Approximate Optimal Periodogram Smoothing for Cepstrum Estimation Using a Penalty Term

Johan Sandberg and Maria Hansson-Sandsten

Mathematical Statistics
Centre for Mathematical Sciences
Lund University

This work was supported by the Swedish Research Council.
Let \( \{x(t), t=1, \ldots, n\} \) denote a weakly stationary random process with zero mean. We define the covariance function and the spectrum by:

\[
    r(\tau) = \mathbb{E} [x(t)x(t + \tau)]
\]

\[
    S(p) = \sum_{\tau=-n+1}^{n-1} r(\tau) e^{-i2\pi \frac{p}{N} \tau} \quad N = 2n - 1
\]
Assume that $S(p) > 0 \ \forall p$. Then the cepstrum is defined by:

$$c(q) = \frac{1}{N} \sum_{p=-n+1}^{n-1} \log (S(p)) e^{i2\pi \frac{p}{N} q}$$
Estimates of the cepstrum may be used in:

- Speech and speaker recognition
- Other audio classification applications, e.g. music genre classification
- Other classification problems, e.g. detecting root fungi in trees
The most common cepstrum estimator is based on the periodogram:

\[ \hat{c}_{\text{per}}(q) = \frac{1}{N} \sum_{p=-n+1}^{n-1} \log \left( \hat{S}_{\text{per}}(p) \right) e^{i2\pi \frac{p}{N} q} \]

where \( S_{\text{per}} \) is the periodogram:

\[ \hat{S}_{\text{per}}(p) = \frac{1}{n} \left| \sum_{t=1}^{n} x(t)e^{-i2\pi t \frac{p}{N}} \right|^2 \]
Our Class of Estimators

In this work, we will consider the following class of cepstrum estimators:

\[
\hat{c}_W(q) = \frac{1}{N} \sum_{p=-n+1}^{n-1} \log \left( \hat{S}_W(p) \right) e^{i2\pi \frac{p}{N} q}
\]

where \( S_W \) is the smoothed periodogram:

\[
\hat{S}_W(p) = \sum_{p'=-n+1}^{n-1} W(p') \hat{S}_{\text{per}}(p - p')
\]

and \( W \) is a smoothing kernel function.
Our aim is to find the lower mean square error (MSE) bound for this family of estimators. Additionally, we would like to find the estimator which gives the lowest MSE. Note: this will depend on the probability distribution of the random process.
Equivalently, we would like to solve, for a fixed $q$:

$$W_{q-opt} = \arg \min_{W\in\{-n+1,...,n-1\}\mapsto\mathbb{R}} F(W)$$

with

$$F(W) \triangleq \mathbb{E} \left[ (c(q) - \hat{c}_W(q))^2 \right]$$

A closed form solution has not been discovered.
The cost function can also be written:

\[
F(W) = \frac{1}{N^2} \mathbb{E} \left[ \left( \sum_{p=-n+1}^{n-1} \log \left( \frac{S(p)}{\hat{S}_W(p)} \right) e^{i2\pi \frac{p}{N} q} \right)^2 \right]
\]

Minimization can be solved using the approximation:

\[
\log(z) \approx 1 - z^{-1}, \quad \text{for } z \approx 1
\]
Approximative solution

New cost function with penalty term:

\[
G(W) \triangleq E \left[ (1 - \rho)n (c(q) - \hat{c}_W(q))^2 \right.
\]

\[
+ \rho \frac{1}{N} \sum_{p=-n+1}^{n-1} \left( \frac{S(p) - \hat{S}_W(p)}{S(p)} \right)^2 \left] \right.
\]
The approximation is more accurate for the new cost function. Under the approximation the optimization can be solved analytically:

\[ W_{q, \rho}^{\text{opt}} = \mathcal{F} (\Psi^{-1} \Phi) \]
Approximative solution

Some details:

\[ A_p = \begin{bmatrix} \sum_{t=1}^{n} x(t)^2 \\ 2\Re \left\{ \sum_{t=1}^{n-1} x(t)x(t+1)e^{-i2\pi \frac{p}{N}} \right\} \\ 2\Re \left\{ \sum_{t=1}^{n-2} x(t)x(t+2)e^{-i2\pi \frac{p}{N}} \right\} \\ \vdots \\ 2\Re \left\{ \sum_{t=1}^{n-(n-1)} x(t)x(t+(n-1))e^{-i2\pi \frac{p}{N}} \right\} \end{bmatrix} \]
Approximative solution

Some details:

\[ Y = \sum_{p=-n+1}^{n-1} \frac{1}{S(p)} A_p e^{i2\pi \frac{p}{N} q} \]

\[ M = \mathbf{E} [YY^T] \]

\[ \Psi = \frac{1 - \rho}{N^2} nM + \rho \frac{1}{N} \sum_{p=-n+1}^{n-1} \frac{\mathbf{E} [A_p A_p^T]}{S(p)^2} \]

\[ \Phi = \rho \frac{1}{N} \sum_{p=-n+1}^{n-1} \frac{\mathbf{E} [A_p^T]}{S(p)} \]
Based on experiments where we have changed the parameters, we have shown how the sample covariance function, estimated from speech signals, can be used to compute a good smoothing kernel for cepstrum estimation. For the solution to be accurate, the smoothing kernel must be computed as in (10), for different values of \( c \) and \( \hat{c} \) of the smoothing kernel function. The approximative solution that we present also includes a penalty term. Fig. 2 shows Monte Carlo computed MSE as a function of \( c \) and \( \hat{c} \).
Based on this, we can propose that AR-processes with parameters estimated from speech signals will provide better performance. This is expected, as the cepstrum estimator \( \hat{c} \) is computed as in (10), behaves on a set of AR(2) processes with poles in the unit circle. A simple example of a data driven approach for cepstrum estimation is shown in Figure 1.

We will now study how the cepstrum estimator \( \hat{c} \) is computed as a function of \( \rho \), and of \( \hat{c} \). The MSE of the cepstrum estimator much. This is expected, as the periodogram is convolved with the smoothing kernel function. On a new realization from the covariance function, we compute the approximate optimal smoothing kernel function. Using this sample covariance functions, we can compute which will make \( \hat{c} \) be very important. The covariance structure \( c \) is computed as in (10), is also shown. The spectrum with \( \omega = 0 \), \( \omega = 0.5 \), \( \omega = 0.1 \), and for the current type of data. The covariance structure \( c \) is not very crucial. The optimal smoothing kernel depends on certain parameters.

The weighting of the penalty term. Fig. 2 shows Monte Carlo computed (2000 simulations) MSE for different cepstrum coefficients. The MSE of three estimators: \( \hat{c} \), \( \hat{c}_{\text{per}} \), and \( \hat{c}_{\text{han}} \). We will conclude that this parameter does not affect enough to provide input to (10), by which a good smoother can be plugged into (10) in order to compute a suitable smoothing kernel function. On a new realization from the covariance function, we compute the approximate optimal periodogram smoothing kernel function. The spectrum with 98 for different cepstrum coefficients. The MSE has considerably lower MSE than the other estimators. We choose the smoothing kernel function. The exact choice does not seem to be very important.
The MSE optimal estimator within our family will depend on the true probability distribution of the random process.

Robustness experiment: Given $K$ independent realizations of a Gaussian process, use their average sample covariance functions to find the MSE optimal estimator.
Robustness Experiment

Average over $q=1, \ldots, 15$

- $\hat{C}_W$
- $\hat{C}_{\text{per}}$
- $\hat{C}_{\text{han}}$

MSE vs $K$ for $q=1, \ldots, 15$

$\hat{C}_W$ has considerably lower MSE than the other estimators.

Robustness Experiment

Thank you!