PROPERTIES OF THE STRUCTURED AUTO-REGRESSIVE TIME-FREQUENCY DISTRIBUTION

Jakob Ängeby

Signal Processing, Department of Applied Electronics
Chalmers University of Technology, 412 96 GÖTEBORG, SWEDEN

ABSTRACT

Primarily the structured auto-regressive (AR) model was introduced as a mean to estimate the parameters of non-stationary signals in additive noise. However, it is straightforward to use the structured AR model as a model-based time-frequency distribution (TFD). It is shown that the structured AR TFD can be interpreted as a member of Cohen's class with a non-stationary adaptive kernel. The interpretation of the structured AR TFD as a member of Cohen's class establishes a link between TFD:s and signal parameter estimation.

1. INTRODUCTION

Power spectral density estimation (PSDE) of wide-sense stationary (WSS) signals was traditionally based on Fourier transforms, until a flurry of research activity the recent decades resulted in a number of alternative, high resolution, approaches. Fourier-based techniques using windowed data competed with model-based approaches. When applicable, model-based approaches were found to be superior to Fourier-based, especially in the case of short data records [6].

In recent years, time-frequency spectral density estimation (TFSDE) of non-stationary signals has received a considerable attention [4, 5]. A large variety of approaches have been suggested, of which almost all are non-parametric and many belong to the well-known Cohen class. Within the Cohen class different TFSDE are characterized by their kernel which, in analogy to windows in Fourier based PSDE, are designed to enhance useful information in the signal while suppressing artifacts and noise. Traditionally, TFD:s have used fixed kernels. However, fixed kernels limit the class of signals for which the TFD performs well. In an attempt to alleviate this drawback, signal-dependent kernels have been proposed [3, 7]. The proposed optimization criterions are, however, *ad hoc*.

The work was supported by *Ericsson Microwave Systems AB* and *Ericsson Infocom Consultants AB*

The structured AR approach to signal parameter estimation was introduced in [1] and the statistical properties for the case of a polynomial phase signal were investigated in [2]: The estimates are consistent and close to efficient. As indicated in [1], the structured AR filter can be used as a TFD. Simulations have shown some encouraging results. However, the properties of the structured AR approach as a TFD have not been investigated, and the relationship to other TFD:s has not been clarified.

2. STRUCTURED AR–MODELING

Here the main results of structured AR modeling are given for easy reference. Details of the derivation can be found in [2]. Consider a complex-valued deterministic signal, say $s_k(\boldsymbol{\vartheta}_0)$, parameterized by $\boldsymbol{\vartheta}_0$ and observed in white Gaussian noise with variance σ^2 at time instants $\{t_k\}_{k=0}^{N-1}$,

$$y(t_k) = s_k(\boldsymbol{\vartheta}_0) + e(t_k). \tag{1}$$

The time dependence of $s_k(\boldsymbol{\vartheta}_0)$ is written as an index in order to keep the notation short. The signal may be sampled non-uniformly in time. Let $s_k(\bar{\boldsymbol{\vartheta}})$ denote a user-defined (scaled) model signal, parameterized by a model signal parameter vector $\bar{\boldsymbol{\vartheta}}$. For numerical reasons the model signal $s_k(\bar{\boldsymbol{\vartheta}})$ is "scaled" such that one of the signal components in $s_k(\bar{\boldsymbol{\vartheta}})$ equals one for $k = 0$, and as a consequence $\dim(\bar{\boldsymbol{\vartheta}}) = \dim(\boldsymbol{\vartheta}_0) - 2$. Let $\bar{\boldsymbol{\vartheta}}_0$ denote the model signal parameter vector that corresponds to $\boldsymbol{\vartheta}_0$. Due to the scaling, $s_k(\bar{\boldsymbol{\vartheta}}_0)$ and $s_k(\boldsymbol{\vartheta}_0)$ differ by a constant multiplicative (complex) factor, say b_0. For example, if $s_k(\boldsymbol{\vartheta}_0)$ is a sum of two chirps, then

$$\begin{aligned}
s_k(\boldsymbol{\vartheta}_0) &= \sum_{l=1}^{2} b_l \exp\{j(a_{l0} + a_{l1}t_k + a_{l2}t_k^2)\},\\
\boldsymbol{\vartheta}_0 &= (a_{10}, a_{11}, a_{12}, a_{20}, a_{21}, a_{22}, b_1, b_2),\\
\bar{\boldsymbol{\vartheta}}_0 &= (a_{11}, a_{12}, (a_{20} - a_{10}), a_{21}, a_{22}, b_2/b_1),\\
b_0 &= b_1 \exp\{ja_{10}\},
\end{aligned}$$

and $s_k(\vartheta_0) = b_0 s_k(\bar{\vartheta}_0)$.
The time-dependent structured AR filter parameter vector of order n that minimizes the analytical prediction error variance is a function of $\bar{\vartheta}$, and can be written as

$$\begin{aligned} \boldsymbol{\theta}_k(\bar{\vartheta}) &= \beta_k(\bar{\vartheta})\, \boldsymbol{s}^*_{k-1}(\bar{\vartheta}) s_k(\bar{\vartheta}), & \text{(2a)} \\ \boldsymbol{s}_{k-1}(\bar{\vartheta}) &\stackrel{def}{=} (s_{k-1}(\bar{\vartheta}), \dots, s_{k-n}(\bar{\vartheta}))^T, & \text{(2b)} \\ \beta_k(\bar{\vartheta}) &\stackrel{def}{=} -\frac{1}{\sigma_0^2 + \boldsymbol{s}^H_{k-1}(\bar{\vartheta}) \boldsymbol{s}_{k-1}(\bar{\vartheta})}, & \text{(2c)} \\ \sigma_0^2 &\stackrel{def}{=} \frac{\sigma^2}{b_0^2}, & \text{(2d)} \end{aligned}$$

where '*' denotes complex conjugate. Note that (boldface) $\boldsymbol{s}_{k-1}(\bar{\vartheta})$ denotes an $(n|1)$ vector, whereas $s_k(\bar{\vartheta})$ is a scalar. Let $\boldsymbol{y}_{k-1}$ denote a vector of n measurements,

$$\boldsymbol{y}_{k-1} \stackrel{def}{=} (y(t_{k-1}), \dots, y(t_{k-n}))^T. \qquad (3)$$

The structured AR filter $\boldsymbol{\theta}_k(\bar{\vartheta})$ is used to predict $y(t_k)$ from $\boldsymbol{y}_{k-1}$, and the prediction is given by

$$\hat{y}_k(\bar{\vartheta}) = -\boldsymbol{\theta}_k^T(\bar{\vartheta})\, \boldsymbol{y}_{k-1}. \qquad (4)$$

Let the prediction error be denoted by $\bar{\varepsilon}_k(\bar{\vartheta})$. Then

$$\begin{aligned} \bar{\varepsilon}_k(\bar{\vartheta}) &= y(t_k) - \hat{y}_k(\bar{\vartheta}) \\ &= y(t_k) + \boldsymbol{\theta}_k^T(\bar{\vartheta})\, \boldsymbol{y}_{k-1}. & \text{(5)} \end{aligned}$$

The signal parameter estimates are found as the minimizing argument of the sum of squared prediction errors,

$$\begin{aligned} \hat{\bar{\vartheta}} &= \arg\min_{\bar{\vartheta}} V_N(\bar{\vartheta}), & \text{(6)} \\ V_N(\bar{\vartheta}) &\stackrel{def}{=} \frac{1}{N-n} \textstyle\sum_{k=n}^{N-1} \mid \bar{\varepsilon}_k(\bar{\vartheta}) \mid^2. & \text{(7)} \end{aligned}$$

The notation in (2a) and (7) is not entirely correct since $\boldsymbol{\theta}_k(\bar{\vartheta})$ actually depends on both $\bar{\vartheta}$ and σ_0^2, c.f. (2c). However, when implementing the structured AR algorithm, σ_0^2 in (2c) is preferably set to a constant, since otherwise the search may become ill-conditioned for high SNR. The performance of the estimator is not significantly decreased even if σ_0^2 is set to value $\pm$10dB from the true value, and the phase parameters are still consistently estimated. Then, after $\bar{\vartheta}$ has been estimated, b_0 and the noise variance σ^2 can be estimated using a straight-forward least squares fit of $s_k(\hat{\bar{\vartheta}})$ to $y(t_k)$.
Note that the signal parameters can be estimated even if the data is non-uniformly sampled and that once $\bar{\vartheta}$ has been estimated, then $\boldsymbol{\theta}_k(\bar{\vartheta})$ is defined for all t_k, i.e. $\boldsymbol{\theta}_k(\bar{\vartheta})$ can be calculated for any t.

3. TIME-FREQUENCY DISTRIBUTION

When the signal parameters have been estimated, the instantaneous phase and frequency are given by $\arg s_k(\hat{\bar{\vartheta}})$ and $\frac{d}{dt} \arg s_k(\hat{\bar{\vartheta}})$, respectively. The structured AR filter contains information on the instantaneous spectral density of the signal and can be used to construct a corresponding TFD. However, note that once the signal parameters have been estimated, all information about the signal is contained in the signal parameter estimates. Any other means to present the information is redundant and serve merely as a tool to visualize and interpret the information. In the following, TFD:s based on the structured AR filter are introduced and their properties are discussed. Let the structured AR parameters given by the estimated signal parameters be denoted by $\{c_l(t_k; \hat{\bar{\vartheta}})\}_{l=1}^n$;

$$\boldsymbol{\theta}_k(\hat{\bar{\vartheta}}) = \left(c_1(t_k; \hat{\bar{\vartheta}}), \dots, c_n(t_k; \hat{\bar{\vartheta}})\right). \qquad (8)$$

A TFD based on the structured AR parameters is achieved as follows:

1. Estimate ϑ_0 .
2. Choose a constant "sampling" interval $\Delta = t_k - t_{k-1}$ and a structured AR length n.
3. Calculate the structured AR parameters from (2a)-(2c).
4. Substitute $\boldsymbol{\theta}_k(\hat{\bar{\vartheta}})$ into a formula relating the AR parameters to the PSD.

The structured AR approach retains the parameters $\hat{\bar{\vartheta}}$ as continuous-time signal parameters, despite the fact that all calculations are performed using sampled data. As a consequence the structured AR TFD will generally be defined in $0 < \omega \leq 2\pi$ without aliasing effects.
A straight-forward choice of formula that relates the structured AR parameters to PSD is the theoretical AR PSD [6]:

$$\hat{P}_{AR}(t_k, w; \hat{\bar{\vartheta}}) = \frac{\hat{\sigma}^2}{|1 + \sum_{l=1}^n c_l(t_k; \hat{\bar{\vartheta}}) e^{-j\omega l}|^2}, \qquad (9)$$

where ω denotes the discrete frequency and $\hat{\sigma}^2$ is the estimated measurement noise variance, c.f. (1).
Another TFD based on the structured AR parameters is achieved by noting that the structured AR parameters are estimated scaled versions of the instantaneous autocorrelation function of $y(t_k)$. Let the (time-varying) covariance function of $y(t_k)$ be denoted by $r_k(u, v)$,

$$\begin{aligned} r_k(u,v) &\stackrel{def}{=} \mathrm{E}[y(t_{k-u}) y^*(t_{k-v})] \\ &= s_{k-u}(\vartheta_0) s^*_{k-v}(\vartheta_0)] + \sigma^2 \delta_{u,v}. \qquad (10) \end{aligned}$$

From (2a)–(2d), it holds true that

$$c_l(t_k;\bar{\vartheta}_0) = \beta_k(\bar{\vartheta}_0)s^*_{k-l}(\bar{\vartheta}_0)s_k(\bar{\vartheta}_0)$$
$$= \frac{\beta_k(\bar{\vartheta}_0)}{b_0^2}s^*_{k-l}(\vartheta_0)s_k(\vartheta_0) = -\frac{\beta_k(\bar{\vartheta}_0)}{b_0^2}r_k(0,l), \quad (11)$$

for $l > 0$. As a candidate for a TFD, consider

$$\rho_{AR}(k,\omega;\hat{\bar{\vartheta}}) \stackrel{def}{=} \sum_{l=1}^{n} \frac{\hat{b}_0^2}{\beta_k(\hat{\bar{\vartheta}})}c_l(t_k;\hat{\bar{\vartheta}})e^{-j\omega l}. \quad (12)$$

Using (11) it follows that

$$\rho_{AR}(k,\omega;\hat{\bar{\vartheta}}) = \sum_{l=1}^{n} r_k(0,l;\hat{\vartheta})e^{-j\omega l} \quad (13a)$$
$$= s_k(\hat{\vartheta})e^{-j\omega k}\left(\sum_{l=1}^{n} s_{k-l}(\hat{\vartheta})e^{-j\omega(k-l)}\right)^* \quad (13b)$$
$$\stackrel{def}{=} s_k(\hat{\vartheta})e^{-j\omega k}\mathcal{F}^*_{n-}\{s_k(\hat{\vartheta})\}, \quad (13c)$$

where $\mathcal{F}^*_{n-}\{s_k(\hat{\vartheta})\}$ denotes the discrete Fourier transform of $s_k(\hat{\vartheta})$ windowed by a rectangular window stretching from $k-1$ to $k-n$. From (13c) it is seen that $\rho_{AR}(k,\omega;\hat{\bar{\vartheta}})$ is the complex energy density of a windowed version of $s_k(\hat{\vartheta})$ and that it can be interpreted as a truncated discrete version of the Page distribution [5]. Since the Page distribution is a member of Cohen's general class of bilinear transforms, there is reason to believe that this is the case also for $\rho_{AR}(k,\omega;\hat{\bar{\vartheta}})$. The discrete version of Cohen's general class of TFD:s [4] can be written as

$$\rho(k,w) \stackrel{def}{=} 2\sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} G(m-k,l)r_m(-l,l)e^{-j2\omega l}, \quad (14)$$

where $G(m-k,l)$ is an implementation specific kernel that gives different TFD:s their properties. By equating (13a) to (14), it is seen that the structured AR TFD (12) can indeed be interpreted as a member of Cohen's class and that the kernel is

$$G(m-k,l;\hat{\vartheta}) = \begin{cases} \frac{1}{2} \mid r_m(-l,l;\hat{\vartheta}) \mid^{-2} \times \\ r^*_m(-l,l;\hat{\vartheta})r_k(0,l;\hat{\vartheta})e^{j\omega l}\delta_{l,m} & l = 1,\dots,n, \\ 0 & \text{else.} \end{cases} \quad (15)$$

$\rho_{AR}(k,\omega;\hat{\bar{\vartheta}})$ is not real-valued and hence an estimate of the instantaneous spectral density is given by the magnitude of $\rho_{AR}(k,\omega;\hat{\bar{\vartheta}})$.

The structured AR TFD:s do not contain any cross-terms. This property is a consequence of that the kernel contains $\delta_{m,l}$, which implies

$$\rho(k,w) = 2\sum_{l=-\infty}^{\infty} \tilde{G}(l-k,l)r_l(-l,l)e^{-j2\omega l} \quad (16a)$$
$$= 2\sum_{l=-\infty}^{\infty} \tilde{G}(l-k,l)s(t_{2l})s^*(t_0)e^{-j2\omega l}, \quad (16b)$$
$$= 2s^*(t_0)\sum_{l=-\infty}^{\infty} \tilde{G}(l-k,l)s(t_{2l})e^{-j2\omega l}, \quad (16c)$$

where $\tilde{G}(l-k,l)$ denotes the kernel without $\delta_{l,m}$. From (16c) it is seen that it is $s(t_{2l})$, i.e. not $s(t_l)s^*(t_{-l})$, that is transformed into the frequency domain, weighted with some weight-function $\tilde{G}(m-k,l)$. Cross-terms appear when a *product* of signals containing multiple components is transformed, which would have been the case in the Wigner-Ville transform.
Formulas (9) and (12) implicitly assume that the signal is an AR process and a moving average (MA) process, respectively. The match between (9) and (12) and the "true" time-frequency spectra is limited by the validity of the model, i.e. how well it describes the signal. From PSDE of WSS signals it is known that an AR model describes the signal using significantly less parameters than when using a corresponding MA model in the case of poorly damped signals, such as cisoids. When using structured AR filters a large number of model parameters corresponds to a large data window. However, using a large data window implies a low time resolution. A good TFD shall provide a high resolution in both time and frequency. Hence, it is reasonable to expect that the theoretical AR PSDE approach (9) will outperform the (MA) approach (12), especially in the case of complex valued signals with non-linear phases. Note that (12) is an estimate of the TFSD of $s_k(\vartheta_0)$. An estimate of the TFSD of $y(t_k)$ is provided by (9).
It is possible to construct other structured AR filters and loss-functions that correspond to well known TFD:s other than the Page distribution. For example, consider a "backward" structured AR implementation, i.e. $s_k(\vartheta_0)$ is estimated from $\{y(t_{k+l})\}_{l=1}^{n}$ and where the signal parameters are found by minimizing the sample variance of the backward prediction errors. The corresponding backward structured AR TFD can be interpreted as the future running transform, c.f. [5]. If the signal parameters are defined as the minimizing arguments of

$$V_N^{fb}(\bar{\vartheta}) = \frac{1}{N-n}\sum_{k=n}^{N-n}\left(\mid \bar{\varepsilon}_k^f(\bar{\vartheta}) \mid^2 + \mid \bar{\varepsilon}_k^b(\bar{\vartheta}) \mid^2\right), \quad (17)$$

where $\bar{\varepsilon}_k^f(\bar{\vartheta})$ and $\bar{\varepsilon}_k^b(\bar{\vartheta})$ denote the forward and backward prediction errors, respectively, then the corresponding TFD defined as

$$\rho_{AR}^{fb}(k,\omega;\hat{\bar{\vartheta}}) \stackrel{def}{=} \sum_{l=1}^{n}\left(\frac{\hat{b}_0^2}{\beta_k^f(\hat{\bar{\vartheta}})}c_l^f(t_k;\hat{\bar{\vartheta}})e^{-j\omega l}+\frac{\hat{b}_0^2}{\beta_k^b(\hat{\bar{\vartheta}})}c_l^b(t_k;\hat{\bar{\vartheta}})e^{j\omega l}\right) \tag{18}$$

can be interpreted as a truncated discrete version of the Rihaczek distribution. Hence, different structured AR filters and loss-functions correspond to different TFD:s.

4. NUMERICAL STUDY

Figure 1 shows a comparison of the structured AR TFD:s given in (9) and (12), the Wigner-Ville transform and the short time Fourier transform (STFT). The signal is the sum of a cisoid and linear FM which is observed in noise.

Figure 1 verifies that the structured AR TFD:s do not contain any cross-terms. The structured theoretical AR TFD (9) has a high resolution in both time and frequency, as expected. The "rugged" appearance of the theoretical AR TFSDE is explained by the fact that when signal cancellation occurs (i.e. zero crossings of the envelope), then the energy of the signal is zero. Structured AR TFD:s are only indirectly influenced by the noise, through the signal parameter estimates $\hat{\bar{\vartheta}}$. Hence, they are robust against noise.

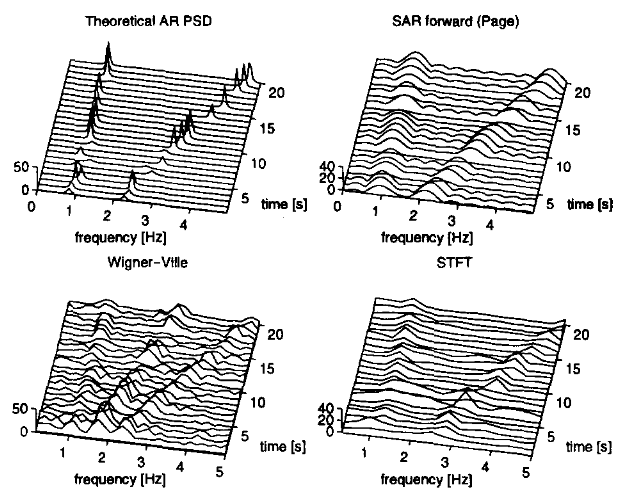

Figure 1: Structured AR TFSDE of a cisoid and linear FM of equal power in noise; SNR=3dB, N=200: (a) Structured theoretical AR, Eq (9), (b) Structured AR, Eq (12), (c) Wigner-Ville and (d) STFT.

5. CONCLUDING REMARKS

It has been shown how the structured AR filter coefficients can be used to construct TFD:s. The structured AR TFD:s are model based and, to the best of the authors knowledge, they are the first completely model-based TFD:s to be reported.

It was shown that $\rho_{AR}(k,\omega;\hat{\bar{\vartheta}})$, defined in (12), can be interpreted as a member of Cohen's class. Hereby a link between TFD:s of Cohen's class and signal parameter estimation has been established. Hence, through Parseval's relation, it can be conjectured that structured AR signal parameter estimation is equivalent to a fit of the corresponding (model-based) structured AR TFD to the time-frequency content of the measurements. Moreover, in applications l and m in Cohen's class of TFD:s (14) is limited such that $|l| < L$ and $|m| < M$. Then the kernel can be viewed as a complex valued matrix of dimension $(2L+1|2M+1)$, [4]. Using Parseval's relation, and the fact that $\mathrm{E}[\bar{\varepsilon}_k^2(\bar{\vartheta})]$ is minimal by construction, it can also be conjectured that the structured AR TFD kernel is optimal, in the sense that no other kernel of the same dimension can produce a TFD that is closer to the true time-frequency spectral density of the measurements. The latter conjecture holds, of course, only if the model is known.

6. REFERENCES

[1] J. Ängeby. Structured auto-regressive instantaneous phase and frequency estimation. In *ICASSP*, volume 3, pages 1768–1771. IEEE, May 1995.

[2] J. Ängeby. Statistical analysis of polynomial phase signal parameter estimates based on structured auto-regressive modeling. In *8:th IEEE Sig. Proc. Workshop on Stat. Sig. and Array Proc.*, pages 444–447, June 1996.

[3] R. G. Baraniuk and D. L. Jones. A signal dependent time-frequency representation: Otimal kernel design. *IEEE Trans. Sig. Proc.*, 41(4):1589–1602, Apr 1993.

[4] B. Boashash, editor. *Time-Frequency Signal Analysis, Methods and Applications.* Longman Cheshire, 1992.

[5] L. Cohen. Time-frequency distributions - a review. *Proc. IEEE*, 77:941–981, July 1989.

[6] S. M. Kay. *Modern Spectral Estimation, Theory and Applications.* Prentice-Hall signal processing series. Prentice Hall, 1988.

[7] T. Sang and W. J. Williams. Renyi information and signal-dependent optimal kernel design. In *ICASSP*, volume 3, pages 997–1000. IEEE, May 1995.