Z_\theta)Z_\theta] = 1$ by integration by parts, one has

$$H(Z_\theta + \delta Z_\theta) \approx H(Z_\theta) + \mathrm{E}[\psi_{Z_\theta}(Z_\theta)Z_\theta^\perp]\delta\theta + \frac{1}{2}\{\mathrm{E}[\mathrm{var}(Z_\theta^\perp|Z_\theta)\psi'_{Z_\theta}(Z_\theta)] - [\mathrm{E}(Z_\theta^\perp|Z_\theta)]'^2 - 1\}(\delta\theta)^2$$

up to second order in $\delta\theta$. Thus, we have

$$\frac{\partial H(Z_\theta)}{\partial\theta} = \mathrm{E}[\psi_{Z_\theta}(Z_\theta)Z_\theta^\perp] \quad \text{and}$$
$$\frac{\partial^2 H(Z_\theta)}{\partial\theta^2} = \mathrm{E}[\mathrm{var}(Z_\theta^\perp|Z_\theta)\psi'_{Z_\theta}(Z_\theta)] - [\mathrm{E}(Z_\theta^\perp|Z_\theta)]'^2 - 1. \tag{1}$$

The above result shows that the values of θ satisfying $\mathrm{E}[\psi_{Z_\theta}(Z_\theta)Z_\theta^\perp] = 0$ are stationary points of $H(Z_\theta)$. Clearly, this is true for $\theta = 0$ and $\theta = \pi/2$ since $Z_0 = -Z_{\pi/2}^\perp = S_1$, $Z_{\pi/2} = Z_0^\perp = S_2$ and S_1 and S_2 are independent. Actually $H(Z_\theta)$ attains a local minimum at $\theta = 0$ unless S_1 is Gaussian [1]. Indeed, the second derivative of $H(Z_\theta)$ at $\theta = 0$ reduces to $\mathrm{var}(S_2)\mathrm{E}[\psi'_{S_1}(S_1)] - 1$. But for any random variable Y, $\mathrm{E}[\psi'_Y(Y)] = \mathrm{E}[\psi_Y^2(Y)]$ by integration by parts and $\mathrm{var}(Y)\mathrm{E}[\psi_Y^2(Y)] \geq 1$ by the Schwartz inequality (noting that $\mathrm{E}[\psi_Y(Y)Y] = 1$ and $\mathrm{E}[\psi_Y(Y)] = 0$), which is strict unless ψ_Y is linear, that is Y is Gaussian. By the same argument, $H(Z_\theta)$ also attains a local minimum at $\theta = \pi/2$ unless S_2 is Gaussian.

We consider the special case where the sources have the same distribution. It can then be seen that there are two other stationary points of $H(Z_\theta)$ at $\theta = \pi/4$, for which $Z_\theta = (S_1 + S_2)/\sqrt{2}$ and $Z_\theta^\perp = (S_2 - S_1)/\sqrt{2}$, and at $\theta = 3\pi/4$, for which $Z_\theta = (S_2 - S_1)/\sqrt{2}$ and $Z_\theta^\perp = -(S_1 + S_2)/\sqrt{2}$. Indeed, since the joint distribution of (S_1, S_2) is the same as that of (S_2, S_1), one has

$$\mathrm{E}\Big[\psi_{(S_1\pm S_2)/\sqrt{2}}\Big(\frac{S_1 \pm S_2}{\sqrt{2}}\Big)\Big(\frac{S_2 \mp S_1}{\sqrt{2}}\Big)\Big] = \mathrm{E}\Big[\psi_{(S_2\pm S_1)/\sqrt{2}}\Big(\frac{S_2 \pm S_1}{\sqrt{2}}\Big)\Big(\frac{S_1 \mp S_2}{\sqrt{2}}\Big)\Big]$$

as the second right hand side is nothing else than the first with S_1 and S_2 permuted. Thus $\mathrm{E}[\psi_{Z_{\pi/4}}(Z_{\pi/4})(Z_{\pi/4}^\perp)]$ vanishes as it equals its opposite and

$$\mathrm{E}[\psi_{-Z_{3\pi/4}}(-Z_{3\pi/4})(Z_{3\pi/4}^\perp)] = \mathrm{E}[\psi_{Z_{3\pi/4}}(Z_{3\pi/4})(Z_{3\pi/4}^\perp)].$$

[1] Gaussian distribution has maximum entropy so that in this case the point $\theta = 0$ corresponds actually to a global maximum of the entropy.

But $Z_{3\pi/4} = (S_2 - S_1)/\sqrt{2}$ has the same distribution as $(S_1 - S_2)/\sqrt{2} = -Z_{3\pi/4}$ and thus its distribution is symmetric. Hence $\psi_{Z_{3\pi/4}} = \psi_{-Z_{3\pi/4}}$ and is an odd function. It follows that $\mathrm{E}[\psi_{Z_{3\pi/4}}(Z_{3\pi/4})(Z_{3\pi/4}^\perp)] = 0$.

To see if the values $\theta = \pi/4$ and $\theta = 3\pi/4$ correspond to local minima of $H(Z_\theta)$, we look at the second derivative of $H(Z_\theta)$. By permuting S_1 and S_2, we have

$$\mathrm{E}(S_2 - S_1|S_1 + S_2) = \mathrm{E}(S_1 - S_2|S_1 + S_2), \tag{2}$$

yielding $\mathrm{E}(Z_{\pi/4}^\perp|Z_{\pi/4}) = 0$. Thus the second derivative (1) of $H(Z_\theta)$ at $\theta = \pi/4$ reduces to $\mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})] - 1$, noting that [7] $\mathrm{E}[\mathrm{E}(Z_{\pi/4}^{\perp 2}|Z_{\pi/4})\psi'_{Z_{\pi/4}}(Z_{\pi/4})] = \mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})]$. If we specialize further to the case where the common distribution of S_1 and S_2 is symmetric, one may replace S_1 by $-S_1$ in the equality (2), which then yields $\mathrm{E}(Z_{3\pi/4}^\perp|Z_{3\pi/4}) = 0$. Then the second derivative (1) of $H(Z_\theta)$ at $\theta = 3\pi/4$ reduces to $\mathrm{E}[Z_{3\pi/4}^{\perp 2}\psi'_{Z_{3\pi/4}}(Z_{3\pi/4})] - 1$. Note that in this symmetric case, the pair $(Z_{\pi/4}, Z_{\pi/4}^\perp)$ has the same joint distribution as $(Z_{3\pi/4}, Z_{3\pi/4}^\perp)$ and therefore $\mathrm{E}[Z_{3\pi/4}^{\perp 2}\psi'_{Z_{3\pi/4}}(Z_{3\pi/4})] = \mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})]$.

In summary, if S_1 and S_2 have the same distribution, $H(Z_\theta)$ attains a local minimum at $\theta = \pi/4$ if and only if

$$\mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})] > 1 \ .$$

If the common distribution of S_1 and S_2 is symmetric, then $\mathrm{E}[Z_{3\pi/4}^{\perp 2}\psi'_{Z_{3\pi/4}}(Z_{3\pi/4})] = \mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})]$ and $H(Z_\theta)$ attains a local minimum at $\theta = 3\pi/4$ if and only if the above expectation is strictly greater than 1.

Inspired by simulations presented in [5], we consider a source pdf which is a mixture of two normal densities

$$p_S(s) = \{\phi[(s+\mu)/\sigma] + \phi[(s-\mu)/\sigma]\}/(2\sigma), \tag{3}$$

where $\phi(s) = \exp(-s^2/2)/\sqrt{2\pi}$ is the standard normal density. The ratio σ/μ is taken small so that (3) is bimodal. Note that the S_i are distributed as $\mu\tilde{S}_i$ where $\tilde{S}_i$ have a pdf of the form (3) with (μ, σ) replaced by $(1, \sigma/\mu)$. It then can be seen that $\mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})] = \mathrm{E}[Y_2^2\psi'_{Y_1}(Y_1)]$ where $Y_1 = \tilde{S}_1 + \tilde{S}_2$, $Y_2 = \tilde{S}_2 - \tilde{S}_1$. For this special case, the functions ψ_{Y_1} and $\mathrm{E}(Y_2^2|Y_1)$ can be obtained in closed form and the expectation term $\mathrm{E}[Y_2^2\psi'_{Y_1}(Y_1)]$ can be expressed as a simple integral and shown to tend to ∞ as σ/μ tends to 0. Detail of calculations and proof of this result are reported in a submitted paper [4].

Figure 1 illustrates the above result: $\mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})]$ is plotted versus σ/μ. One can see that when σ/μ decreases beyond the value $\tau = 0.63$ (approximately) this expectation becomes greater than 1. Figure 2 illustrates the example presented in this section (common distribution for S_1 and S_2 given by eq. (3)). One

ISSPA 2005, 8th International Symposium on Signal Processing and its Applications
Sydney (Australia), 28-31 August 2005, pp. 37-40.

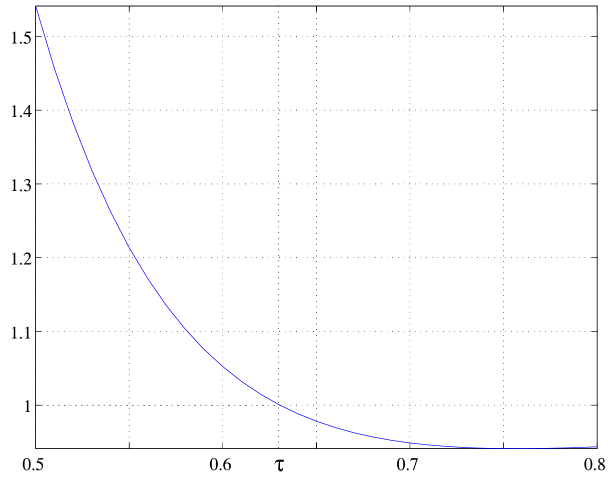

Figure 1: Plot of $\mathrm{E}[Z_{\pi/4}^{\perp 2}\psi'_{Z_{\pi/4}}(Z_{\pi/4})]$ versus σ/μ.

can remark that spurious entropy minima exist if the ratio σ/μ is lower than τ. On the contrary, no spurious entropy minima can be observed if $\sigma/\mu > \tau$.

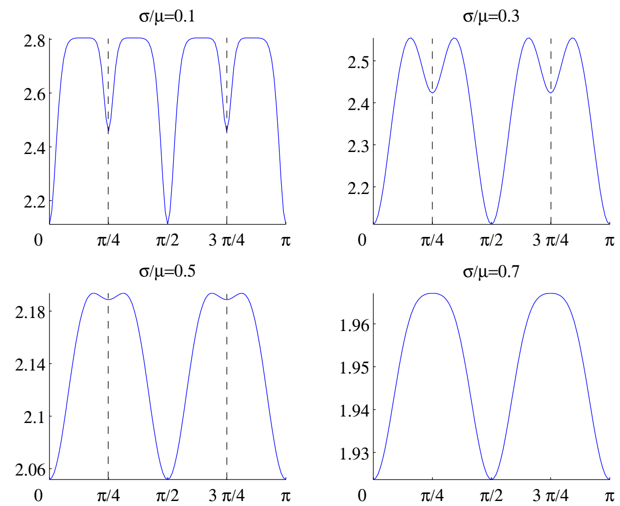

Figure 2: Plot of $H(Z_\theta) - \log(\sigma)$ versus θ for several values of the ratio σ/μ. Vertical dashed lines indicate spurious minima.

3. THE ENTROPY APPROXIMATION APPROACH

This approach relies on an entropy approximation of a strongly multimodal pdf, of the form

$$p_Y(y) = \sum_{n=1}^{N} \pi_n K_n(y) \ , \tag{4}$$

where $\pi_1, \ldots, \pi_N$ are probabilities and $K_1, \ldots, K_N$ are pdfs with "nearly disjoint" supports. Due to space limitation, proofs of results are omitted and will be report in a future paper.

Lemma 1 *Let p_Y be given by (4). Then*

$$H(p_Y) \leq h(\pi) + \sum_{n=1}^{N} \pi_n H(K_n) \ , \tag{5}$$

where $H(p_Y)$ denotes (by abuse of notation) the entropy of a random variable with density p_Y and $h(\pi) = -\sum_{i=1}^{N} \pi_i \log \pi_i$ is the entropy of a discrete random variable taking N distinct values with probabilities $\pi_1, \ldots, \pi_N$. In addition, assume that $\sup K_n = \sup_{y\in\mathbb{R}} K_n(y) < \infty$ ($1 \leq n \leq N$) and let $\Omega_1, \ldots, \Omega_N$ be disjoint subsets which approximately cover the supports of $K_1, \ldots, K_N$, such that

$$\epsilon_n = \int_{\mathbb{R}\setminus\Omega_n} K_n(y)dy$$

and

$$\epsilon'_n = \int_{\mathbb{R}\setminus\Omega_n} K_n(y) \log \frac{\sup K_n}{K_n(y)} dy$$

are small. Then, we also have

$$H(p_Y) \geq h(\pi) + \sum_{n=1}^{N} \pi_n H(K_n) - \sum_{n=1}^{N} \pi_n \epsilon'_n - \sum_{n=1}^{N} \pi_n \Big[\log \Big(\frac{\max_{1\leq n\leq N} \sup K_n}{\pi_n \sup K_n} \Big) + 1 \Big] \epsilon_n. \tag{6}$$

We are interested in the case where the densities K_n in (4) are of the form

$$K_n(y) = (1/\sigma) K[(y - \mu_n)/\sigma] \tag{7}$$

where K is a bounded density of finite entropy. The parameter σ is taken small with respect to

$$d = \min_{m\neq l} |\mu_m - \mu_l| \tag{8}$$

so that the density (4) is multimodal. Taking Ω_n in the above Lemma to be an interval centered at μ_n of length d, it results that the ϵ_n and ϵ'_n do not depend on n and are given by

$$\int_{|x|\geq d/(2\sigma)} K(x)dx \quad \text{and} \quad \int_{|x|\geq d/(2\sigma)} K(x) \log \frac{\sup K}{K(x)} dx$$

respectively. Thus one gets the following corollary.

Corollary 1 *Let p_Y be given by (4) with K_n of the form (7). Then with the same notation as in Lemma 1 and d given by (8), $H(p_Y) - \log\sigma$ is bounded above by $h(\pi) + H(K)$ and converges to this bound as $\sigma/d \to 0$.*

ISSPA 2005, 8th International Symposium on Signal Processing and its Applications
Sydney (Australia), 28-31 August 2005, pp. 37-40.

Let us return to function $H(Z_\theta)$. We consider the case where the sources S_1 and S_2 have densities p_{S_1} and p_{S_2} which are mixtures of normal densities with the same variance σ. One may model S_i as $U_i+\sigma V_i$ where U_1, U_2, V_1, V_2 are independent random variables, U_1 and U_2 are discrete and V_1 and V_2 are standard normal variables. Hence Z_θ is distributed as $U(\theta)+\sigma V$ where $U(\theta)=\cos\theta U_1+\sin\theta U_2$ and V is a standard normal variable independent of U_1 and U_2. The density p_{Z_θ} is thus of the form described by (4) and (7) with $K=\phi$, the standard normal density. Therefore, from Corollary 1, $H(Z_\theta)\simeq h[U(\theta)]+\log\sigma+H(\phi)$ for small σ, where $h[U(\theta)]$ is the discrete entropy of $U(\theta)$. Since the discrete entropy depends only on the probabilities associated to the values of the random variable, but not on these values themselves, one can prove:

Lemma 2 *Let $U(\theta)=\cos\theta U_1+\sin\theta U_2$ where (U_1,U_2) is a pair of discrete random variables taking a finite number of values. Then $h[U(\theta)]=h(U_1,U_2)$ except for a finite number of values of θ for which $h[U(\theta)]<h(U_1,U_2)$.*

The above result shows that $h[U(\theta)]$ is essentially a constant function of θ, except for some jumps downward at certain values of θ. Since $H(Z_\theta)$ may be approximated, for small σ, by $h[U(\theta)]$ plus a term not depending on θ, one may expect, for σ small enough, that $H(Z_\theta)$ admits local minima near the points of downward jumps of $h[U(\theta)]$.

For instance, consider again the source pdf (3): in this example U_1 and U_2 take values in $\{-1,1\}$ with probabilities $\{1/2,1/2\}$. Then, provided that $\theta\notin\{k\pi/4|k\in Z\}$, $U(\theta)$ has 4 distinct values with probabilities equal to 1/4: $h[U(\theta)]=\log(4)$. If $\theta\in\{\pi/4,3\pi/4\}$, $U(\theta)$ takes only 3 distinct values with probabilities $\{1/4,1/2,1/4\}$: $h[U(\theta)]=(\log(2)+\log(4))/2<\log(4)$. Finally, if $\theta\in\{0,\pi/2\}$, $U(\theta)$ takes 2 distinct values, with probabilities equal to 1/2, and in this case, $h[U(\theta)]=\log(2)<\log(4)$.

Lemma 3 *Let Z_θ be distributed as $U(\theta)+\sigma V$ where $U(\theta)$ is as in Lemma 2 and V is a standard normal variable independent of (U_2,U_2). Let $\theta_1,\ldots,\theta_p$ be the values of $\theta\in[0,\pi)$ for which $h(U(\theta))<h(U_2,U_2)$. Then for σ sufficiently small $H(Z_\theta)$ admits p local minima in $[0,\pi)$, which converge respectively to $\theta_1,\ldots,\theta_p$, as $\sigma\to 0$.*

The above Lemma shows the existence of spurious entropy minima when the Gaussian modes overlap becomes negligible: this overlap depends on the ratio of the intermodal distances to the mode standard deviation σ, as can be seen from Corollary 1. Figure 3 shows such minima when σ is *small enough*. In this approach, the term *small enough* remains vague since no numerical threshold τ is available.

4. CONCLUSION

In this paper, two different arguments are used to prove that spurious minima of entropy may exist in the blind source separation context when dealing with source pdfs that are "multimodal enough". The first approach, that uses a series development of the entropy, allows one to precise the terms "multimodal enough" in terms of expectation and score functions, but is only applicable for the case of same (and symmetric) source pdfs. By contrast, in the second approach, the terms "multimodal enough" remain vague and a "threshold" cannot be determined. But this method is more general in the sense that it can deal with asymetric and different source pdfs.

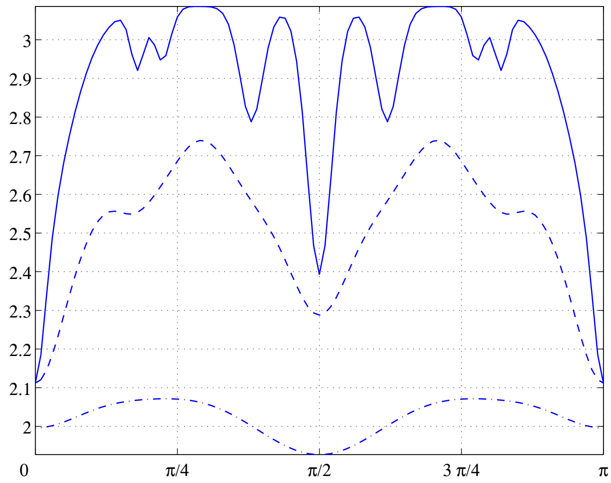

Figure 3: Plot of $H(Z_\theta)-\log(\sigma)$ for $\sigma=0.2$ (solid), $\sigma=0.6$ (dashed) and $\sigma=1.5$ (dash-dotted); $U_1=\{-\sqrt{103}/4,\sqrt{103}/4\}$ with probabilities $[1/2,1/2]$ and $U_2=\{0,2,8\}$ with probabilities $[1/2,3/8,1/8]$.

5. REFERENCES

[1] Comon, P. "Independent component analysis – a new concept?," *Signal Processing*, **36**, pp. 287–314, 1994.

[2] Cover, Th. and Thomas, J. A. *Elements of Information Theory*. New-York: Wiley, 1991

[3] Pham, D. T. "Entropy of a Variable Slightly Contaminated With Another." *IEEE Sig. Proc. Lett.*, To appear.

[4] Pham, D. T. and Vrins, F. "Local minima of information-theoretic criteria in blind source separation" Submitted to *IEEE Sig. Proc. Lett.*.

[5] Vrins, F. Lee, J.A. and Verleysen, M. "Can we always trust entropy minima in the ICA context ?" submitted to *EUSIPCO 2005*.

[6] Cruces, S. Cichocki, A. and Amari, S.,"The minimum entropy and cumulants based contrast functions for blind source extraction." *LNCS 2085*, 786–793, 2001.

[7] Gray, R. and Davisson, L. "An Introduction to Statistical Signal Processing." *Cambridge University Press*, p. 208, 2004.

[8] Hyvärinen, A. "Survey on independent component analysis." *Neural Computing Surveys*, **2**, 94–128, 1999.