HAL
open science

Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients

Georgy Boshnakov, Sophie Lambert-Lacroix

To cite this version:

Georgy Boshnakov, Sophie Lambert-Lacroix. Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients. Journal of Time Series Analysis, 2009, 30 (5), pp.467-486. 10.1111/j.1467-9892.2009.00619.x . hal-00354373

HAL Id: hal-00354373

https://hal.science/hal-00354373v1

Submitted on 23 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients

Georgi N. Boshnakov
School of Mathematics
The University of Manchester
Oxford Road
Manchester M13 9PL, UK

Sophie Lambert-Lacroix
LJK, BP 53,
Université de Grenoble et CNRS
38041 Grenoble cedex 9, France

January 17, 2009

Abstract

We consider the maximum entropy extention of a partially specified autocovariance sequence of a periodically correlated process. The sequence may be specified on a non-contiguous set. We give a method which solves the problem completely—it gives the positive definite solution when it exists and reports that it does not exist otherwise. The method is numerically reliable even when the solution is "almost" semidefinite. It also works when only positive semidefinite extention(s) exist.

1 Introduction

The maximum entropy principle provides an appealing framework for the specification of complete models from partial information. In some sense, models chosen by this principle do not impose more restrictions than necessary to accomodate the available information.

The maximum entropy method was introduced to stationary time series by Burg in the influential works [2], [3]. Given a contiguous set of autocovariances for lags $0, \ldots, p$, the maximum entropy solution is an autoregressive process of order p with those autocovariances. In this case the problem is linear and the solution can be obtained by solving the Yule-Walker equations with the Levinson-Durbin algorithm. This result holds in the multivariate case as well. If the lags are not contiguous the problem is, in general, non-linear but the solution is still an autoregression of order equal to the maximum specified lag.

Lambert-Lacroix [9] generalised this result to periodically correlated processes (pc-processes) — given a contiguous set of autocovariances, the maximum entropy solution is a periodic autoregression process but for one or more seasons the order of the autoregression equations may not coincide with the maximal lags of the given autocovariances. This is in stark contrast with the stationary case. Maximum entropy estimation based on sample partial autocorrelations was studied in [8].

For univariate stationary processes, the particular case of non-contiguous lags when autocovariances are given on a lattice was studied by Politis [11] and the case of general gaps by Rozario and Papoulis [14].

Maximisation of the *entropy rate* of a process is closely related but different from the maximization of the entropy of a random vector. The maximum entropy problem for a random vector may be cast as a maximisation of a determinant and there is a large body of literature on matrix completion for determinant maximisation, see Johnson [7] for a survey. The maximisation of the entropy rate of a process may be related to limits of ratios of determinants. We use such techniques in some of our proofs.

In this paper we give a method which solves the maximum entropy problem for any given set of autocovariances of a periodically correlated process. The method gives the solution, if it exists, and reports that no solution exists otherwise.

The paper is organized as follows. Section 2 presents some basic results about pc-processes. Section 3 discusses the periodic autoregression model and its parameterisations. Section 4 introduces the periodic autocovariance problem, the main subject of this paper. Our method is presented in Section 5 and its properties and further details are given in Section 6. Numerical results illustrating the behavior of the method are presented in Section 7. Most proofs and some auxilliary results are given in Appendix A. Appendix B contains formulae for the calculation of the gradient and Hessian of the entropy.

We denote by $\mathbb{N}$ the set of the non-negative integers. We work mostly with double indexed sequences. If two sequences, say a and b, are defined over non-overlapping sets of indices, then $a \cup b$ stands for the merged sequence.

2 Periodically correlated processes

A zero-mean process $\{X_t, t \in \mathbb{N} \setminus \{0\}\}$ is periodically correlated of period T if its autocovariance function $R(u,v) = E\left\{X_u \overline{X_v}\right\}$ is T-periodic, i.e.

$$R(u+T, v+T) = R(u,v) \qquad \text{for all } (u,v) \in \mathbb{N}^2 \tag{1}$$

(see [5]). The matrix $\boldsymbol{R}_t = \{R(i,j)\}_{i,j=1,\ldots,t}$ is the covariance matrix of the random vector $[X_1, \ldots, X_t]^T$. Hence, $\boldsymbol{R}_t$ is positive semidefinite (p.s.d.) for every $t \in \mathbb{N} \setminus \{0\}$. If $\boldsymbol{R}_t$ is positive definite (p.d.) for every t, then the process is said to be not locally deterministic. Otherwise $\boldsymbol{R}_t$ is singular for some t and the process is locally deterministic.

It is convenient to think about the autocovariances in terms of the *seasons* $t = 1, \ldots, T$ and the *lags* $k \in \mathbb{N}$. Each pair $(u,v) \in \mathbb{N}^2$ may be represented as $(u,v) = (mT+t, mT+t-k)$ for some $t \in \{1, \ldots, T\}$, $m \in \mathbb{N}$, and integer k. From Equation (1) it follows that $R(mT+t, mT+t-k)$ does not depend on m. So, we may introduce the notation

$$R_t(k) = R(mT+t, mT+t-k), \qquad t \in \{1, \ldots, T\}, \ m \in \mathbb{N}, \ k\text{—integer}.$$

Moreover, it is sufficient to consider pairs (u,v) with $u \geq v$ since $R(u,v) = \overline{R(v,u)}$. For such pairs k is non-negative. If t is one of the seasons, $1, \ldots, T$, and k is a non-negative integer lag, then (t,k) will be called a *season-lag pair*.

The T functions $R_1(\cdot)$, $\ldots$, $R_T(\cdot)$, considered as functions on $\mathbb{N}$, completely parameterise the second order structure of the pc-process in the sense that for each (u,v) there is exactly one season-lag pair (t,k) such that $R(u,v) = R_t(k)$ when $u \geq v$ and $R(u,v) = \overline{R_t(k)}$ when $u < v$. In other words, the doubly indexed sequence $\{R_t(k)\}$, $t \in \{1, \ldots, T\}$, $k \in \mathbb{N}$, enumerates the autocovariances in a non-redundant way.

An equivalent parameterisation is given by the partial autocorrelations (pacf) $\{\beta_t(k)\}$, $t = 1, \ldots, T$, $k \in \mathbb{N}$ (see [9] for details).

Let $\{X_t\}$ be a pc-process and let $v_t(k)$ be the variance of the prediction error of X_t in terms of the k previous values $X_{t-1}, \ldots, X_{t-k}$. Then for any given $t \in \{1, \ldots, T\}$ the sequence $\{v_{mT+t}(mT+t-1)\}_{m=1}^{\infty}$ is convergent as $m \to \infty$ since it is monotonically decreasing and bounded from below by 0. Let

$$\sigma_t^2 = \lim_{m \to \infty} v_{mT+t}(mT+t-1), \qquad t = 1, \ldots, T. \tag{2}$$

An expression for σ_t^2 in terms of the partial autocorrelations is (see [9])

$$\sigma_t^2 = R_t(0) \prod_{n=1}^{\infty} (1 - |\beta_t(n)|^2), \qquad t = 1, \ldots, T.$$

It can be shown [6, p. 119] that for a Gaussian not locally deterministic pc-process X the entropy rate is equal to

$$h(X) = \frac{1}{2}\log(2\pi e) + \frac{1}{2T}\sum_{t=1}^{T}\log\sigma_t^2.$$

If σ_t^2 is equal to 0 for some t, then the entropy is defined to be $-\infty$.

Since we are considering only second order properties and the first term is a constant, we can define the entropy rate of a process with autocovariance sequence R by

$$h(R) = \frac{1}{T}\sum_{t=1}^{T}\log\sigma_t^2. \tag{3}$$

3 Periodic autoregression

A periodically correlated process $\{X_t\}$ is said to be a periodic autoregression (PAR) process if its dynamics are described by the equation

$$X_t - \sum_{i=1}^{p_t}\phi_{t,i}X_{t-i} = \varepsilon_t, \qquad t \in \mathbb{N}, \tag{4}$$

where $\{\varepsilon_t\}$ is a periodic white noise sequence with $\mathrm{Var}\{\varepsilon_t\} = \sigma_t^2$. The model orders, p_t, and the parameters, $\phi_{t,i}$ and σ_t^2, are periodic with period T, in the sense that $p_{t+T} = p_t$, $\phi_{t+T,i} = \phi_{t,i}$ and $\sigma_{t+T}^2 = \sigma_t^2$ for all t and i. For convenience we define $\phi_{t,i} = 0$ for $i > p_t$. The extended notation PAR($p_1,\ldots,p_T$) is used to emphasise the model order and normally implies that $\phi_{t,p_t} \neq 0$ for $t \in \{1,\ldots,T\}$. If ϕ_{t,p_t} may be zero for some t in $\{1,\ldots,T\}$, then we use the notation PAR($\leq p_1,\ldots,\leq p_T$). We assume also that the model is causal, i.e. ε_t is uncorrelated with X_s when $t > s$.

The standard parameterisation of the PAR($p_1,\ldots,p_T$) model is:

$$\{\,\phi_{t,k}, \sigma_t^2,\ t \in \{1,\ldots,T\},\ k \in \{1,\ldots,p_t\}\}. \tag{5}$$

The periodic Levinson-Durbin algorithm (see [9]) computes, among other things, the standard parameters of a PAR($p_1,\ldots,p_T$) model from the first few autocovariances,

$$\{\,R_t(k),\ t \in \{1,\ldots,T\},\ k \in \{0,1,\ldots,p_t\}\}. \tag{6}$$

Hence, the autocovariances listed in Equation (6) are an alternative set of parameters for the PAR($p_1,\ldots,p_T$) model.

Yet another parameterisation of the PAR(p_1,...,p_T) model is given by the partial autocorrelations (see [9])

$$\{\, \beta_t(k),\ t \in \{\, 1, \ldots, T \,\},\ k \in \{\, 0, 1, \ldots, p_t \,\} \,\}.$$

Similarly to the stationary case, the periodic partial autocorrelation coefficient $\beta_t(k)$ of a PAR(p_1,...,p_T) process is zero when $k > p_t$.

The autocovariances and the standard parameters of a PAR(p_1,...,p_T) process are related by the periodic Yule-Walker equations [10],

$$R_t(k) = \sum_{i=1}^{p_t} \phi_{t,i} R_t(k-i) + \delta_k \sigma_t^2, \qquad t = 1, \ldots, T, \quad k \in \mathbb{N}, \tag{7}$$

where $\delta_k = 1$ if $k = 0$ and $\delta_k = 0$ otherwise.

If a periodically correlated process is PAR(p_1,...,p_T), then $v_t(k) = v_t(p_t) = \sigma_t^2$ for $k \geq p_t$ (see [9]), i.e. the limits in Equation (2) are reached at finite lags. This is consistent with the notation used for the variance of the periodic white noise sequence in Equation (4).

4 The periodic autocovariance extension problem

A periodic autocovariance sequence is completely specified if its values are given for all season-lag pairs, i.e. for all (t, k) such that $t \in \{\, 1, \ldots, T \,\}$ and $k \in \mathbb{N}$. Let I be a set of season-lag pairs and $K = \{R_t(k)\}_{(t,k)\in I}$ be a sequence defined on I. Let Γ be the set of all periodic autocovariance sequences whose values coinside with $R_t(k)$ for $(t, k) \in I$ (Γ may be empty). Each element of Γ is a completion (or extension) of K. The maximum entropy extension is the one whose entropy rate is maximal in Γ.

ME(K,I) problem. Given a set I of season-lag pairs and a sequence K defined on I, find the completion of K whose entropy rate is maximal or show that such a completion does not exist.

The maximum entropy solution of the ME(K,I) problem is a periodic autocovariance function which will be denoted by $\gamma^{\mathrm{ME}}(K)$ and its entropy by $h_{\mathrm{ME}}(K)$.

Let $\tau(t) = \{\, k \mid (t, k) \in I \,\}$ for $t = 1, \ldots, T$. In the context of the completion problem, $\tau(t)$ is the set of those non-negative lags, k, for which $R_t(k)$ is fixed. With this notation we may rewrite I and K equivalently as

$$I = \{\, (t, k) \mid t \in \{\, 1, \ldots, T \,\},\ k \in \tau(t) \,\},$$
$$K = \{\, R_t(k),\ t \in \{\, 1, \ldots, T \,\},\ k \in \tau(t) \,\}.$$

We will assume that the lag zero autocovariances are given for all seasons, i.e. $(t,0) \in I$ for $t \in \{1,\ldots,T\}$, or, equivalently, $0 \in \tau(t)$ for $t \in \{1,\ldots,T\}$. Without this restriction the completion problem has no solution since the entropy rate may be made arbitrarily large by setting the non-specified variances to large values. We assume also that the set I is finite. Then for each $t \in \{1,\ldots,T\}$, $\tau(t)$ is a subset of the set $\{0,1,\ldots,\max\tau(t)\}$, where $\max\tau(t)$ is the maximal element of $\tau(t)$.

Obviously, the completion problem ME(K,I) has no solution if the set Γ is empty. It can be shown [1] that if Γ contains at least one positive definite sequence, then a positive definite solution exists and is unique. In this case we say that K is p.d.-completable. The remaining possibility is that Γ contains only positive semi-definite sequences, all with entropy equal to $-\infty$. We will take the view that any member of Γ is a (p.s.d.) solution in this case and say that K is p.s.d.-completable. K is completable if it is either p.d. or p.s.d.-completable. Finally, we say that K is p.d.-PAR($p_1,\ldots,p_T$) completable if it has a p.d. PAR completion of order $(p_1,\ldots,p_T)$, or less.

The completion problem requires the specification of values for the infinitely many season-lag pairs that are not in I. We will show that the completion may be accomplished by working on an extention $E_c(I)$ of I defined as follows.

Definition 1. Let I be a set of season-lag pairs. The smallest constrained season-lag set containing I is a set, $E_c(I)$, of season-lag pairs,

$$E_c(I) = \{\,(t,k) \mid t = 1,\ldots,T,\ k = 0,\ldots,p_t\,\}\,, \tag{8}$$

where $(p_1,\ldots,p_T)$ are the smallest non-negative integers satisfying the constraints $p_1 \le p_T + 1$ and $p_t \le p_{t-1} + 1$ for $t = 2,\ldots,T$, and such that $E_c(I) \supseteq I$.

In particular, $p_t \ge \max\tau(t)$ and p_t may be larger than $\max\tau(t)$ for one or more ts. For example, let $T = 2$ and $I = \{\,(1,0),(2,0),(2,2)\,\}$. Then $p_1 = 1$, $p_2 = 2$, and $E_c(I) = \{\,(1,0),(1,1),(2,0),(2,1),(2,2)\,\}$. So, in this case $E_c(I) \neq I$. Note that $\max\tau(1) = 0 < p_1$.

The constraints on $(p_1,\ldots,p_T)$ in the definition of $E_c(I)$ ensure that $\tau(t)$ is a subset of $\{0,1,\ldots,p_t\}$ for each $t \in \{1,\ldots,T\}$. Let $\tau^c(t) = \{0,1,\ldots,p_t\} \setminus \tau(t)$ be the complement of $\tau(t)$ in the set $\{0,1,\ldots,p_t\}$.

It is natural to refer to the elements of $E_c(I) \setminus I$ as *gaps* since these are season-lag pairs with no prespecified values in K.

The importance of the set $E_c(I)$ was discovered by Lambert-Lacroix [9] who introduced it (using somewhat different notation) and showed that if $I = E_c(I)$, then the maximum entropy solution is a PAR($\le p_1,\ldots,\le p_T$)

model. The problem ME(K,I) is linear in this case and the solution can be obtained by setting all partial autocorrelations for larger lags to zero and using periodic Levinson-Durbin recursions to obtain the autocovariances [9]. The following theorem summarises some results from [9] to which we refer later. For more general settings with linear solutions see [1].

Theorem 1. *Let $p_1, \ldots, p_T$ be non-negatve integers satisfying the constraints $p_1 \leq p_T + 1$ and $p_t \leq p_{t-1} + 1$ for $t = 2, \ldots, T$. Let $R = \{ R_t(k), (t,k) \in I \}$ be a sequence defined on the set*

$$I = \{ (t,k) \mid t \in \{ 1, \ldots, T \}, k \in \{ 0, 1, \ldots, p_t \} \}.$$

Then the following claims hold:

1. *$I = E_c(I)$.*
2. *The periodic Levinson-Durbin algorithm determines whether R is p.d.-completable or not.*
3. *If R is p.d.-completable, then*
   (a) *R is p.d.-PAR($p_1, \ldots, p_T$) completable;*
   (b) *The entropy of the p.d.-PAR($p_1, \ldots, p_T$) completion of R is strictly greater than the entropy of any other completion of R;*
   (c) *The p.d.-PAR($p_1, \ldots, p_T$) completion is a PAR($\leq p_1, \ldots, \leq p_T$) model whose standard parameters can be computed with the periodic Levinson-Durbin algorithm.*

We loosely refer to the case considered by Theorem 1 as the *contiguous case* since for each $t \in \{ 1, \ldots, T \}$ the set I contains all season-lag pairs (t,k) from the contiguous set $k = 0, 1, \ldots, p_t$. Note however that for the equality $I = E_c(I)$ we need also the condition on the p_ts.

If $I \neq E_c(I)$, then the completion problem is in general non-linear but, as we discuss in the following sections, the solution is still a PAR($\leq p_1, \ldots, \leq p_T$) process. It is therefore sufficient to complete the missing values $R_t(k)$ for $t = 1, \ldots, T$, $k \in \tau^c(t)$, and obtain the remaining autocovariances with the periodic Levinson-Durbin algorithm. We propose a procedure that gives (a numerical approximation to) the solution, if it exists, and reports that no solution exists otherwise.

5 The method

To put the results in context we give here a description of the proposed method and its main properties. Further details and justification of the claims in this section are provided in the following sections.

The solution of the ME(K,I) problem is the acf of a PAR($\leq p_1$,...,$\leq p_T$) model, where $p_1, \ldots, p_T$ satisfy the constraints given in Definition 1. This model is determined uniquely from the autocovariances on the set $E_c(I)$. In particular, the entropy of the PAR model may be determined from these autocovariances. The entropy may be considered as a function on the gaps $(t, k) \in E_c(I) \setminus I$ since the autocovariances are fixed for $(t, k) \in I$. It is therefore sufficient to work on the gaps $E_c(I) \setminus I$. The following sections establish that this function has some very desirable properties for numerical optimization when a p.d. completion exists. The problematic task is to find an initial value or show that the problem has no solution. We do not attack this task directly. Instead, we solve a sequence of modified problems constructed so that they have p.d. solutions and the solution of each one serves as the initial value for the next one. The initial value for the first problem is arbitrary. The most important property of this procedure is probably that when a p.d. completion for K exists it solves the original problem (and therefore can be terminated) after a finite number of steps.

The modified problems are obtained by adding sufficiently large constants to the lag zero autocovariances. Namely, let c be any real number and $K^c = \{ R_t^c(k), (t, k) \in I \}$ be a modification of K obtained by adding a constant, c, to $R_t(0)$ for $t = 1, \ldots, T$, i.e.

$$R_t^c(k) = \begin{cases} R_t(0) + c & \text{for } k = 0, \\ R_t(k) & \text{for } k \neq 0 \text{ and } k \in \tau(t). \end{cases} \tag{9}$$

The modified problems are ME(K^{c_i},I) for suitably chosen c_i's.

Our method consists of the following steps.

(1) Determination of PAR order Determine the set $E_c(I)$ (see Equation (8)), i.e. find $p_1, \ldots, p_T$ satisfying the constraints of Definition 1. The maximum entropy solution, if it exists, is a PAR($\leq p_1$,...,$\leq p_T$) process which is uniquely determined by its autocovariances for season-lag pairs in $E_c(I)$. It is therefore sufficient to find values for the autocovariances for the gaps $(t, k) \in E_c(I) \setminus I$.

(2) Initialisation Fill the gaps $(t,k) \in E_c(I) \setminus I$ with arbitrary values $w_t^{c_0}(k)$, e.g. zeroes, and let $w^{c_0} = \{w_t^{c_0}(k)\}_{(t,k)\in E_c(I)\setminus I}$. Set $c_0 = \infty$ and $n = 0$.

(3) Iteration (repeat while necessary)

(i) Set $n = n+1$ and check if the completion obtained at step $n-1$ as the solution of the ME($K^{c_{n-1}}$,$E_c(I)$) problem may be used as initial value for the unmodified ME(K,I) problem, i.e. if the sequence R^{init} obtained by merging K and $w^{c_{n-1}}$ provides a completion of K. The sequence R^{init} can be written more formally as $R^{\text{init}} = \{R_t^{\text{init}}(k)\}_{(t,k)\in E_c(I)}$, where

$$R_t^{\text{init}}(k) = \begin{cases} R_t(k) & \text{for } (t,k) \in I, \\ w_t^{c_{n-1}}(k) & \text{for } (t,k) \in E_c(I) \setminus I. \end{cases}$$

Alternatively,

$$\begin{aligned} R^{\text{init}} &= K \cup w^{c_{n-1}} \\ &= \{\, R_t(k),\ (t,k) \in I \,\} \cup \{\, w_t^{c_{n-1}}(k),\ (t,k) \in E_c(I) \setminus I \,\} \end{aligned}$$

(ii) If the answer in (i) is *yes*, then set $c_n = 0$. Otherwise set c_n to a value in the interval $(0, c_{n-1})$ such that the modified R^{init}

$$\begin{aligned} R^{\text{init}} &= K^{c_n} \cup w^{c_{n-1}} \\ &= \{\, R_t^{c_n}(k),\ (t,k) \in I \,\} \cup \{\, w_t^{c_{n-1}}(k),\ (t,k) \in E_c(I) \setminus I \,\} \end{aligned}$$

provides a completion of K^{c_n} (see the remarks below).

(iii) Solve the ME(K^{c_n},$E_c(I)$) problem using R^{init} as initial value. This gives the gap "fillers" $w^{c_n} = \{\, w_t^{c_n}(k) \mid (t,k) \in E_c(I) \setminus I \,\}$.

The theory developed in the following sections shows that the procedure outlined above may be turned into an algorithm with excellent properties. Its main features are summarised in the following remarks.

Remark 1. If $c_n = 0$ for some n, then $K \cup w^{c_n}$ provides the maximum entropy completion of K, i.e. the solution of the ME(K,I) problem.

Remark 2. If K is p.d.-completable, then c_n is guaranteed to become 0 after a finite number of iterations of step (3). Together with Remark 1 this means that if the maximum entropy p.d. completion exists, then it will be found in a finite number of repetitions of step (3).

Remark 3. The solution of the modified problem ME(K^{c_n},$E_c(I)$) at any iteration $n \geq 1$ gives an excellent initial value for the following iteration, $n+1$.

Remark 4. The sequence c_n is monotonically decreasing.

Remark 5. If the solution is semi-definite, then c_n monotonically decreases to 0.

Remark 6. If the ME(K,I) problem has no solution, then c_n decreases monotonically to a positive limit.

Remark 7. If a p.d. solution of the ME(K,I) problem exists, then there exists a positive constant δ such that at each step c_n can be reduced by at least δ. This ensures that after a finite number of iterations it will be possible to set $c_n = 0$. It is possible, at each iteration, to determine the smallest admissible value of c_n for step (3ii). A simple heuristic rule is sufficient however since we always check first that $c_n = 0$ is admissible. Moreover, zero often becomes an admissible value for c_n after the first iteration and in numerical experiments with the wildest choices of the starting value, w^{c_0}, and c_0, the value $c_n = 0$ becomes admissible after only a few steps (when a p.d. solution exists, of course).

Remark 8. It is not necessary to determine c_0 optimally. If it is larger than necessary, then at the next step it will be possible to reduce it more.

Remark 9. The method works very well numerically and, in particular, when the maximum entropy solution is "almost" p.s.d. or even exactly p.s.d., although the latter case is not likely to happen except for artificial examples.

Remark 10. For stationary processes, the idea to maximise the entropy of a modified autocovariance sequence was employed by Rozario and Papoulis [14]. They assumed from the outset that a p.d. solution exists and did not notice the finite number of iterations property. Thus, our treatment is not only a generalisation of the method to the periodic case but it also solves the problem completely by detecting when a solution does not exist or is only p.s.d..

6 Properties and details

6.1 Determination of the PAR order

The values of $p_1, \ldots, p_T$ and the set $E_c(I)$ can be determined from I by ensuring that the constraints of Definition 1 are satisfied. To achieve this, set initially p_i to the maximum element of $\tau(i)$, for $i = 1, \ldots, T$, and let s be such that $p_s = \max(p_1, \ldots, p_T)$. For convenience, extend $p_1, \ldots, p_T$ periodically,

so that $p_0 \equiv p_T$, $p_{-1} \equiv p_{T-1}$, and so on. Then set $p_{s-i} = \max(p_{s-i+1} - 1, p_{s-i})$ for $i = 1, \ldots, T-1$ (in this order) to ensure that the p's are as required.

6.2 Maximization of the entropy rate

In this section we give details about Step 3(iii) of the method described in Section 5. At this step we maximise the entropy rate of a sequence which is known to be p.d.-completable starting with an initial completion determined at Steps 3(i–ii).

Let K^c be a sequence defined on the season-lag set I and let γ_0 be a p.d. sequence whose restriction to I coincides with K^c. Let $E_c(I)$ and $p_1, \ldots, p_T$ be as defined in Definition 1. In this section we consider the solution of the ME(K^c,I) problem given the initial solution γ_0.

Let Γ^c_{PAR} be the set of all p.d.-PAR($\leq p_1$,...,$\leq p_T$) completions of K^c. Let $\gamma \in \Gamma^c$ be any completion of K^c. Consider the restriction of γ on $E_c(I)$. The season-lag set $E_c(I)$ is a contiguous set of the form

$$E_c(I) = \{\, (t,k) \mid t = 1, \ldots, T,\ k = 0, \ldots, p_t \,\},$$

where $p_1, \ldots, p_T$ satisfy the constraints in Definition 1. From part 2. of Theorem 1 it follows that there exists a unique PAR($\leq p_1$,...,$\leq p_T$) acf, γ_{PAR}, which coincides with γ on $E_c(I)$ and whose entropy, $h(\gamma_{\text{PAR}})$, is strictly larger than that of any other sequence which coincides with γ on $E_c(I)$. So for any γ in Γ^c, one can associate a completion γ_{PAR} in Γ^c_{PAR} such that

$$h(\gamma) \leq h(\gamma_{\text{PAR}}),$$

with equality if and only if $\gamma = \gamma_{\text{PAR}}$. Hence, we have the following lemma.

Lemma 1. *The maximum entropy solution of the ME(K^c,I) problem is a PAR($\leq p_1$,...,$\leq p_T$) acf, i.e.*

$$\operatorname*{argmax}_{\gamma \in \Gamma^c} h(\gamma) = \operatorname*{argmax}_{\gamma \in \Gamma^c_{PAR}} h(\gamma).$$

Note that $p_1, \ldots, p_T$ are associated with the set $E_c(I)$ which contains I but, in general, is larger.

Consider the problem ME($K^c \cup \omega$,$E_c(I)$) , where ω is an admissible filling of the gaps $E_c(I) \setminus I$. The sequence $K^c \cup \omega$ is given on the constrained set $E_c(I)$ which satisfies the requirements of Theorem 1. Hence, for any given ω, the solution to the ME($K^c \cup \omega$,$E_c(I)$) problem is a PAR acf which may be obtained using the PLD algrorithm.

Let $h_{\mathrm{ME}}(K^c \cup \omega)$ be the entropy rate of the solution of the ME($K^c \cup \omega$,$E_c(I)$) problem. The function $h_{\mathrm{ME}}(K^c \cup \omega)$ may be considered as a function of the gaps only, since K^c is fixed. So, we will write simply $h_{\mathrm{ME}}(\omega)$. To solve the ME(K^c,I) problem we need to maximise $h_{\mathrm{ME}}(\omega)$ over all admissible ωs. This maximisation is greatly facilitated by the fact that $h_{\mathrm{ME}}(\omega)$ is very well behaved—it is strictly concave on the set of the admissible fillings of the gaps, and it is continuous and smooth. We discuss these properties below.

Any convex linear combination of admissible gap fillings is admissible. For reference, we formulate this as a lemma, see Appendix A.1 for the proof.

Lemma 2. *The set of the admissible gap fillings is convex.*

The following result (and its proof) is valid under more general conditions but we formulate it in a form sufficient for our purposes.

Lemma 3. *Let the process $\{V_t\}$ be defined by the equation*

$$V_t = \theta X_t + (1-\theta) Y_t,$$

where $\{X_t\}$ and $\{Y_t\}$ are processes with acf K_1 and K_2, respectively, and θ is a random variable taking values 1 and 0 with probability α and β, respectively, and independent of $\{X_t\}$ and $\{Y_t\}$. Let $\hat{X}_t = \sum_{i=1}^{t-1} \alpha_{t,i}^X X_{t-i}$ be the best linear predictor of X_t given $X_{t-1}, \ldots, X_1$ and let $\sigma_{t,X}^2$ be the variance of the corresponding prediction error. Define similarly $\sigma_{t,Y}^2$, $\alpha_{t,i}^Y$ and $\sigma_{t,V}^2$, $\alpha_{t,i}^V$. Then

1. *The autocovariance function of V_t is equal to $\alpha K_1 + \beta K_2$.*
2. *$\sigma_{t,V}^2 \geq \alpha \sigma_{t,X}^2 + \beta \sigma_{t,Y}^2$ for each $t > 1$, with equality if and only if $\alpha_{t,i}^X = \alpha_{t,i}^Y$ for $i = 1, \ldots, t-1$.*
3. *$\log(\sigma_{t,V}^2) \geq \alpha \log(\sigma_{t,X}^2) + \beta \log(\sigma_{t,Y}^2)$, with equality if and only if $\alpha_{t,i}^X = \alpha_{t,i}^Y$ for $i = 1, \ldots,$ and $\sigma_{t,X}^2 = \sigma_{t,Y}^2$.*

Note that for equality in part (3) of the lemma it is not sufficient to require that $\sigma_{t,X}^2 = \sigma_{t,Y}^2$, the prediction coefficients should also be the same. The following corollary is used to show the concavity of $h_{\mathrm{ME}}(\omega)$.

Corollary 1. *In the notation of Lemma 3, let X and Y be PAR($\leq p_1, \ldots, \leq p_T$) processes. Then for any t_0 greater than $\max(p_1, \ldots, p_T)$ we have*

$$\sum_{t=t_0+1}^{t_0+T} \log(\sigma_{t,V}^2) \geq \alpha \sum_{t=t_0+1}^{t_0+T} \log(\sigma_{t,X}^2) + \beta \sum_{t=t_0+1}^{t_0+T} \log(\sigma_{t,Y}^2)$$

with equality if and only if the autocovariance functions of X and Y are the same.

Lemma 4. *$h_{ME}(\omega)$ is strictly concave.*

It follows that $h_{\mathrm{ME}}(\omega)$ is continuous since it is a concave function defined on an open set (see e.g. [13, p. 93, Theorem D]). From the periodic Yule-Walker equations we can also see that its first and second order partial derivatives (as well as higher order ones) exist. Together with concavity the existence of the first order partial derivatives implies differentiability of $h_{\mathrm{ME}}(\omega)$ (see [13, p. 101, Theorem D]). We also may conclude that the Hessian matrix of $h_{\mathrm{ME}}(\omega)$ is positive definite (see [13, p. 103, Theorem F]).

The above results suggest that methods based on first and second derivatives, such as gradient methods and Newton-Raphson's method, should be sufficiently efficient for maximisation of $h_{\mathrm{ME}}(\omega)$. The entropy is a very complicated function of the gaps and it is hardly possible to write down useful expressions for its derivatives for a general pattern of the gaps. So, we develop recursions for the first and second order derivatives based on the periodic Levinson-Durbin algorithm. Details are given in Appendix B.

6.3 The modified ME problems

The essence of Step 3 of our method (see Section 5) is that instead of searching for a completion of K to be used as an initial value for numerical maximisation of the entropy for the ME(K,I) problem, we deal with the simple problem of finding a modification, K^c, of K for which a given initial value is a completion, and solve the modified problem ME(K^c,I). When this process is iterated, the modification needed becomes smaller and smaller. The properties of the resulting procedure and further details are discussed in this section.

The following theorem is the backbone of our method. Its proof is given in Appendix A.4.

Theorem 2. *Let c be any real number and $K^c = \{ R_t^c(k), (t,k) \in I \}$ be a modification of K obtained by adding c to $R_t(0)$ for $t = 1, \ldots, T$, i.e.*

$$R_t^c(k) = \begin{cases} R_t(0) + c & \text{for } k = 0, \\ R_t(k) & \text{for } k \neq 0 \text{ and } k \in \tau(t). \end{cases}$$

Then there is a constant ω such that K^c is p.d.-completable *for $c > \omega$, K^c is not completable for $c < \omega$, and K^c is* p.s.d.-completable *for $c = \omega$.*

Of course, if K is p.d.-completable, then $\omega < 0$. If K is p.d.-completable, then K^c is completable for any $c > 0$ and $h_{\mathrm{ME}}(K^c) > h_{\mathrm{ME}}(K)$. Note that the inequality is strict.

For any fixed sequence K, $h_{\mathrm{ME}}(K^c)$ can be viewed as a function of c, defined over (ω, ∞), where ω is defined as in Theorem 2. The following lemma shows that $h_{\mathrm{ME}}(K^c)$ is a well behaved function.

Lemma 5. *$h_{ME}(K^c)$ is strictly concave and continuous as a function of c.*

The following property is intuitively clear but the "strict" claim is somewhat delicate.

Lemma 6. *$h_{ME}(K^c)$ is strictly increasing as a function of c.*

The above properties provide the basis for optimisation w.r.t. c. Let $\gamma^{c,\mathrm{ME}}$ be the acf associated to the entropy rate $h_{\mathrm{ME}}(K^c)$. The following result shows that $\gamma^{c_n,\mathrm{ME}} \to \gamma^{c_0,\mathrm{ME}}$ when $c_n \to c_0$. In particular, this property gives an explanation as to why the solution, $\gamma_n^{c,\mathrm{ME}}$, obtained at step n is a good initial value for maximisation at step $n+1$.

Lemma 7. *$\gamma^{c,ME}$ is continuous as a function of c defined on (ω, ∞).*

Theorem 2 shows, in particular, that by trying larger and larger cs we eventually will find a p.d.-completable K^c which can be used at the first iteration of our method. At the nth iteration of step (3) of our method we need to modify the completion from the preceding iteration if it cannot be used directly as initial value for our original problem. Moreover, we wish to modify with a smaller c. In general, we wish $c_n \to 0$. We show below that the c_ns can be chosen to be monotonically decreasing. Moreover, when a p.d. completion exists, there is a universal positive constant d^* such that c_n can be always chosen so that $c_n < c_{n-1} - d^*$. The constant is universal in the sense that it does not depend on the previous cs. So, after finite number of iterations the value $c_n = 0$ will become admissible.

Let $\gamma^{c,\mathrm{ME}}$ be the maximum entropy p.d. completion of K^c. Then for sufficiently large δ the sequence $\{\gamma_t(k)\}_{(t,k)\in I}$ defined by

$$\gamma_t(k) = \begin{cases} R_t(0) + \delta + a & \text{for } k = 0,\ t = 1, \ldots, T, \\ R_t(k) & \text{for } k \neq 0 \text{ and } (t,k) \in I, \\ \gamma_t^{c,\mathrm{ME}}(k) & k \in \tau^c(t), \end{cases} \tag{10}$$

is p.d.-PAR$(p_1, \ldots, p_T)$ completable for any $a > 0$. Let $q(c)$ be the infimum of the δ's for which this is the case. It is clear that $q(c) \leq c$. Then $d_c = c - q(c)$ is the largest value that can be subtracted from the lag zero autocovariances of K^c keeping at least the p.s.d. property. However, if it turns out that $q(c) = c$, then this would mean that $\gamma^{c,\mathrm{ME}}$ is not a completion for any c_0 smaller than

c and therefore cannot be used as initial value for any mep problem with smaller c. This would be a problem for the maximisation procedure.

Lemma 10 in Appendix A.9 show that $q(c) < c$ (strict inequality). Hence, it is always possible to use the mep solution for some value of c as an initial value for a mep problem with smaller c.

Let $d^* = \inf_{c \geq \max(0,\omega)} d_c$, where ω is the one defined in Theorem 2. The constant d^* is a universal lower bound on the possible reduction of c_n at Step (3) of our method. From the discussion so far we know that $d^* \geq 0$. The following theorem shows that the inequality is strict, see Appendix A.9 for the proof.

Theorem 3. *If K is p.d.-completable, then $d^* > 0$.*

Theorem 3 shows that at iteration n of Step (3) of the method c_{n-1} can be reduced by at least, say, $d^*/2$ by setting $c_n = c_{n-1} - d^*/2$. Therefore, after a final number of iterations the value 0 will become admissible value for c_n, i.e. R^{init} specified at Step (3i) will provide a completion of K to be used as initial value for the unmodified ME(K,I) problem and after the maximisation step (3iii) the algorithm will terminate.

Informally, the positiveness of d^* means that the iterations of step (3) of the method may use a monotonically decreasing sequence of cs and they will never get stuck at a particular value of c. This property is due to the fact that we are solving a maximum entropy problem at each step and would not hold if we did not do that.

7 Numerical results

We illustrate our method on five examples with $T = 2$ and autocovariances given up to lag 3 (see Table 7) except for a gap at $(1, 2)$ in Examples 1–4 and $(2, 1)$ in Example 5. We have

$$I = \begin{cases} \{\,(1,0),(1,1),(1,3), \quad (2,0),(2,1),(2,2),(2,3)\,\} & \text{(Examples 1–4)}, \\ \{\,(1,0),(1,1),(1,2),(1,3),\ (2,0),(2,1),(2,3)\,\} & \text{(Example 5)}. \end{cases}$$

The constrained set $E_c(I)$ is the same for all examples:

$$E_c(I) = \{\,(1,0),(1,1),(1,2),(1,3),\ (2,0),(2,1),(2,2),(2,3)\,\}\,.$$

The examples were constructed by computing autocovariances of PAR models specified with their partial autocorrelations. We therefore know the solutions, when they exist, and are able to compare them to the numerical results.

k	0	1	2	3
		Example 1		
$R_1(k)$	1	0.5	?	-0.04892034
$R_2(k)$	1	0.3	-0.09784067	-0.0293522
		Example 2		
$R_1(k)$	1	0.9	?	0.1307306
$R_2(k)$	1	0.3	0.1452563	0.04357688
		Example 3		
$R_1(k)$	1	0.999999	?	0.2995947
$R_2(k)$	1	0.3	0.299595	0.0898785
		Example 4		
$R_1(k)$	1	0.5	?	0.4880678
$R_2(k)$	1	0.3	0.9761356	0.2928407
		Example 5		
$R_1(k)$	1	2	0.7	0.9
$R_2(k)$	1	0.3	?	0.9

Table 1: Autocovariances for the numerical examples. The gaps are designated with question marks.

In the first three examples the maximum entropy p.d. completion exists. The first example illustrates the case when the starting values give at once a p.d. completion, i.e. we can set $c_1 = 0$. In the second example the initial values do not give a p.d. solution and we need to solve a modified problem at the first step. The first two examples represent typical situations while the third example whose solution is "almost" p.s.d. represents a borderline case. This example is arguably artificial but it demonstrates vividly the reliability of the method. The sequence c_n decreases quickly and at the fifth iteration we can set $c_5 = 0$.

In the fourth example, the solution exists but is p.s.d.. Finally, in the last example there are no solutions.

We did the calculations with the statistical system **R** (see [12]). We used rather conservative stopping criteria in order to study more reliably the behaviour of the method. In all cases we use zeroes as initial values for the missing values.

Example 1 Here the gap is at $(1, 2)$, i.e. $t = 1$, $k = 2$. The "true" maximum entropy solution is obtained for $R_1(2) = 0.15$. Table 2 gives results

c_n	i	$R_1(2)$	entropy	gradient
0	0	0.0000	-0.2660	0.3814
	1	0.1355	-0.2384	0.0349
	2	0.1499	-0.2382	3.3148e-05
	3	0.1500	-0.2382	2.8209e-14

Table 2: Results for Example 1. The initial value $R_1(2) = 0$ is admissible, so $c_1 = 0$ and the optimal solution $R_1(2) = 0.15$ is reached quickly.

obtained with our procedure for this example. The first column gives the values tried for c_n. The results for each iteration at the corresponding values of c_n are given in the following columns, starting with the iteration number, then the value for the missing autocovariance coefficient, the entropy, and the value of the gradient.

Here it is possible to set immediately $c_n = 0$ and after 3 iterations the procedure leads to the expected solution, $R_1(2) = 0.15$.

Example 2 This example is similar to the previous one but this time the value zero was not admissible for c_1. The results are given in Table 3. The optimal value, 0.27, of $R_1(2)$ was reached after two iterations (with $c_1 = 1$ and $c_2 = 0$, respectively) of step 3 of our method.

c_n	i	$R_1(2)$	entropy	gradient
1	0	0.0000	0.5623	0.0679
	1	0.1619	0.5679	0.00146
	2	0.1655	0.5679	1.1178e-08
	3	0.1655	0.5679	-7.2435e-16
0	0	0.1655	-1.1520	5.4523
	1	0.2154	-0.9776	2.0404
	2	0.2604	-0.9262	0.3238
	3	0.2699	-0.9247	0.0018
	4	0.2700	-0.9247	3.7871e-10
	5	0.2700	-0.9247	3.4084e-10

Table 3: Results for Example 2. The initial value $R_1(2) = 0$ is NOT admissible. However the optimal value, $R_1(2) = 0.16552$, obtained for the modified problem with $c_1 = 1$ is admissible for the original problem, so $c_2 = 0$. The optimal solution here is $R_1(2) = 0.27$.

Example 3 In this example the solution is "almost" p.s.d.. It would have been exactly p.s.d. if $R_1(1) = 1$, a difference of only 0.000001 from the value used in this example, see Table 7. The optimal solution for $R_1(2)$ is 0.2999997. Table 4 gives results obtained for this example.

Here the procedure is slower than in Examples 1 and 2 but, bearing in mind the close proximity to singularity, the fact that the solution was obtained with only 5 iterations of step 3 of the method seems an excellent result.

Example 4 This example is similar to Example 1 but it is set up so that the solution is p.s.d.. The value 0.15 is an optimal solution for the missing $R_1(0)$, as in Example 1.

For this singular example the procedure leads to a sequence of c_n values that monotonically decreases to 0. Table 5 gives the results obtained for the first eight values of c_n as well as for its 80th value. For each value of c_n, we give only the results for the last iteration of the Newton-Raphson's algorithm. Table 5 is in agreement with the theory which says that $c_n \to 0$ in the p.s.d. case.

Example 5 Our last example illustrates the case when the given sequence is not completable. Table 6 gives the results obtained for the first eight values of c_n as well as for its 83th value. For each values of c_n, as in the previous example, we give only the results for the last iteration of the Newton-

c_n	i	$R_1(2)$	entropy	gradient
1.000000	0	0.0000000	0.5184	0.1046
	1	0.2152802	0.5300	0.0045
	2	0.2253450	0.5300	3.5921e-07
	3	0.2253458	0.5300	-6.5702e-13
0.050000	0	0.2253458	-1.6121	18.3505
	1	0.2452564	-1.3671	8.2140
	2	0.2738726	-1.2187	2.8691
	3	0.2960202	-1.1841	0.3400
	4	0.2992355	-1.1836	0.0007
	5	0.2992426	-1.1836	-1.7116e-12
0.002500	0	0.2992426	-2.7068	30.9271
	1	0.2999629	-2.6953	1.3467
	2	0.2999965	-2.6953	0.0001
	3	0.2999965	-2.6953	-2.1290e-12
0.000125	0	0.2999965	-4.1905	48.7861
	1	0.2999996	-4.1904	0.0147
	2	0.2999996	-4.1904	5.2170e-10
0.000000	0	0.2999996	-6.6560	16725.18
	1	0.2999997	-6.6554	33.9126

Table 4: Results for Example 3. The solution here is almost p.s.d.. Despite that the algorithm finds an admissible solution after solving 4 modified problems and after only one additional iteration gives the optimal value, $R_1(2) = 0.2999997$.

c_n	i	$R_1(2)$	entropy	gradient
1.000000	3	0.1362538	0.5102	-8.2608e-09
0.050000	3	0.1616193	-1.2162	-2.2817e-10
0.002500	3	0.1511813	-2.7275	3.7530e-10
0.000125	2	0.1500620	-4.2261	4.6311e-12
6.250e-06	2	0.1500031	-5.7240	-1.5291e-13
3.125e-07	2	0.1500002	-7.2219	-6.4537e-07
1.563e-08	1	0.1500000	-8.7197	-5.8958e-08
7.813e-10	1	0.1500000	-10.2176	-9.5701e-09
...	...	...	...	...
7.994e-16	2	0.1500000	-17.1731	-0.0012

Table 5: Results for Example 4. The solution here is p.s.d.. In this case c_n converges to 0 but can never be set equal to 0. In floating point calculations this cannot be detected exactly but the decision rules used for our examples stopped the algorithm after solving the 80th modified problem (with $c_{80} = 7.994 \cdot 10^{-16}$) and declared that the solution is p.s.d..

Raphson's algorthm. The procedure converges quickly to the value $c = 1.069440 > 0$, which shows that no solution exists.

Remark on p.s.d. solutions. From numerical viewpoint there is little distinction between the "exact" and "almost" p.s.d. case. One might expect erratic behaviour of numerical algorithms in these cases. It is therefore interesting that our method not only successfully deals with them but it also distinguishes the two cases.

For example, in the singular case illustrated Example 4 the solution changes very little from about c_{15}, but it does change. On the other hand, the "almost" p.s.d. case of Example 3 reached the value $c_n = 0$ for $n = 5$ iterations and finished successfully. We run Examples 3–5, and many other similar examples, for far more iterations that would be needed in practice in order to get a better understanding of the behaviour of the method. Our results were qualitatively similar—the method finishes successfully and fails to distinguish "almost" from "exact" p.s.d. only when the solution is extremely close to being p.s.d.. Note that the "distance" of 10^{-6} from singularity in Example 3 is not extremely close in this context. In some examples it seemed not possible to make the gradient at the declared solution sufficiently small. However, the huge value of the gradient in examples like Example 5 is not surprising since as we are approaching a p.s.d. solution the entropy diverges to $-\infty$, see Appendix C. It may be better to maximize e^h in such cases and

c_n	i	$R_1(2)$	entropy	gradient
2.000000	4	0.4514833	0.7068	-7.8642e-08
1.073418	4	0.4477606	-2.9850	-1.9047e-05
1.070620	8	0.4431905	-4.1350	-0.3180
1.069847	6	0.4417908	-5.1751	-4.8700e-05
1.069569	8	0.4412777	-6.2893	-9.9963e-05
1.069480	7	0.4411105	-7.3921	-12.6808
1.069451	6	0.4410558	-8.4504	-14.0478
1.069440	1	0.4410556	-11.1154	-1077458
...	...	...	...	...
1.069440	1	0.4410555	-20.8842	-3.3077e+14

Table 6: Results for Example 5. The sequence c_n converges to a positive value since the maximum entropy problem has no solution here.

we did some limited experiments but the results are similar.

So, we may confidently conclude that our method is very reliable even in cases close to singularity. Since it also detects when the completion problem has no solution, it is a complete all round method.

A Proofs of theorems and lemmas

A.1 Proof of Lemma 2

We need to show that any convex linear combination of admissible gap fillings is also admissible gap filling. Indeed, let $w^{(1)}$ and $w^{(2)}$ be two admissible gap fillings. Then $K_1 = K \cup w^{(1)}$ and $K_2 = K \cup w^{(2)}$ are positive definite.

Let $w^{()} = \alpha w^{(1)} + \beta w^{(2)}$, where $\alpha + \beta = 1$ and $\alpha, \beta \geq 0$. Obviously, $K_3 = K \cup w^{()} = \alpha(K \cup w^{(1)}) + \beta(K \cup w^{(2)})$. It remains to show that K_3 is autocovariance sequence. Indeed, let $\{X_t\}$ be a process with acf K_1, $\{Y_t\}$ a process with acf K_2, and θ a random variable taking values 1 and 0 with probability α and β, respectively, and independent of $\{X_t\}$ and $\{Y_t\}$. Consider the process $\{V_t\}$ defined by the equation

$$V_t = \theta X_t + (1 - \theta) Y_t.$$

The autocovariance function of V_t is equal to $\alpha(K \cup w^{(1)}) + \beta(K \cup w^{(2)})$. Hence $w^{()}$ is admissible.

A.2 Proof of Lemma 3

Parts (1) and (2) are obtained as follows.

$$\begin{aligned}
R_v(t,k) &= E(\theta X_t + (1-\theta)Y_t)(\theta X_{t-k} + (1-\theta)Y_{t-k}) \\
&= E(\theta^2 X_t X_{t-k} + (1-\theta)^2 Y_t Y_{t-k}) \\
&= E(\theta^2)E(X_t X_{t-k}) + E(1-\theta)^2 E(Y_t Y_{t-k}) \\
&= \alpha R_X(t,k) + \beta R_Y(t,k) \qquad \text{(proving part 1).}
\end{aligned}$$

$$\begin{aligned}
\sigma^2_{t,V} &= \mathrm{Var}(V_t - \hat{V}_t) \\
&= E(V_t - \hat{V}_t)^2 \\
&= E(E((V_t - \hat{V}_t)^2|\theta)) \\
&= \alpha E((V_t - \hat{V}_t)^2|\theta = 1) + \beta E((V_t - \hat{V}_t)^2|\theta = 0) \\
&\geq \alpha E((V_t - \sum_i \alpha^X_{t,i} V_{t-i})^2|\theta = 1) + \beta E((V_t - \sum_i \alpha^Y_{t,i} V_{t-i})^2|\theta = 0) \\
&= \alpha\sigma^2_{t,X} + \beta\sigma^2_{t,Y}.
\end{aligned}$$

The inequality above will be strict unless $\alpha^V_{t,i} = \alpha^X_{t,i} = \alpha^Y_{t,i}$ for all i, as claimed in part 2.

(3) Taking natural logarithms on both sides of the inequality in part (2) of the lemma we get

$$\log(\sigma^2_{t,V}) \geq \log\left(\alpha\sigma^2_{t,X} + \beta\sigma^2_{t,Y}\right). \tag{11}$$

On the other hand, the generalised inequality between the arithmetic and geometric means gives

$$\alpha\sigma^2_{t,X} + \beta\sigma^2_{t,Y} \geq \exp(\alpha\log(\sigma^2_{t,X}) + \beta\log(\sigma^2_{t,Y})),$$

where the inequality is strict unless $\sigma^2_{t,X} = \sigma^2_{t,Y}$. Taking logarithms from both sides gives

$$\log(\alpha\sigma^2_{t,X} + \beta\sigma^2_{t,Y}) \geq \alpha\log(\sigma^2_{t,X}) + \beta\log(\sigma^2_{t,Y}). \tag{12}$$

From Equations (11) and (12) we get the required inequality. The inequality becomes equality if and only if equality holds in both equations, which in turn happens if and only if both conditions in part (3) are met.

A.3 Proof of Lemma 4

Indeed, let $w^{(1)}$ and $w^{(2)}$ be two different gap fillings and $\omega = \alpha w^{(1)} + \beta w^{(2)}$, where α and β are positive with $\alpha + \beta = 1$. Then by Lemma 1 $h_{\mathrm{ME}}(w^{(1)})$

and $h_{\mathrm{ME}}(w^{(2)})$ correspond to PAR completions. From Corollary 1 we obtain that

$$\frac{1}{T}\sum_{t=t_0+1}^{t_0+T}\log(\sigma^2_{t,\omega}) > \alpha\frac{1}{T}\sum_{t=t_0+1}^{t_0+T}\log(\sigma^2_{t,w^{(1)}}) + \beta\frac{1}{T}\sum_{t=t_0+1}^{t_0+T}\log(\sigma^2_{t,w^{(2)}}),$$

where the inequality is strict since the two autocovariance sequences are different. For sufficiently large t_0 the two sums on the right-hand side are equal to T times $h_{\mathrm{ME}}(w^{(1)})$ and $h_{\mathrm{ME}}(w^{(2)})$, respectively (see Equation (3)). So, for large t_0,

$$\frac{1}{T}\sum_{t=t_0+1}^{t_0+T}\log(\sigma^2_{t,\omega}) > \alpha h_{\mathrm{ME}}(w^{(1)}) + \beta h_{\mathrm{ME}}(w^{(2)}),$$

The left-hand side is smaller or equal to $h_{\mathrm{ME}}(\omega)$ since ω is a completion. Hence,

$$h_{\mathrm{ME}}(\omega) > \alpha h_{\mathrm{ME}}(w^{(1)}) + \beta h_{\mathrm{ME}}(w^{(2)}),$$

as required for strict concavity.

A.4 Proof of Theorem 2

If K^{c_1} is p.d.-completable, then K^{c_2} is p.d.-completable for any $c_2 > c_1$. Indeed, suppose that K^{c_1} is p.d.-completable and c_2 is any number greater than c_1. Let γ^{c_1} be any completion of K^{c_1}. Let $\{X_t\}$ be a process with acf γ^{c_1}. Let $\{u_t\}$ be white noise with mean 0 and variance $c_2 - c_1$. Define a process $\{Y_t\}$ by $Y_t = X_t + u_t$. The autocovariance function of $\{Y_t\}$ is positive definite and is given by the equation

$$\begin{cases} R_t(0) + c_1 + (c_2 - c_1) & \text{for } k = 0, \\ \gamma_t^{c_1}(k) & \text{for } k \neq 0. \end{cases}$$

So, the autocovariance function of $\{Y_t\}$ is a positive definite completion of K^{c_2}. Hence, K^{c_2} is p.d.-completable.

A similar argument can be used to show that if K^{c_1} is not p.d.-completable, then K^{c_2} is not p.d.-completable for any $c_2 < c_1$. The set of c's for which K^c is p.d.-completable is bounded from below since the lag 0 autocovariances, $R_t(0) + c$, must be positive.

Let $\omega = \inf_c \{\, c \mid K^c \text{ is p.d.-completable} \,\}$. Then trivially K^c is not completable for $c < \omega$. If we assume that there is a $c > \omega$ such that K^c is not completable, then using the definition of ω as an infimum we can find c_1 such

that $\omega < c_1 < c$ and K^{c_1} is p.d.-completable. But then K^c must be p.d.-completable too since $c > c_1$, a contradiction. Hence, K^c is p.d.-completable for $c > \omega$.

To prove the last claim of the theorem, i.e. that K^c is p.s.d.-completable for $c = \omega$, first note that K^ω is at least p.s.d. since a p.s.d. completion can be obtained as a limit of a convergent sequence of p.d. completions. Assume that K^ω is p.d.-completable and let γ be a p.d. completion of K^ω. Since the entropy rate is continuous with respect to $R(t,0)$ on the set of positive definite completions we can find sufficiently small ε such that the sequence $\{r_t(k)\}$

$$r_t(k) = \begin{cases} \gamma_t(0) - \varepsilon & \text{for } k = 0, \\ \gamma_t(k) & \text{for } k \neq 0. \end{cases}$$

is p.d.. But $\gamma_t(0) - \varepsilon = R_t(0) + \omega - \varepsilon$ and hence $\{r_k\}$ is a p.d. completion for $K^{\omega-\varepsilon}$. This contradicts the choice of ω. So, K^ω is not p.d..

A.5 Proof of Lemma 5

First let us remark that for $\alpha+\beta = 1$, $\alpha, \beta \in [0,1]$, $\alpha K^{c_1} + \beta K^{c_2} = K^{\alpha c_1 + \beta c_2}$. So for c_1, $c_2 \in (\omega, \infty)$, Lemma 4 leads to

$$\alpha h_{\mathrm{ME}}(K^{c_1}) + \beta h_{\mathrm{ME}}(K^{c_2}) < h_{\mathrm{ME}}(\alpha K^{c_1} + \beta K^{c_2}) = h_{\mathrm{ME}}(K^{\alpha c_1 + \beta c_2}).$$

That is the function that associates $h_{\mathrm{ME}}(K^c)$ to c is strictly concave. As a consequence, this function is also continuous at any point in (ω, ∞).

A.6 Proof of Lemma 6

We use the following lemma in the proof of Lemma 6.

Lemma 8. *Let $\{X_t\}$ and $\{U_t\}$, $t \in \mathbb{Z}$, be two uncorrelated processes (possibly non-stationary), i.e. $\mathrm{E}\, X_t U_s = 0$ for all $t \neq s$. Let also $Y_t = X_t + U_t$. Then*

$$\sigma^2_{Y,t} \geq \sigma^2_{X,t} + \sigma^2_{U,t}, \quad t \in \mathbb{Z},$$

where $\sigma^2_{Y,t}$, $\sigma^2_{X,t}$, and $\sigma^2_{U,t}$ are the variances of the innovation processes of $\{Y_t\}$, $\{X_t\}$, and $\{U_t\}$, respectively.

Proof of Lemma 8. Let $\mathcal{M}^Z(t)$ be the closed linear subspace $\overline{sp}\{Z_s, s < t\}$ associated to any process $\{Z_t\}$. Here it is convenient to use a geometrical approach by considering the inner product $\langle U, V\rangle = E\{UV\} = Cov\{U,V\}$. The norm associated with this inner product is defined by $\|U\|^2 = Var(U)$.

We denote by $P_{\mathcal{M}}(Z_t)$ the orthogonal projection (for this inner product) of the variable Z_t on the space $\mathcal{M}$.

Since any element of $\mathcal{M}^Y(t)$ is a linear combination of $X_s + U_s$, $s < t$, we have

$$\mathcal{M}^Y(t) \subset \mathcal{M}^X(t) \oplus \mathcal{M}^U(t),$$

and

$$\begin{aligned} \|Y_t - P_{\mathcal{M}^X(t)\oplus\mathcal{M}^U(t)}(Y_t)\|^2 &= \inf_{v\in\mathcal{M}^X(t)\oplus\mathcal{M}^U(t)} \|Y_t - v\|^2 \\ &\leq \inf_{v\in\mathcal{M}^Y(t)} \|Y_t - v\|^2 \\ &\leq \sigma^2_{Y,t}. \end{aligned} \tag{13}$$

Furthermore, since $\{X_t\}$ and $\{U_t\}$ uncorrelated processes and the spaces $\mathcal{M}^X(t)$ and $\mathcal{M}^U(t)$ are orthogonal, we obtain

$$\begin{aligned} P_{\mathcal{M}^X(t)\oplus\mathcal{M}^U(t)}(Y_t) &= P_{\mathcal{M}^X(t)}(Y_t) + P_{\mathcal{M}^U(t)}(Y_t) \\ &= P_{\mathcal{M}^X(t)}(X_t) + P_{\mathcal{M}^U(t)}(U_t). \end{aligned}$$

This leads to

$$\begin{aligned} \|Y_t - P_{\mathcal{M}^X(t)\oplus\mathcal{M}^U(t)}(Y_t)\|^2 &= \|X_t - P_{\mathcal{M}^X(t)}(X_t)\|^2 + \|U_t - P_{\mathcal{M}^U(t)}(U_t)\|^2, \\ &= \sigma^2_{X,t} + \sigma^2_{U,t}, \end{aligned}$$

and the inequality (13) gives the result. □

Proof of Lemma 6. Let c_1 and c_2 be two reals in (ω, ∞) such that $c_1 < c_2$. Recall that $\gamma^{c_1,\mathrm{ME}}$ is the acf associated with the entropy rate $h_{\mathrm{ME}}(K^{c_1})$. Let $\{X_t\}$ be a process with acf $\gamma^{c_1,\mathrm{ME}}$. Let $\{U_t\}$ be a white noise uncorrelated with $\{X_t\}$, with mean 0 and variance $\sigma^2_{U,t} = c_2 - c_1 > 0$. Let us define a process $\{Y_t\}$ by $Y_t = X_t + U_t$. Then the acf γ^{c_2} of $\{Y_t\}$ is p.d.and is given by

$$\gamma_t^{c_2}(k) = \begin{cases} R_t(0) + c_1 + (c_2 - c_1) & \text{for } k = 0, \\ \gamma_t^{c_1,\mathrm{ME}}(k) & \text{for } k \neq 0. \end{cases}$$

So γ^{c_2} is a p.d. completion of K^{c_2} and its entropy rate, $h(\gamma^{c_2})$, does not exceed the maximum possible, i.e.

$$h(\gamma^{c_2}) \leq h_{\mathrm{ME}}(K^{c_2}). \tag{14}$$

Note that $h(\gamma^{c_2})$ is also the entropy rate of the process $\{Y_t\}$. Furthermore, Lemma 8 shows that

$$\sigma^2_{Y,t} \geq \sigma^2_{X,t} + \sigma^2_{U,t}, \quad t = 1, \ldots T.$$

But the variance, $\sigma^2_{U,t}$, of $\{U_t\}$ is strictly positive. Hence,

$$\sigma^2_{Y,t} > \sigma^2_{X,t}, \quad t = 1, \dots T.$$

Together with Equation (3) this shows that

$$h(\gamma^{c_2}) > h(\gamma^{c_1,\mathrm{ME}}) = h_{\mathrm{ME}}(K^{c_1}).$$

The last inequality together with Equation (14) gives $h_{\mathrm{ME}}(K^{c_1}) < h_{\mathrm{ME}}(K^{c_2})$, as required.

□

A.7 Proof of Lemma 7

For any i the sequence $\gamma^{c_i,\mathrm{ME}}$ is the acf of a PAR($\leq p_1$,...,$\leq p_T$) (see Section 6.2). Also, the coefficients $\gamma_t^{c_i,\mathrm{ME}}(k) = R_t(k)$ are fixed when $(t,k) \in I$. So, it is sufficient to consider $\gamma_t^{c_i,\mathrm{ME}}(k)$ for $(t,k) \in E_c(I) \setminus I$ only. Below we use the fact that the set $E_c(I) \setminus I$ is finite.

Let $\{c_i\}$ be any sequence that converges to $c \in (\omega, \infty)$. Then the sequence $\{c_i\}$ is bounded and $\gamma_t^{c_i,\mathrm{ME}}(0) = R_t(0) + c_i$. Since $\gamma_t^{c_i,\mathrm{ME}}(0) \leq \max_{t\in\{1,\dots,T\}} R_t(0) + \max_i c_i \equiv M$, it follows that $\gamma_t^{c_i,\mathrm{ME}}(k)$ is bounded as well. Therefore $|\gamma_t^{c_i,\mathrm{ME}}(k)| \leq \gamma_t^{c_i,\mathrm{ME}}(0) \leq M$ for all t, k and c_i. If we now assume that $\gamma_t^{c_i,\mathrm{ME}}(k)$ does not converge to $\gamma_t^{c,\mathrm{ME}}(k)$ for $(t,k) \in E_c(I) \setminus I$, then there exist subsequences converging to different limits. However, the entropies of these (different) limiting sequences should be the same since $h_{\mathrm{ME}}(K^{c_i})$ converges to $h_{\mathrm{ME}}(K^c)$ by the continuity claim in Lemma 5. Thus we would have several sequences having the maximum entropy rate in K^c, a contradiction with the uniqueness (see Lemma 4 and [13, p. 123, Theorem A]).

A.8 Two technical lemmas

Lemma 9. *Let γ be a positive definite periodic acf of a PAR(p_1,...,p_T) model. There exists a constant ε such that any sequence r defined for $t = 1, \dots, T$, $k = 0, 1, \dots, p_t$, and satisfying $|r_t(k) - \gamma_t(k)| < \varepsilon$ is p.d.-PAR($p_1, \dots, p_T$) completable.*

Proof of Lemma 9. γ is p.d.-PAR($p_1, \dots, p_T$) completable if and only if the corresponding prediction error variances are positive. The latter depend continuously on γ as can be seen from the periodic Yule-Walker equations and the PLD algorithm. Therefore small changes in the autocovariances will leave the prediction error variances positive, i.e. the changed autocovariance sequence will be p.d.-PAR($p_1, \dots, p_T$) completable. Standard reasoning then leads to the claim of the lemma. □

We will not necessarily get a p.d. sequence if we replace $\gamma_t^{c,\mathrm{ME}}(k)$ on the last line of Equation (10) by the maximum entropy solution for another c. But the p.d. property holds in a neighbourhood of c as Lemma 10 shows.

Lemma 10. *For each c there exists a neighbourhood $(c-\varepsilon_0, c+\varepsilon_0)$ such that the function*

$$\begin{cases} R_t(0) + q(c) + a & \text{for } k = 0, \\ R_t(k) & \text{for } k \neq 0 \text{ and } (t,k) \in I, \end{cases}$$

is p.d.-PAR($p_1, \ldots, p_T$) completable for any ε such that $|\varepsilon| < \varepsilon_0$.

Proof of Lemma 10. This is a corollary of Lemma 9. □

A.9 Proof ot Theorem 3

We will prove the claim by contradiction. Assume that $d^* = 0$ and note that we consider c's over a finite interval. Hence, there exists a convergent sequence $\{c_k\}$ such that $d_k \to 0$. Let $c^* = \lim_{k\to\infty} c_k$. Let $\gamma^{c_k,\mathrm{ME}}$ be the ME autocovariance sequence corresponding to c_k. Lemma 7 shows that $\gamma^{c_k,\mathrm{ME}} \to \gamma^{c^*,\mathrm{ME}}$. Let $\delta = (c^* - q(c^*))/2$ where $q(\cdot)$ is as defined above corresponding to $\gamma^{c^*,\mathrm{ME}}$. Then

$$\begin{cases} R_t(0) + q(c^*) + \delta & \text{for } k = 0, \\ R_t(k) & \text{for } k \neq 0 \text{ and } (t,k) \in I, \\ \gamma_t^{c^*,\mathrm{ME}}(k) & k \in \tau^c(t) \end{cases} \tag{15}$$

is p.d.-PAR($p_1, \ldots, p_T$) completable. By Lemma 9 for sufficiently large k

$$\begin{cases} R_t(0) + q(c^*) + \delta & \text{for } k = 0, \\ R_t(k) & \text{for } k \neq 0 \text{ and } (t,k) \in I, \\ \gamma_t^{c_k,\mathrm{ME}}(k) & k \in \tau^c(t) \end{cases} \tag{16}$$

is also p.d.-PAR($p_1, \ldots, p_T$) completable.

On the other hand, $c_k \to c^*$ and $d_k = c_k - q(c_k) \to 0$. Hence $q(c_k) > q(c^*) + \delta$ for sufficiently large k.

So, $q(c^*) + \delta$ is smaller than $q(c_k)$ and

$$\begin{cases} R_t(0) + q(c^*) + \delta & \text{for } k = 0, \\ R_t(k) & \text{for } k \neq 0 \text{ and } (t,k) \in I, \\ \gamma_t^{c_k,\mathrm{ME}}(k) & k \in \tau^c(t) \end{cases} \tag{17}$$

is p.d.-PAR($p_1, \ldots, p_T$) completable, a contradiction with the definition of $q()$.

B PLD algorithm

Here we give the formulae for the calculation of the elements of the gradient and the Hessian matrix of the entropy rate which are needed for the Newton-Raphson's algorithm. In this section we use somewhat different notation from the rest of the paper since there are more quantities than we needed in the discussion of the theory. We denote by $\sigma_k^{f2}(n)$ and $\sigma_k^{b2}(n)$ the variances of the nth-order forward and backward partial innovations $\varepsilon_k^f(n)$ and $\varepsilon_k^b(n)$, defined by

$$\varepsilon_k^f(n) = \sum_{j=0}^{n} a_k^f(n,j) X_{k-j}, \quad a_k^f(n,0) = 1,$$
$$\varepsilon_k^b(n) = \sum_{j=0}^{n} a_k^b(n,j) X_{k-n+j}, \quad a_k^b(n,0) = 1.$$

where the filters $\{a_k^f(n,\cdot)\}$ and $\{a_k^b(n,\cdot)\}$ can be determined recursively in n, see below. So, $\sigma_k^{f2}(n)$ replaces the notation $v_k(n)$ for the variance of the prediction error introduced in Section 2. With this notation the expression for the entropy rate is

$$h(R) = \frac{1}{T} \sum_{k=1}^{T} \log \sigma_k^{f2}(p_k).$$

Its first and second derivatives are

$$\begin{aligned} \frac{\partial h(R)}{\partial R_t(l)} &= \frac{1}{T} \sum_{k=1}^{T} \frac{\frac{\partial \sigma_k^{f2}(p_k)}{\partial R_t(l)}}{\sigma_k^{f2}(p_k)}, \\ \frac{\partial^2 h(R)}{\partial R_t(l) \partial R_s(m)} &= \frac{1}{T} \sum_{k=1}^{T} \left[\frac{\frac{\partial^2 \sigma_k^{f2}(p_k)}{\partial R_t(l) \partial R_s(m)}}{\sigma_k^{f2}(p_k)} - \frac{\frac{\partial \sigma_k^{f2}(p_k)}{\partial R_t(l)} \frac{\partial \sigma_k^{f2}(p_k)}{\partial R_s(m)}}{(\sigma_k^{f2}(p_k))^2} \right]. \end{aligned}$$

for $t, s = 1, \ldots, T,\ l, m \in \tau_g$. So we need to compute the first and second derivatives of the variances of the innovations with respect to the acf coefficients corresponding to gaps. Analytic closed form expressions of these variances in terms of the acf coefficients are not available but they can be calculated recursively using the Periodic Levinson-Durbin (PLD) algorithm (see [9]). The PLD algorithm computes, besides the variances, the filters $\{a_k^f(n,\cdot)\}$ and $\{a_k^b(n,\cdot)\}$ which are also need for the calculation of the derivatives.

This algorithm is based in part on the computation of the filters $a_k^f(n,\cdot)$ and $a_k^b(n,\cdot)$ given by,

$$\varepsilon_k^f(n) = \sum_{j=1}^{n} a_k^f(n,j) X_{k-j}, \quad a_k^f(n,0) = 1,$$

$$\varepsilon_k^b(n) = \sum_{j=1}^{n} a_k^b(n,j) X_{k-n+j}, \quad a_k^b(n,0) = 1$$

which are determined recursively.

We differentiate the PLD equations to develop PLD-type recursions for the first and second order derivatives.

In the formulae below we use the convention $\sum_{j=1}^{0} \ldots = 0$. Also, when the subscript $k-1$ is equal to 0, it is replaced by T. $\delta_{i,j}$ is the Kronecker symbol, that is $\delta_{i,j} = 1$ if $i = j$, $\delta_{i,j} = 0$ otherwise. We denote by $t[T]$ the integer r_t in $[1,\ldots,T]$ such that $t = q_t T + r_t$ and $q_t \in \mathbb{Z}$.

Algorithm 1 summarises the calculations of the derivatives. It does not include formulae since many of them are quite long. The formulae are given in the following subsections and referred to in Algorithm 1.

B.1 Initialisation step

For $k = 1,\ldots,T$, $t = 1,\ldots,T$, $l,m \in \tau_g$, set the following initial values.

$$\beta_k(0) = \sigma_k^{f2}(0) = \sigma_k^{b2}(0) = R_k(0) \tag{18}$$

$$\frac{\partial \sigma_k^{f2}(0)}{\partial R_t(l)} = \frac{\partial \sigma_k^{b2}(0)}{\partial R_t(l)} = \frac{\partial^2 \sigma_k^{f2}(0)}{\partial R_t(l) \partial R_s(m)} = \frac{\partial^2 \sigma_k^{b2}(0)}{\partial R_t(l) \partial R_s(m)} = 0 \tag{19}$$

B.2 PLD step for the parameters

These are the standard PLD calculations, see [9].

The equations for step n are:

$$A_k(n) = R_k(n) + \sum_{j=1}^{n-1} a_k^f(n-1,j) R_{(k-j)[T]}(n-j), \tag{20a}$$

$$\beta_k(n) = \frac{A_k(n)}{\sigma_k^f(n-1)\sigma_{k-1}^b(n-1)}, \tag{20b}$$

Algorithm 1: Outline of the calculations of first and second order derivatives based on the PLD algorithm. An entry like Equations XX stands for the group of equations numbered (XXa), (XXb), ..., while Equation XX stands for a single equation.

```
for k = 1 to T do ;                                   // Initialisation step

    Equation (18)
    for t = 1 to T, and l, m ∈ τ_g do
        Equation (19)
for n = 1 to max_{i=1,...T} p_i do ;                  // Main loop

    for k = 1 to T, and n ≤ p_k do
        Equations (20)
        if  n > 1 then   for j = 1 to n − 1 do
            Equations (21)
        Equations (22)
        if  n > 1 then   for j = 1 to n − 1 do
            Equations (23)
        for t, s = 1 to T and l, m ∈ τ_g do
            if l > n ∨ m > n ∨ (k = t ∧ n = l) ∨ (k = s ∧ n = m) then
                Equation (24a)
            else
                Equation (24b)
            Equations (25)
        Equations (26)
```

$$\sigma_k^{f2}(n) = \sigma_k^{f2}(n-1) - \frac{A_k(n)^2}{\sigma_{k-1}^{b2}(n-1)}, \tag{20c}$$

$$\sigma_k^{b2}(n) = \sigma_{k-1}^{b2}(n-1) - \frac{A_k(n)^2}{\sigma_k^{f2}(n-1)}, \tag{20d}$$

$$a_k^f(n,n) = -\frac{A_k(n)}{\sigma_{k-1}^{b2}(n-1)}, \tag{20e}$$

$$a_k^b(n,n) = -\frac{A_k(n)}{\sigma_k^{f2}(n-1)}, \tag{20f}$$

The following calculations are done after the above when $n > 1$.

$$a_k^f(n,j) = a_k^f(n-1,j) + a_k^f(n,n)a_{k-1}^b(n-1,n-j), \tag{21a}$$

$$a_k^b(n,j) = a_{k-1}^b(n-1,j) + a_k^b(n,n)a_k^f(n-1,n-j) \tag{21b}$$

$$\text{(for } j = 1,\ldots,n-1.)$$

B.3 Gradient PLD step

For $t = 1,\ldots,T,\ l \in \tau_g$,

$$\frac{\partial A_k(n)}{\partial R_t(l)} = \begin{cases} 0, & \text{when } l > n, \\ 1, & \text{when } k = t \text{ and } n = l, \\ \delta_{t,(k-n+l)[T]} a_k^f(n-1,n-l) & \\ \quad + \sum_{j=1}^{n-1} \frac{\partial a_k^f(n-1,j)}{\partial R_t(l)} R_{(k-j)[T]}(n-j), & \text{otherwise.} \end{cases} \tag{22a}$$

$$\frac{\partial \sigma_k^{f2}(n)}{\partial R_t(l)} = \frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)} - \frac{2A_k(n)\frac{\partial A_k(n)}{\partial R_t(l)}}{\sigma_{k-1}^{b2}(n-1)} + \frac{A_k(n)^2 \frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^2}, \tag{22b}$$

$$\frac{\partial \sigma_k^{b2}(n)}{\partial R_t(l)} = \frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)} - \frac{2A_k(n)\frac{\partial A_k(n)}{\partial R_t(l)}}{\sigma_k^{f2}(n-1)} + \frac{A_k(n)^2 \frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^2}, \tag{22c}$$

$$\frac{\partial a_k^f(n,n)}{\partial R_t(l)} = -\frac{\frac{\partial A_k(n)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))} + \frac{A_k(n)\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^2}, \tag{22d}$$

$$\frac{\partial a_k^b(n,n)}{\partial R_t(l)} = -\frac{\frac{\partial A_k(n)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))} + \frac{A_k(n)\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^2}, \tag{22e}$$

And, if $n > 1$, for $j = 1, \ldots, n-1$,

$$\frac{\partial a_k^f(n,j)}{\partial R_t(l)} = \frac{\partial a_k^f(n-1,j)}{\partial R_t(l)} + \frac{\partial a_k^f(n,n)}{\partial R_t(l)} a_{k-1}^b(n-1,n-j) + a_k^f(n,n)\frac{\partial a_{k-1}^b(n-1,n-j)}{\partial R_t(l)}, \tag{23a}$$

$$\frac{\partial a_k^b(n,j)}{\partial R_t(l)} = \frac{\partial a_{k-1}^b(n-1,j)}{\partial R_t(l)} + \frac{\partial a_k^b(n,n)}{\partial R_t(l)} a_k^f(n-1,n-j) + a_k^b(n,n)\frac{\partial a_k^f(n-1,n-j)}{\partial R_t(l)}. \tag{23b}$$

B.4 Hessian PLD step

For $t, s = 1, \ldots, T,\ l, m \in \tau_g$,

If $l > n$ or $m > n$ or ($k = t$ and $n = l$) or ($k = s$ and $n = m$), put

$$\frac{\partial^2 A_k(n)}{\partial R_t(l)\partial R_s(m)} = 0, \tag{24a}$$

Otherwise put

$$\frac{\partial^2 A_k(n)}{\partial R_t(l)\partial R_s(m)} = \delta_{t,(k-n+l)[T]}\frac{\partial a_k^f(n-1,n-l)}{\partial R_s(m)} + \delta_{s,(k-n+m)[T]}\frac{\partial a_k^f(n-1,n-m)}{\partial R_t(l)} + \sum_{j=1}^{n-1}\frac{\partial^2 a_k^f(n-1,j)}{\partial R_t(l)\partial R_s(m)} R_{(k-j)[T]}(n-j) \tag{24b}$$

$$\frac{\partial^2 a_k^f(n,n)}{\partial R_t(l)\partial R_s(m)} = \frac{\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_s(m)}\frac{\partial A_k(n)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^2} - \frac{\frac{\partial^2 A_k(n)}{\partial R_t(l)\partial R_s(m)}}{(\sigma_{k-1}^{b2}(n-1))} + \frac{\frac{\partial A_k(n)}{\partial R_s(m)}\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^2} - 2\frac{A_k(n)\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_s(m)}\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^3} + \frac{A_k(n)\frac{\partial^2 \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)\partial R_s(m)}}{(\sigma_{k-1}^{b2}(n-1))^2}, \tag{25a}$$

$$\frac{\partial^2 a_k^b(n,n)}{\partial R_t(l)\partial R_s(m)} = \frac{\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_s(m)}\frac{\partial A_k(n)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^2} - \frac{\frac{\partial^2 A_k(n)}{\partial R_t(l)\partial R_s(m)}}{(\sigma_k^{f2}(n-1))} + \frac{\frac{\partial A_k(n)}{\partial R_s(m)}\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^2} - 2\frac{A_k(n)\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_s(m)}\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^3} + \frac{A_k(n)\frac{\partial^2 \sigma_k^{f2}(n-1)}{\partial R_t(l)\partial R_s(m)}}{(\sigma_k^{f2}(n-1))^2} \tag{25b}$$

$$\frac{\partial^2 \sigma_k^{f2}(n)}{\partial R_t(l)\partial R_s(m)} = \frac{\partial^2 \sigma_k^{f2}(n-1)}{\partial R_t(l)\partial R_s(m)} - 2\frac{\frac{\partial A_k(n)}{\partial R_s(m)}\frac{\partial A_k(n)}{\partial R_t(l)}}{\sigma_{k-1}^{b2}(n-1)} - 2\frac{A_k(n)\frac{\partial^2 A_k(n)}{\partial R_t(l)\partial R_s(m)}}{\sigma_{k-1}^{b2}(n-1)}$$

$$- 2\frac{A_k(n)^2\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^3} + 2\frac{A_k(n)\frac{\partial A_k(n)}{\partial R_t(l)}\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_s(m)}}{(\sigma_{k-1}^{b2}(n-1))^2} \tag{25c}$$

$$+ 2\frac{A_k(n)\frac{\partial A_k(n)}{\partial R_s(m)}\frac{\partial \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)}}{(\sigma_{k-1}^{b2}(n-1))^2} + \frac{A_k(n)^2\frac{\partial^2 \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)\partial R_s(m)}}{(\sigma_{k-1}^{b2}(n-1))^2}, \tag{25d}$$

$$\frac{\partial^2 \sigma_k^{b2}(n)}{\partial R_t(l)\partial R_s(m)} = \frac{\partial^2 \sigma_{k-1}^{b2}(n-1)}{\partial R_t(l)\partial R_s(m)} - 2\frac{\frac{\partial A_k(n)}{\partial R_s(m)}\frac{\partial A_k(n)}{\partial R_t(l)}}{\sigma_k^{f2}(n-1)} - 2\frac{A_k(n)\frac{\partial^2 A_k(n)}{\partial R_t(l)\partial R_s(m)}}{\sigma_k^{f2}(n-1)}$$

$$- 2\frac{A_k(n)^2\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^3} + 2\frac{A_k(n)\frac{\partial A_k(n)}{\partial R_t(l)}\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_s(m)}}{(\sigma_k^{f2}(n-1))^2} \tag{25e}$$

$$+ 2\frac{A_k(n)\frac{\partial A_k(n)}{\partial R_s(m)}\frac{\partial \sigma_k^{f2}(n-1)}{\partial R_t(l)}}{(\sigma_k^{f2}(n-1))^2} + \frac{A_k(n)^2\frac{\partial^2 \sigma_k^{f2}(n-1)}{\partial R_t(l)\partial R_s(m)}}{(\sigma_k^{f2}(n-1))^2}, \tag{25f}$$

And, if $n > 1$, for $j = 1, \ldots, n-1$,

$$\frac{\partial^2 a_k^f(n,j)}{\partial R_t(l)\partial R_s(m)} = \frac{\partial^2 a_k^f(n-1,j)}{\partial R_t(l)\partial R_s(m)} + a_k^f(n,n)\frac{\partial^2 a_{k-1}^b(n-1,n-j)}{\partial R_t(l)\partial R_s(m)}$$

$$+ a_{k-1}^b(n-1,n-j)\frac{\partial^2 a_k^f(n,n)}{\partial R_t(l)\partial R_s(m)} + \frac{\partial a_k^f(n,n)}{\partial R_t(l)}\frac{\partial a_{k-1}^b(n-1,n-j)}{\partial R_s(m)}$$

$$+ \frac{\partial a_k^f(n,n)}{\partial R_s(m)}\frac{\partial a_{k-1}^b(n-1,n-j)}{\partial R_t(l)}, \tag{26a}$$

$$\frac{\partial^2 a_k^b(n,j)}{\partial R_t(l)\partial R_s(m)} = \frac{\partial^2 a_{k-1}^b(n-1,j)}{\partial R_t(l)\partial R_s(m)} + a_k^b(n,n)\frac{\partial^2 a_k^f(n-1,n-j)}{\partial R_t(l)\partial R_s(m)}$$

$$+ a_k^f(n-1,n-j)\frac{\partial^2 a_k^b(n,n)}{\partial R_t(l)\partial R_s(m)} + \frac{\partial a_k^b(n,n)}{\partial R_t(l)}\frac{\partial a_k^f(n-1,n-j)}{\partial R_s(m)}$$

$$+ \frac{\partial a_k^b(n,n)}{\partial R_s(m)}\frac{\partial a_k^f(n-1,n-j)}{\partial R_t(l)}. \tag{26b}$$

C Gradient of e^{entropy}

Let h be the entropy rate and consider e^h. Maximising h and e^h is equivalent. Let y be any of the parameters w.r.t. which we are maximising. We have $\frac{\partial}{\partial y}e^h = e^h\frac{\partial h}{\partial y}$. Hence, $\frac{\partial h}{\partial y} = e^{-h}\frac{\partial}{\partial y}e^h$. Therefore the gradient of h is e^{-h} times the gradient of e^h. Near the optimal values both partial derivatives are "close" to zero but differ by a factor of e^{-h}. So, near the p.s.d. border where h goes to $-\infty$ and e^{-h} goes to ∞, the derivative of H can be large even very close to the optimal value.

Acknowledgements

Part of this work was supported by the Interuniversity Attraction Pole (IAP) research network in Statistics P5/24. The first author gratefully acknowledges the support of LJK-IMAG at Joseph Fourier University in Grenoble, France, where part of this work was completed during his visit in February–March 2007.

References

[1] G. N. Boshnakov and Bisher M. Iqelan, *Non-Gaussian maximum entropy processes*, Probability and Statistics Group, Research Report No. 3, School of Mathematics, University of Manchester, 2007.

[2] J. P. Burg, *Maximum entropy spectral analysis*, Proceedings of 37th meeting of society of exploration geophysicists, 1967, pp. 34–41.

[3] J.P. Burg, *Maximum entropy spectral analysis.*, Ph.D. Thesis, Stanford, CA, 1975.

[4] S. Dégerine and Lambert-Lacroix S, *Partial autocorrelation function of a nonstationary time series*, Journal of Multivariate Analysis (2003), 46–59.

[5] E.G. Gladishev, *Periodically correlated random sequences*, Soviet Math. **2** (1961), 385–388.

[6] Bisher M. Iqelan, *Periodically correlated time series: models and examples*, Ph.D. Thesis, 2007.

[7] Charles R. Johnson, *Matrix completion problems: A survey.*, Proc. Sympos. Appl. Math. **40** (1990), 171–198.

[8] S. Lambert-Lacroix, *On periodic autoregressive process estimation*, IEEE Transactions on Signal Processing **48** (2000), no. 6, 1800–1803.

[9] S. Lambert-Lacroix., *Extension of autocovariance coefficients sequence for periodically correlated processes*, Journal of Time Series Analysis **26** (2005), no. 6, 423–435.

[10] M. Pagano, *On periodic and multiple autoregression*, Ann. Statist. **6** (1978), 1310–1317.

[11] Dimitris N. Politis, *On the maximum entropy problem with autocorrelations specified on a lattice.*, IEEE Trans. Signal Process. **41** (1993), no. 4, 1715–1716.

[12] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[13] A.Wayne Roberts and Dale E. Varberg, *Convex functions.*, Pure and Applied Mathematics, 57. New York-London: Academic Press, a subsidiary of Harcourt Brace Jovanovich, Publishers. XX, 300 p. $ 19.50, 1973 (English).

[14] Novellone Rozario and Athanasios Papoulis, *Spectral estimation from nonconsecutive data.*, IEEE Trans. Inf. Theory **33** (1987), 889–894.