Spectral Envelope Estimation using a Penalized Likelihood Criterion

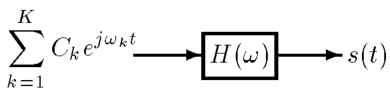
Olivier Cappé, Marine Oudot and Eric Moulines

ENST Département Signal / CNRS URA 820
46 rue Barrault, 75634 Paris Cedex 13, France[*]
email: `cappe@sig.enst.fr`

ABSTRACT

Finding a smooth spectral envelope that connects estimated sinusoids is a topic of major importance in audio signal processing. In this paper, a penalized likelihood criterion is introduced for the estimation of the spectral envelope in the presence of measurement noise. Various simulation results are presented that highlight the efficiency of the proposed performance criterion.

1. Introduction



The model above (filter excited by sinusoidal source) is perhaps the most widespread representation in audio. If the modulus of the C_k can be considered as constant in the frequency range of $H(\omega)$, the magnitude frequency response $|H(\omega)|$ is identifiable up to a gain factor. In the following, we should refer to $S(\omega) = |H(\omega)|^2$ as the spectral envelope.

From a signal processing point of view, estimating the parameters of such a model from the measured signal $s(t)$ is a difficult inverse problem because the information is reduced to the values of the spectral envelope at frequencies ω_k. Moreover, it is more realistic to consider that in practical situations the signal $s(t)$ itself is non-observable, and that only a noise corrupted version of the signal is available. This notion of noise is particularly important since it accounts both for the limitations of the measurement device and for the fact that the model above is not 100% correct! While the first factor can often be neglected, the second one always plays a role to some extent.

Several methods have been proposed to estimate the spectral envelope: [1] relies on a non parametric envelope model, while the techniques developed in [2] and [3] are based on the most usual speech envelope parameterizations (as an all-pole filter or with cepstral coefficients). In practice however, these techniques are often plagued with poor conditioning or instability problems due to the ill-posed nature of the model. This stems from the fact that it is not possible to estimate the spectral envelope from a discrete set of values only without imposing some constraints on the envelope. Usually, these constraints are entirely fixed by the parameterization, for instance by using a low order all pole model. In [4], we proposed a more explicit solution to this problem based on the use of a function of the envelope (the so-called penalty) which formalizes our a priori knowledge of the envelope behavior. We however missed an opportunity to point out that problems similar to that considered here do receive extensive coverage in the statistical literature on smoothing (splines, kernels, etc) [5], [6].

In this paper, we develop a performance criterion for assessing the goodness of fit of the envelope which is motivated by the statistical analysis of the model under consideration. This criterion is obtained as a large sample approximation of the likelihood for the estimated values of $|H(\omega_k)|^2$. The envelope itself is estimated by numerical optimization of a penalized version of the criterion. Simulations shows that this new method largely outperforms the one proposed in [4] when considering the average log spectral distortion to the true (unknown) spectral envelope.

2. Likelihood criterion

Switching to a phase/quadrature representation, we consider that the observed data consists of

$$x(t) = \sum_{k=1}^{K} [a_k \cos \omega_k t + b_k \sin \omega_k t] + n(t) \quad (1 \le t \le T) \tag{1}$$

where $\omega_k, \ldots, \omega_K$ are the frequencies of the sinusoids. The noise $n(t)$ is modeled as a stationary random process with power spectral density $\Pi_n(\omega)$. We assume that the frequencies of the sinusoids have been measured beforehand and can be considered as "exact" (this is usually the case, at least for harmonic models).

The amplitude of the phase and quadrature components of the sinusoids (a_k and b_k) can be estimated ei-

* This work was supported by the Centre National d'Etude des Télécommunications (CNET), under contract 95-PE-7697.

ther via the Fourier transform or with a time-domain least squares criterion. Under standard mixing assumptions on $n(t)$, it is well known that when the size T of the analysis window is sufficiently large: *(i)* Both approaches become equivalent; *(2)* The estimated values $\hat{a}_k$ et $\hat{b}_k$ are asymptotically independent and normally distributed with mean a_k and b_k, respectively, and variance $n_k/2$, where $n_k = 4\Pi_n(\omega_k)/T$ is the apparent noise level at frequency ω_k. We use the term *apparent* to distinguish between the noise level n_k as can measured on the Fourier transform of $x(1), \ldots, x(T)$[1] and the PSD $\Pi_n(\omega_k)$ which is independent of T.

The estimated squared magnitude of the kth sinusoid $x_k = (\hat{a}_k)^2 + (\hat{b}_k)^2$ is distributed as a scaled non-central χ^2 distribution with two degrees of freedom, whose probability density is given by

$$p(x_k) = \frac{1}{n_k} \exp[-\frac{s_k + x_k}{n_k}] \, I_0 \left(2\sqrt{\frac{s_k x_k}{n_k^2}} \right) \quad (2)$$

where $s_k = a_k^2 + b_k^2$ is the true value of squared magnitude of the kth sinusoid and $I_0()$ stands for the modified Bessel functions of the first kind and order 0. The name "non-central χ^2" corresponds to the fact that we are considering the distribution of the sum of squarred Gaussian random variables with non-zero means. Surprisingly, the obtained distribution only depends on the sum of the squared means (non-centrality parameter) and the number of Gaussian variables (degrees of freedom: 2 in the present case). The compact expression with the modified Bessel function only holds when the number of degrees of freedom is 2 [7] (for a much more down-to-hearth derivation of (2) see [8]).

With the independence assumption, the negative log-likelihood of the K amplitude estimates $\mathcal{L}(x_{1,\ldots,K}|S)$ can be written as

$$\sum_{k=1}^{K} \left[\log n_k + \frac{s_k + x_k}{n_k} - \log I_0 \left(2\sqrt{\frac{s_k x_k}{n_k^2}} \right) \right] \quad (3)$$

where the notation S refers to the spectral envelope and $s_k = S(\omega_k)$.

In order to guarantee the smoothness of the estimated envelope, we use the same type of penalty as in [4], [6]:

$$\mathcal{R}(S) = \int_{-\pi}^{\pi} \left[\frac{d \log S(\omega)}{d\omega} \right]^2 d\omega \quad (4)$$

The envelope S is estimated by minimizing the penalized likelihood criterion

$$\mathcal{L}_\lambda(S|x_{1,\ldots,K}) = \mathcal{L}(x_{1,\ldots,K}|S) + \lambda \mathcal{R}(S)$$

[1] In practice, the asymptotic behavior is reached more rapidly when using a tapering window $w(t)$. In this case, n_k should be scaled by $(1/T \sum w^2(t))/(1/T \sum w(t))^2$.

where λ is a smoothing parameter which controls the balance between fitting the observed data $x_1, \ldots, x_K$ and satisfying the regularity conditions corresponding to $\mathcal{R}(S)$.

The above notation suggests an interpretation in terms of Bayesian maximum a posteriori estimation where $\mathcal{R}(S)$ plays the role of the prior. When the envelope S is parameterized with cepstral coefficients, the parallel goes one step further since $\mathcal{R}(S)$ becomes a quadratic form [4], [6], which means that we are assuming a Gaussian prior on the cepstral coefficients (with zero mean and variance decreasing as $1/n^2$ where n is the index of the cepstral coefficient).

3. Optimization of the penalized likelihood criterion

A full discussion of the numerical techniques suitable for minimizing $\mathcal{L}_\lambda(S|x_{1,\ldots,K})$ would be beyond the scope of the present paper. We thus just state the most important points (omitting their justification):

- Whatever parameterization is used for S, both the gradient and the Hessian of $\mathcal{L}(x_{1,\ldots,K}|S)$ can be derived analytically. Computing the gradient and the Hessian only requires the evaluation of the special functions $I_0(u)$ and $I_1(u)/I_0(u)$. in practice, these functions are evaluated using the approximations given in [9] for large u and are tabulated for low values of u.
- In general, $\mathcal{L}(x_{1,\ldots,K}|S)$ does not correspond to a convex criterion so that $\mathcal{L}_\lambda(S|x_{1,\ldots,K})$ must be optimized using a quasi-Newton algorithm with embeded one-dimensional line searches.
- When S is parameterized with cepstral coefficients, the Hessian of $\mathcal{L}(x_{1,\ldots,K}|S)$ is positive definitive in the neighborhood of the true value S_0 (with probability tending to 1 as the noise level decreases). Thus, there are no significant local minima when minimizing $\mathcal{L}_\lambda(S|x_{1,\ldots,K})$ as long as the algorithm is started from a plausible value of S (ie. one sufficiently close to S_0).
- For the all-pole parameterization of S, the above argument does not hold anymore and in practice, the criterion $\mathcal{L}(x_{1,\ldots,K}|S)$ seems to be less regular. Moreover, the exact calculation of the penalty $\mathcal{R}(S)$ is not feasible. We did however obtain satisfying results with an approximate version of $\mathcal{R}(S)$ based on the discretization of the integral in (4). The obtained envelopes can be made free of the "overshooting" problems traditionally observed with the all-pole parameterization [2].

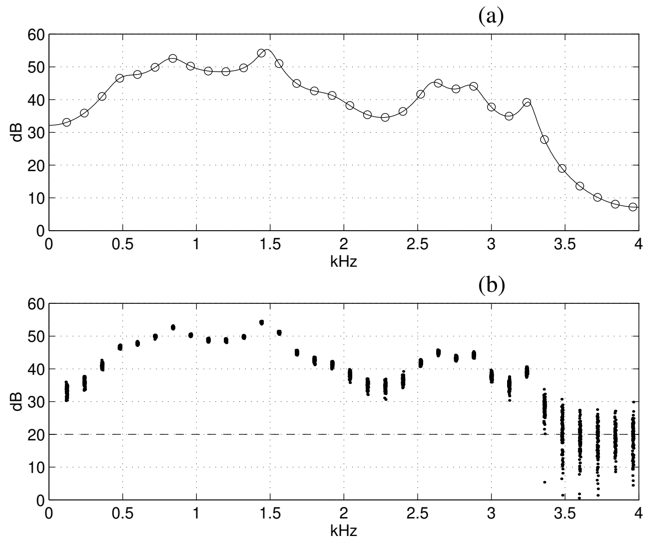


Figure 1: **(a)** Unknown spectral envelope $S_0(\omega)$ (solid line) with the values at the harmonic frequencies ω_k (circles). **(b)** Scatterplot of the measured envelope values x_k for an apparent noise level n of 20 dB (100 draws of the additive noise).

4. Simulation results

In this section, we consider the results of the proposed procedure (referred to as PLE, for Penalized Likelihood Estimation) when using the cepstral parameterization of the envelope S:

$$S(\omega_k) = s_k = \exp[c_0 + 2\sum_{n=1}^{p} c_n \cos \omega_k n]$$

with $p = K - 1$ (which means that without smoothing the estimated envelope exactly fits the measured values).

The Regularized Discrete Cepstrum (RDC) method of [4] is based on the minimization of the following criterion:

$$\sum_{k=1}^{K} (\log x_k - \log s_k)^2 + \rho \mathcal{R}(S) \qquad (5)$$

which corresponds to a (penalized) least-squares regression in the log-spectral domain. The RDC method is computationally attractive because (5) reduces to a linear regression problem with respect to the cepstral parameters c_n.

For the simulations, we use a synthetic harmonic signal obtained from the spectral envelope shown on figure 1-(a). For the sake of simplicity, we consider the case where the additive noise is white and we denote the apparent noise level by n since it does not depend any more on the frequency ω_k. Fig. 1-(b) shows the values of x_k estimated for 100 successive realizations of the noise (a small random jitter of $\pm$ 5 Hz and ± 0.2 dB has been added to each measured value to give an idea of the distribution of the points). This figures clearly shows that the RDC least-squares criterion misses the point because precisely fitting the measured envelope values is not necessarily a good idea: Values that are well above the apparent noise level are indeed reliable whereas the lowest ones exhibit erratic variations of high amplitude [8].

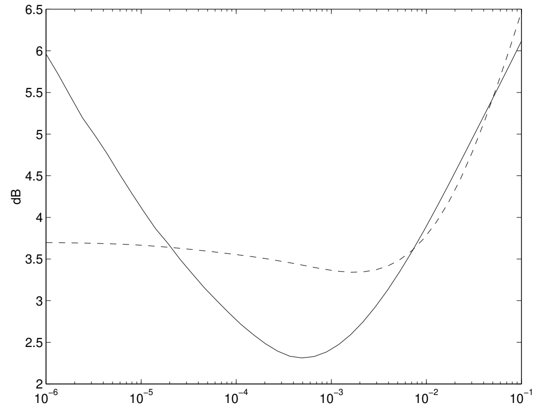


Figure 2: $\mathcal{D}(S_0, \hat{S})$ as a function of the smoothing parameters for $n = 20$ dB. Solid line: PLE (abscissa is λ). Dashed line: RDC (abscissa is ρ). Both curves are averaged over 50 simulations.

Comparing the two procedures is a bit tricky since both of them include tuning parameters (λ and ρ) which modify to some extent the estimated envelope. The automating determination of this type of smoothing parameters is a question of much interest which has yet received no simple and definitive answer [5], [6]. In the present case, we can circumvent the problem since we know what the estimated envelope should look like: λ and ρ can be fixed by minimizing a suitable distance criterion between the original envelope S_0 and the estimated envelope $\hat{S}$. We selected a criterion of the form:

$$\mathcal{D}^2(S_0, \hat{S}) = \frac{1}{2\pi} \int_{\pi}^{\pi} \left(\log |S(\omega)| - \log |\hat{S}(\omega)| \right)^2 d\omega$$

which can be computed through its cepstral domain equivalent (squared distance between the cepstral parameters of both envelopes). Note that from a statistical point of view, this criterion is not sensible since it gives the same importance to all the spectral errors regardless of the local signal to noise ratio $S_0(\omega)/n$. For a potential user of the method however, this is perhaps the most natural criterion.

Fig. 2 shows the average distortion values attained by the two methods: For values of λ between $2e-5$ and $2e-3$ the proposed method yields an average error that is uniformly lower than that obtained with the RDC for all values of ρ. In theory, one would expect the optimal value of ρ to increase with the noise level (more noise necessitates more smoothing) [5] whereas the optimal value of λ should remain constant (considering the Bayesian interpretation).

Noise level n (dB)	10	20	30	40
RDC, ρ opt.	1.6	3.4	6.6	10.9
PLE, $\lambda = 5e-4$	1.5	2.3	5.3	9.1

Table 1: Average values of the distortion to the true envelope (in dB) for different levels of the apparent noise.

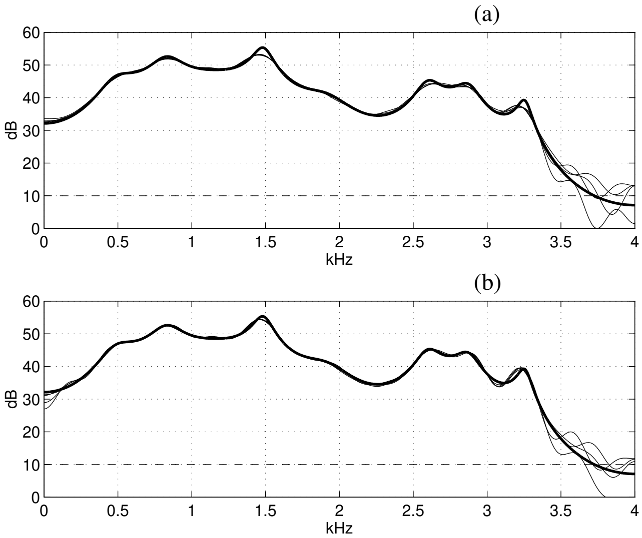


Figure 3: Actual envelope (strong line) and 4 estimated envelopes for $n = 10$ dB. **(a)** RDC, **(b)** PLE.

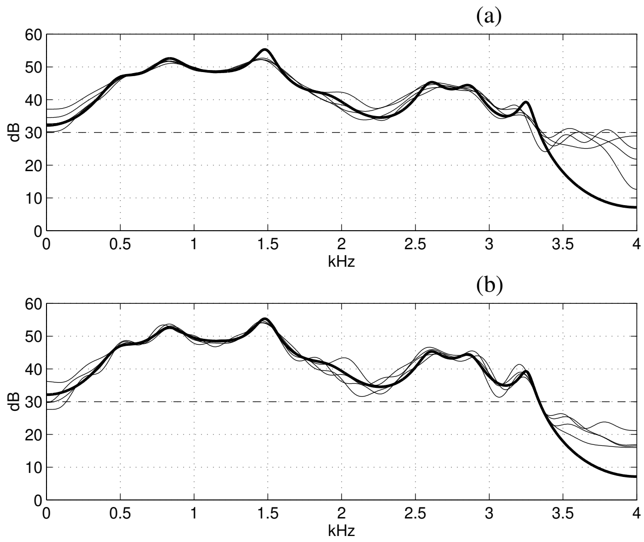


Figure 4: Actual envelope (strong line) and 4 estimated envelopes for $n = 30$ dB. **(a)** RDC, **(b)** PLE.

In practice, the behavior of both curves when varying the noise level was not so clear, but the situation was comparable to that of figure 2 in the sense that the results shown on table 1 indicate that the PLE with a fixed value of $\lambda = 5e-4$ outperforms the RDC with optimized ρ. The difference between both methods tends to decrease with the noise level. Perhaps more significant are the shapes of the estimated envelopes shown on figs. 3 and 4 for two different noise levels. Compared to the RDC, the PLE produces envelopes that are at the same time closer to the true envelope (look in particular at the peaks located at 1 and 3.25 kHz), and less influenced by the erratic fluctuations due to the noise component.

5. Conclusion

The PLE procedure raises a number of important questions which should be considered for future work: The estimation of the noise characteristics and the tuning of the smoothing parameter are challenging topics; Reducing the computation associated with the PLE is also a concern of great practical interest.

References

1. D. Paul, "The spectral envelope estimation vocoder," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 29, pp. 786–7941, august 1981.
2. A. El-Jaroudi and J. Makhoul, "Discrete all pole modeling," *IEEE Trans. Signal Processing*, vol. 39, pp. 411–423, Feb. 1991.
3. T. Galas and X. Rodet, "Generalized functional approximation for source-filter system modeling," in *Proc. Eurospeech*, (Genova), pp. 1085–1088, 1991.
4. O. Cappé and E. Moulines, "Regularization techniques for discrete cepstrum estimation," *IEEE Signal Processing Letters*, vol. 3, pp. 100–102, Apr. 1996.
5. F. O'Sullivan, "A statistical perspective on ill-posed inverse problems," *Statistical Science*, vol. 1, no. 4, pp. 502–518, 1986.
6. Y. Pawitan and F. O'Sullivan, "Non parametric spectral density estimation using penalized whittle likelihood," *Journal of the American Statistical Association*, vol. 89, pp. 600–610, june 1994.
7. N. L. Johnson and S. Kotz, *Continuous Univariate Distributions*, vol. 2. Wiley-Interscience, 1970.
8. O. Cappé and J. Laroche, "Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings," *IEEE Trans. Speech and Audio Processing*, vol. 3, Jan. 1995.
9. M. Abramowitz and I. A. Stegun, eds., *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Applied Mathematics Series, 55, National Bureau of Standards, 1964.