ON FLAT-TOP KERNEL SPECTRAL DENSITY ESTIMATORS FOR HOMOGENEOUS RANDOM FIELDS*

by

Dimitris N. Politis and Joseph P. Romano
Purdue University Stanford University

Technical Report #94-14

Department of Statistics
Purdue University

June 1994

△

On Flat-top Kernel Spectral Density Estimators for Homogeneous Random Fields *

Dimitris N. Politis
Department of Statistics
Purdue University
W. Lafayette, IN 47907

Joseph P. Romano
Department of Statistics
Stanford University
Stanford, CA 94305

Abstract

The problem of nonparametric estimation of the spectral density function of a partially observed homogeneous random field is addressed. In particular, a class of estimators with favorable asymptotic performance (bias, variance, rate of convergence) is proposed. The proposed estimators are actually shown to be $\sqrt{N}$-consistent if the autocovariance function of the random field is supported on a compact set, and close to $\sqrt{N}$-consistent if the autocovariance function decays to zero sufficiently fast for increasing lags.

Keywords. Bartlett's estimator, bias reduction, mean squared error, lag-windows, nonparametric spectral estimation, random fields.

*Research partially supported by a Purdue Research Foundation Faculty Grant.

1. Introduction

Let $\{X(t), t \in Z^n\}$ be a random field in n dimensions, with $n \in Z^+$, i.e., a collection of real valued random variables $X(t)$ defined on a probability space $(\Omega, \mathcal{A}, P)$, and indexed by the variable $t = (t_1, \ldots, t_n)$. The random field $\{X(t)\}$ will be assumed to be homogeneous (stationary, shift invariant), and to possess zero mean, i.e., $EX(t) = 0$, autocovariance function $R(s) = X(t)X(t+s)$, for $s = (s_1, s_2, \ldots, s_n) \in Z^n$, and the spectral density function defined as

$$f(w) = \frac{1}{(2\pi)^n} \sum_{t \in Z^n} R(t) e^{-i(w \cdot t)},$$

where $w = (w_1, \ldots, w_n) \in [-\pi, \pi]^n$, and $(w \cdot t) = \sum_{i=1}^n w_i t_i$ is the inner product in n-dimensional Euclidean space.

After observing a finite stretch of data, i.e., $\{X(t), t \in E_N\}$, where E_N is the rectangle consisting of the points $t = (t_1, t_2, \ldots, t_n) \in Z^n$ such that $1 \leq t_k \leq N_k$, and where N_k, for $k = 1, 2, \ldots, n$ are some positive integers, the problem at hand is estimating the spectral density $f(w)$ at some predetermined point w. To get an asymptotically consistent estimator of $f(w)$ one can employ the original time series idea of M. S. Bartlett (1946) that can be described as follows: choose integers $M_1, \ldots, M_n$ (that may depend on $N_1, \ldots, N_n$) and consider all rectangles consisting of points t such that $1 \leq u_k \leq t_k \leq U_k \leq N_k$, where the integers u_k, U_k satisfy $U_k - u_k + 1 = M_k$, for $k = 1, 2, \ldots, n$; calculate a periodogram from each of these smaller rectangles, and then average all such periodograms to obtain an estimator that, under some standard regularity conditions, is consistent for the spectral density function.

Analogously to the time series case (cf. Priestley (1981)), it can be shown that Bartlett's estimator can be written in an approximate (but asymptotically equivalent) form as the lag-window spectral estimator given below:

$$\tilde{f}_M^B(w) = \frac{1}{(2\pi)^n} \sum_{s \in Z^n} \lambda_M^B(s) \tilde{R}(s) e^{-i(w \cdot s)}, \tag{1}$$

where $\tilde{R}(s)$ are the (biased) sample autocovariances defined by $\tilde{R}(s) = N^{-1} \sum_{t \in Z^n} \tilde{X}(t)\tilde{X}(t+s)$; in the above, $N = \prod_{i=1}^n N_i$,

$$\tilde{X}(t) = \begin{cases} X(t) & \text{if } t \in E_N \\ 0 & \text{else} \end{cases}$$

and the Bartlett lag-window in n dimensions is defined as

$$\lambda_M^B(s) = \left((1 - \frac{|s_1|}{M_1})(1 - \frac{|s_2|}{M_2}) \cdots (1 - \frac{|s_n|}{M_n}) \right)^+, \tag{2}$$

where $(x)^+ = \max(x, 0)$ is the positive part function.

Nevertheless, in order to talk about consistency in this n-dimensional setting some clarifications are appropriate; to this effect we define Condition C_0 below.

Condition C_0: There are constants $c^*, c_* > 0$ such that $c_* < \frac{M_i}{M_k} < c^*$, $c_* < \frac{N_i}{N_k} < c^*$, for any $i, k = 1, 2, \ldots, n$, and $M = \prod_{i=1}^n M_i \to \infty$ as $N = \prod_{i=1}^n N_i \to \infty$, but with $\frac{M}{N} \to 0$.

Under Condition C_0 and some moment and weak dependence conditions it can be shown that

$$Bias(\tilde{f}_M^B(w)) \equiv E\tilde{f}_M^B(w) - f(w) = O(M^{-1/n}), \tag{3}$$

and

$$Var(\tilde{f}_M^B(w)) = O(M/N); \tag{4}$$

see, e.g., Theorem 2 and Corollary 2 in Politis and Romano (1993)). Note that although the M/N order of magnitude of $Var(\tilde{f}_M^B(w))$ is quite typical for all lag-window spectral estimators (see, e.g., Rosenblatt (1985), Zhurbenko (1986)), the order of magnitude of $Bias(\tilde{f}_M^B(w))$ is unnecessarily large, and this results into poor MSE (Mean Squared Error) performance for $\tilde{f}_M^B(w)$. For example, spectral estimators constructed using different lag-windows that satisfy some conditions and are supported on the set $\{s$ such that $\max_k |s_k|/M_k \leq 1\}$ have bias of order $O(M^{-2/n})$; cf. Yuan and Subba Rao (1993).

In the present paper, a new class of lag-window spectral estimators for random fields will be proposed possessing very low asymptotic bias in cases where the true spectral density function is smooth enough. The new estimators are related to the Bartlett estimator in that they are essentially obtained by a linear combination of two Bartlett estimators with different bandwidths. The estimators are defined in Section 2, and their performance is evaluated in Section 3; Section 4 consists of some practical comments, and the Appendix contains all technical proofs.

2. Spectral estimators based on flat-top lag-windows

Since the early papers by Parzen (1957a,b) it has been well understood that the asymptotic behavior of a spectral estimator hinges on the behavior of its lag-window around the origin $s = 0$. So we define another lag-window λ_M of 'pyramidal' shape by

$$\lambda_M(s) = \left(1 - \max_k \frac{|s_k|}{M_k}\right)^+ ; \tag{5}$$

note that equation (5) makes sense even if the M_k's are not necessarily integers. It is apparent that $\lambda_M(s) \simeq \lambda_M^B(s)$ if s is in the neighborhood of zero, or equivalently, for any s, provided $\min_k M_k$ is big enough. The lag-window λ_M maintains similar asymptotic properties[1] to $\lambda_M^B(s)$ albeit it is easier to work with; in Figures 1 and 2, the lag-windows $\lambda_M^B(s)$ and $\lambda_M(s)$ are plotted for comparison in the case $n = 2$, and with $M_1 = 30$, $M_2 = 20$.

Let c be a constant in $(0,1)$, and let $m_k = cM_k$, for $k = 1, 2, \ldots, n$; also let $\lambda_m(s) = \left(1 - \max_k \frac{|s_k|}{m_k}\right)^+$ be the corresponding pyramidal lag-window. We now introduce a 'flat-top' lag-window by defining

$$\lambda_{M,m}(s) = \frac{1}{1-c}\lambda_M(s) - \frac{c}{1-c}\lambda_m(s). \tag{6}$$

The lag-window $\lambda_{M,m}(s)$ is designed to be 'flat', i.e., constant, in a neighborhood of zero, like a pyramid with its top chopped off; see Figure 3 where $\lambda_{M,m}(s)$ is plotted in the case $n = 2$, with $M_1 = 30$, $M_2 = 20$, and $c = 1/2$.

A family of flat-top lag-window spectral estimators $\{\hat{f}_c(w); c \in (0,1)\}$ is now defined by

$$\hat{f}_c(w) = \frac{1}{(2\pi)^n} \sum_{s \in Z^n} \lambda_{M,m}(s)\hat{R}(s)e^{-i(w \cdot s)}, \tag{7}$$

[1] Compare equations (3) and (4) to (9) and (10) in what follows.

where $\hat{R}(s) = N\tilde{R}(s)/\left(\prod_{k=1}^{n}(N_k - |s_k| + 1)\right)$ are the unbiased[2] sample autocovariances. Observe that

$$\hat{f}_c(w) = \frac{1}{1-c}\tilde{f}_M(w) - \frac{c}{1-c}\tilde{f}_m(w), \tag{8}$$

where $\tilde{f}_M(w) = \frac{1}{(2\pi)^n}\sum_{s\in Z^n}\lambda_M(s)\hat{R}(s)e^{-i(w\cdot s)}$, and $\tilde{f}_m(w) = \frac{1}{(2\pi)^n}\sum_{s\in Z^n}\lambda_m(s)\hat{R}(s)e^{-i(w\cdot s)}$; in other words, $\hat{f}_c(w)$ is just a linear combination of two estimators of the same type, but having different bandwidths. As was previously alluded, the estimator $\tilde{f}_M(w)$ has similar asymptotic properties to the Bartlett estimator $\tilde{f}_M^B(w)$; in particular, under Condition C_0 and some moment and weak dependence conditions it can be proved that

$$Bias(\tilde{f}_M(w)) = O(M^{-1/n}), \tag{9}$$

and

$$Var(\tilde{f}_M(w)) = O(M/N). \tag{10}$$

It is now easy to see that $Var(\hat{f}_c(w)) = O(M/N)$ as well; this follows because $Var(\tilde{f}_m(w)) = O(c^n M/N) = O(M/N)$, and because $Cov(\tilde{f}_M(w), \tilde{f}_m(w) = O(M/N)$ by the Cauchy inequality. Since the M/N order of magnitude for the variance is not reduced, achieving a favorable MSE performance for $\hat{f}_c(w)$ hinges on the order of magnitude of its bias. In the next section it will be shown that $\hat{f}_c(w)$ has dramatically smaller bias than the Bartlett estimator; in some sense, $\hat{f}_c(w)$ is 'bias-corrected', i.e., $\tilde{f}_m(w)$ is used to eliminate a good part of the bias of $\tilde{f}_M(w)$ in equation (8), to render a less biased estimator.

[2]Note that $\tilde{R}(s) \simeq \hat{R}(s)$ if $\max_i |s_i|$ is small compared to the N_i's; however, by Condition C_0, the M_i's are of a smaller order of magnitude than the N_i's, and thus it does not make much difference whether we use $\tilde{R}(s)$ or $\hat{R}(s)$ in equation (7). There is another reason that sometimes points to using $\tilde{R}(s)$ in a sum of the type of equation (1) or (7), namely that so doing yields an almost surely nonnegative spectral estimator *provided the lag-window has an everywhere nonnegative Fourier transform*; see, for example, Politis and Romano (1992). Since this is not the case with the lag-window $\lambda_{M,m}$ we have opted to use the unbiased $\hat{R}(s)$ to avoid an unnecessary complication; see Section 4 for a further discussion on the nonnegativity issue.

3. Performance of the flat-top lag-window spectral estimators

In this section, the performance of the flat-top lag-window spectral estimators will be assessed under a range of weak dependence conditions that are defined in what follows.

Condition C_1: There are positive constants $B, c_1, c_2, \ldots, c_n$ and a constant $K > n$ such that $R(s) \le B(\max_i \frac{|s_i|}{c_i})^{-K}$.

Condition C_2: There are positive constants $K, B, c_1, c_2, \ldots, c_n$ such that $R(s) \le Be^{-K \max_i \frac{|s_i|}{c_i}}$.

Condition C_3: There are positive constants $c_1, c_2, \ldots, c_n$ such that $R(s) = 0$, if $\max_i \frac{|s_i|}{c_i} \ge 1$.

Conditions C_1 to C_3 are some of the most commonly used weak dependence conditions based on second moments alone; they are given in increasing order of strength, i.e., if Condition C_2 holds, then Condition C_1 holds as well, and if Condition C_3 holds, then Conditions C_1 and C_2 hold as well. Even under the weakest of the three Conditions, namely C_1, it is easily seen that $\sum_{s \in Z^n} |R(s)| < \infty$, which implies that the spectral density function exists and is continuous. In fact, Conditions C_1 to C_3 can be interpreted as different conditions on the smoothness of the spectral density $f(w)$; cf. Minakshisundaram and Szasz (1947), Wainger (1965), Katznelson (1968), Butzer and Nessel (1971), Stein and Weiss (1971), and the references therein.

Note that the introduction of the constants $c_1, c_2, \ldots, c_n$ allows for the possibility that the autocovariances tend to zero at different rates along the different directions in Z^n; it is quite natural that the construction of the spectral estimator i.e., the choice of the M_i's, will reflect this fact. In general, it would seem preferable that a large M_j should be used for direction j if the corresponding c_j is large as compared to $c_1, c_2, \ldots, c_n$, i.e., in a direction along which the autocovariance decays more slowly; hence, a reasonable choice for the M_i's would satisfy Condition C_M as follows:

Condition C_M: $\frac{M_i}{M_j} = \frac{c_i}{c_j}$, for $i,j = 1,2,\ldots,n$, where the c_j's are the same positive constants appearing in Conditions C_1, C_2, or C_3.

Obviously, just assuming $\frac{M_i}{M_j} \to \frac{c_i}{c_j}$ would be equivalent to Condition C_M in terms of asymptotic considerations alone, and would allow the use of integer values for the M_j's. Nonetheless, in case that either $\frac{M_i}{M_j} = \frac{c_i}{c_j}$ or $\frac{M_i}{M_j} \to \frac{c_i}{c_j}$, the task of properly choosing the M_i's for $i = 1,2,\ldots,n$, reduces to choosing just one number, namely their product M; more on the subject of choosing M will be found in Section 4.

Under different combinations of conditions, the performance of the family of flat-top lag-window estimators $\{\hat{f}_c(w); c \in (0,1)\}$ is quantified in the sequence of theorems that follows.

Theorem 1 *Under Conditions C_0, C_1, and C_M, it follows that*

$$\sup_{w\in[-\pi,\pi]^n} |Bias(\hat{f}_c(w))| = O(M^{1-\frac{K}{n}}).$$

Let w be some point in $[-\pi,\pi]^n$; under further assumptions sufficient to ensure[3] the validity of equation (10), and letting $M \sim A_w N^{n/(2K-n)}$ for some constant[4] $A_w > 0$, the asymptotic order of the Mean Squared Error of $\hat{f}_c(w)$ is given by $MSE(\hat{f}_c(w)) = O(N^{2(n-K)/(2K-n)})$.

Theorem 2 *Under Conditions C_0, C_2, and C_M, and by letting $M_i \sim dc_i \log N_i$, for $i = 1,2,\ldots,n$, where d is a constant such that $d > n/(2K)$, it follows that*

$$\sup_{w\in[-\pi,\pi]^n} |Bias(\hat{f}_c(w))| = O(\frac{(\log N)^{n-1}}{\sqrt[n]{N^{dK}}}) = o(\frac{1}{\sqrt{N}}).$$

[3]There is a variety of different such sufficient conditions; see, for example, Rosenblatt (1985), Zhurbenko (1986), Politis and Romano (1993), or Yuan and Subba Rao (1993).

[4]A_w may depend on w because in general the variance of a lag-window estimator of $f(w)$ is asymptotically proportional to $(M/N)f^2(w)$.

Let w be some point in $[-\pi,\pi]^n$; under further assumptions sufficient to ensure the validity of equation (10) it follows that $MSE(\hat{f}_c(w)) = O(\frac{(\log N)^n}{N})$

Theorem 3 *Assume Condition C_3 and that $N \to \infty$ in such a way that there are positive constants c_*, c^* such that $c_* < \frac{N_i}{N_k} < c^*$, for any $i,k = 1,2,\ldots,n$; also assume that the m_i's and the M_i's are* constants *satisfying $M_i \geq m_i \geq c_i$ for $i = 1,2,\ldots,n$. Then it follows that*

$$\sup_{w\in[-\pi,\pi]^n} |Bias(\hat{f}_c(w))| = 0.$$

Let w be some point in $[-\pi,\pi]^n$; under further assumptions sufficient to ensure the validity of equation (10) it follows that $MSE(\hat{f}_c(w)) = O(1/N)$.

It is should be noted at this point that $\hat{f}_c(w)$ is not only consistent for $f(w)$, but its rate of convergence –as measured by the MSE– is very fast. Under Condition C_1, i.e., if the autocovariances decay like a power, the rate of convergence[5] of $\hat{f}_c(w)$ depends on how fast the autocovariances decay, or equivalently, on how smooth the true spectral density is, e.g., how many derivatives it possesses. In other words, $\hat{f}_c(w)$ 'adapts' to the smoothness of the spectral density under consideration, and its performance is seen to improve if the spectral density is more favorable, i.e., smoother. For comparison, the standard spectral density estimators studied in Yuan and Subba Rao (1993) do *not* share this property since, as mentioned in the Introduction, their bias remains of order $O(M^{-2/n})$ even though the true spectral density might possess a great number of derivatives.

In essense, the $O(M^{-2/n})$ order of the bias of standard estimators is intimately connected with insisting that the lag-window estimator is surely nonnegative; cf. Priestley

[5]In the case of a time series, i.e., $n = 1$, the rate of convergence of $\hat{f}_c(w)$ is optimal as can be seen by Samarov's (1977) lower bound for the precision of a spectral density estimate. However, an analogous lower bound seems to be unavailable for the case of a random field, i.e., $n > 1$; see also Rosenblatt (1985, p. 146).

(1981). The estimator $\hat{f}_c(w)$ is *not* necessarily nonnegative but in giving up nonnegativity it gains accuracy. In particular, if $K = n + 2$, the $Bias(\hat{f}_c(w)) = O(M^{-2/n})$ which is the order of bias of the standard nonnegative estimators, whereas if $K > n + 2$, $Bias(\hat{f}_c(w)) = O(M^{(n-K)/n}) = o(M^{-2/n})$. Note that if K is very large, i.e., the spectral density is very smooth, the rate of convergence of $\hat{f}_c(w)$ corresponding to the optimal choice of $M \sim A_{f,w}N^{n/(2K-n)}$ can be very close to $\sqrt{N}$. In the case of exponential decay of the autocovariances i.e., under Condition C_2, $\hat{f}_c(w)$ is $\sqrt{N/(\log N)^n}$-consistent, whereas under Condition C_3, i.e., if $\{X(t)\}$ is essentially a Moving Average of an uncorrelated random field, $\hat{f}_c(w)$ is actually exactly $\sqrt{N}$-consistent!

4. Practical comments and conclusions

4.1. The question of nonnegativity. An issue that was briefly mentioned in Section 3 concerns the nonnegativity of spectral estimators; since $f(w) \geq 0$, it is natural to desire that its estimator be nonnegative as well. However, $\hat{f}_c(w)$ is not guaranteed to be nonnegative as can be intuitively seen by looking at the expression for $\hat{f}_c(w)$ as a difference in equation (8).

Nevertheless, the solution to this problem is quite obvious: if in a practical situation $\hat{f}_c(w)$ turns out to be negative, the practitioner would naturally prefer to use 0 as the estimator of $f(w)$, rather than $\hat{f}_c(w)$ itself. In other words, the practitioner would use the estimator $\hat{f}_c^+(w) = \max(\hat{f}_c(w), 0)$. It is easy to show that

$$MSE(\hat{f}_c^+(w)) \leq MSE(\hat{f}_c(w)),$$

from which it follows that $\hat{f}_c^+(w)$ inherits the favorable asymptotic properties of $\hat{f}_c(w)$; as a matter of fact, $\hat{f}_c^+(w)$ will be equal to $\hat{f}_c(w)$ with probability that tends to one, and in such a way[6] that $E(\hat{f}_c^+(w) - \hat{f}_c(w))^2 \leq 2MSE(\hat{f}_c(w))$; cf. Politis and Romano (1992). Since the $MSE(\hat{f}_c(w))$ is shown to be of very small order in Theorems 1, 2, and 3, it is apparent that $\hat{f}_c^+(w)$ is a very accurate estimator of $f(w)$ that has the additional desirable property of being nonnegative.

4.2. Selecting the design parameter c. In practice, only one estimator from the family $\{\hat{f}_c(w); c \in (0,1)\}$ or $\{\hat{f}_c^+(w); c \in (0,1)\}$ will be used to carry out the required estimation of $f(w)$, i.e., one must choose the parameter c. Note that to address the problem of choosing c properly may require more careful asymptotic considerations than looking at the rate of convergence and the order of magnitude of the MSE of $\hat{f}_c(w)$ since the latter do not depend at all on c as long as $c \in (0,1)$. As it can be shown (cf. Politis

[6] Under some additional assumptions it can actually be shown that $E(\hat{f}_c^+(w) - \hat{f}_c(w)^2) = o(MSE(\hat{f}_c(w)))$.

and Romano (1992)), the choice of c actually influences the proportionality constant implicit in the relationship $Var(\hat{f}_c(w)) = O(M/N)$, and is also expected to influence the proportionality constant in the relationship $Bias(\hat{f}_c(w)) = O(M^{1-\frac{K}{n}})$ in Theorem 1; therefore, the choice of c may influence the *exact* MSE of $\hat{f}_c(w)$, although the order of magnitude of the MSE remains unchanged.

Nevertheless, it is quite intuitive that values of c near the extremes 0 and 1 should be avoided for the following reasons: in the extreme case where $c = 0$, $\hat{f}_c(w)$ actually reduces to $\tilde{f}_M(w)$ which has unfavorable bias properties (see equation (9)), while in the extreme case where $c = 1$ the lag-window $\lambda_{M,m}(s)$ becomes rectangular (see equation (13)), and this again results into poor bias performance since the corresponding spectral window (the Fourier transform of $\lambda_{M,m}(s)$ with $m = M$) has many prominent sidelobes that promote 'leakage'; more details in the special case $n = 1$ can be found in Politis and Romano (1992) where the simple choice $c = 1/2$ is recommended, and shown to perform well in simulated experiments.

4.3. Choosing the bandwidth parameter M using the data. While a permanent choice of the design parameter c can be made once and for all and used in connection with different data sets, the bandwidth parameter M must be chosen in a data-driven way for optimal performance in applications; see Woodroofe (1970) for an example in the related context of probability density estimation. We will now focus on choosing M, assuming that the parameter c has been selected already; note that in view of Condition C_M, only the product $M = \prod_{i=1}^{n} M_i$ must be specified, since the individual M_i's should be proportional to the constant c_i's appearing in Conditions C_1, C_2, and C_3.

The key to this difficult practical problem lies in noting that although Condition C_3 is the strongest of the three Conditions C_1, C_2, and C_3, even under Conditions C_1 or C_2, Condition C_3 is seen to still hold *approximately*; in other words, if one of Conditions C_1 or C_2 holds, it follows that there are positive constants $c_1, c_2, \ldots, c_n$ such that $R(s) \simeq 0$, if $\max_i \frac{|s_i|}{c_i} \geq 1$.

Hence, we may focus on Condition C_3 and its corresponding Theorem 3; note, however, that Theorem 3 gives an *exact*[7] prescription for an optimal choice of the the m_i's in terms of the c_i's, namely $m_i = c_i$, for $i = 1, \ldots, n$. Therefore, since $M_i = m_i/c$, the M_i's are *also* exactly prescribed. Note that in any practical situation the constant c_i's are unknown as well, but they can be estimated from the data; for example, if is observed that $\tilde{R}(s) \simeq 0$, if $\max_i \frac{|s_i|}{\hat{c}_i} \geq 1$, then the $\hat{c}_i$'s are estimates of the c_i's in Condition C_3.

To summarize, the following heuristic procedure is suggested for choosing the M_i bandwidths: *after having picked the parameter c, let $M_i = \hat{c}_i/c$, for $i = 1, \ldots, n$, where the $\hat{c}_i$'s are some estimates of the c_i's appearing in Condition C_3.* Although this simple empirical guideline has been shown to perform well in simulated experiments in the time series case (Politis and Romano (1992)), it goes without saying that it does not address the problem fully; the difficult problem of optimally choosing the bandwidth of $\hat{f}_c(w)$ is still open for more work, including a theoretical analysis of the performance of $\hat{f}_c(w)$ and $\hat{f}_c^+(w)$ when the bandwidth is chosen in such an adaptive, data-driven fashion.

[7] Actually, Theorem 3 states that by choosing $m_i \geq c_i$, $i = 1, \ldots, n$, zero bias is achieved; but since the variance of $\hat{f}_c(w)$ is proportional to M/N, it is obvious that to reduce the MSE we should choose the m_i's as small as possible, i.e., $m_i = c_i$.

Appendix: Technical proofs.

PROOF OF THEOREM 1. Let w be any point in $[-\pi,\pi]^n$ and note that

$$Bias(\hat{f}_c(w)) = E\hat{f}_c(w) - f(w)$$

$$= \frac{1}{(2\pi)^n}\sum_{s\in Z^n}\lambda_{M,m}(s)E\hat{R}(s)e^{-i(w\cdot s)} - \frac{1}{(2\pi)^n}\sum_{s\in Z^n}R(s)e^{-i(w\cdot s)}$$

$$= \frac{1}{(2\pi)^n}\sum_{s\in Z^n}\left(\lambda_{M,m}(s)-1\right)R(s)e^{-i(w\cdot s)}, \tag{11}$$

where it was used that $E\hat{R}(s) = R(s)$. Now consider the following partition of Z^n, namely $Z^n = \cup_{i=1}^n \left(A_i \cup \bar{A}_i\right)$, where $A_i = \{s$ such that $\frac{s_i}{M_i} = \max_k \frac{|s_k|}{M_k}\}$, and $\bar{A}_i = \{s$ such that $\frac{-s_i}{M_i} = \max_k \frac{|s_k|}{M_k}\}$. Note that the A_i's and $\bar{A}_i$'s are essentially disjoint except for potentially some points on their boundaries, e.g., in the case where $\frac{s_1}{M_1} = \frac{s_2}{M_2} = \max_k \frac{|s_k|}{M_k}$, etc.; we could construct disjoint versions of the A_i's and $\bar{A}_i$'s by letting $A_1^* = A_1$, $A_2^* = A_2 - (A_2 \cap A_1^*)$, $A_3^* = A_3 - (A_3 \cap (A_2^* \cup A_1^*))$, and so on, but this is quite unnecessary because the contribution of the boundaries on a sum of the type of (11) is negligible. Therefore, we can write

$$Bias(\hat{f}_c(w)) \sim \sum_{A_1} + \sum_{A_2} + \cdots + \sum_{A_n} + \sum_{\bar{A}_1} + \sum_{\bar{A}_2} + \cdots + \sum_{\bar{A}_n}, \tag{12}$$

where for $j = 1, 2, \ldots, n$

$$\sum_{A_j} = \frac{1}{(2\pi)^n}\sum_{s\in A_j}\left(\lambda_{M,m}(s)-1\right)R(s)e^{-i(w\cdot s)},$$

and

$$\sum_{\bar{A}_j} = \frac{1}{(2\pi)^n}\sum_{s\in \bar{A}_j}\left(\lambda_{M,m}(s)-1\right)R(s)e^{-i(w\cdot s)}.$$

Note that $\lambda_{M,m}(s) = \lambda_{M,m}(-s)$, as well as $R(s) = R(-s)$, for all s, and therefore $\sum_{A_j} = \sum_{\bar{A}_j}$, for $j = 1, 2, \ldots, n$. We now proceed to analyze in detail the term $\sum_{A_1}$, the analysis of the terms $\sum_{A_2}, \ldots, \sum_{A_n}$ being similar.

Observe that $\sum_{A_1} = \sum_{A_1^1} + \sum_{A_1^2} + \sum_{A_1^3}$, where

$$\sum_{A_1^j} = \frac{1}{(2\pi)^n} \sum_{s \in A_1 \cap B_j} (\lambda_{M,m}(s) - 1)\, R(s) e^{-i(w \cdot s)},$$

for $j = 1, 2, 3$, and where $B_1 = \{s \text{ such that } \max_k |s_k|/m_k \leq 1\}$, $B_2 = \{s$ such that $\max_k |s_k|/m_k > 1$ and $\max_k |s_k|/M_k \leq 1\}$, and $B_3 = \{s \text{ such that } \max_k |s_k|/M_k > 1\}$.

We can now have a more transparent formula for the lag-window $\lambda_{M,m}(s)$, namely

$$\lambda_{M,m}(s) = \begin{cases} 1 & \text{if } s \in B_1 \\ \frac{M_j - |s_j|}{M_j - m_j} & \text{if } s \in A_j \cap B_2 \text{ or if } s \in \bar{A}_j \cap B_2 \\ 0 & \text{if } s \in B_3. \end{cases} \tag{13}$$

Since $\lambda_{M,m}(s) = 1$, for all $s \in B_1$, it follows that $\sum_{A_1^1} = 0$. Similarly, $\lambda_{M,m}(s) = 0$, for all $s \in B_3$, and therefore

$$|\sum_{A_1^3}| \leq \frac{1}{(2\pi)^n} \sum_{s \in A_1 \cap B_3} |R(s)| \leq \frac{B}{(2\pi)^n} \frac{\prod_{i=2}^n M_i}{M_1^{n-1}} \sum_{s_1 = M_1}^{\infty} s_1^{n-1} (\frac{s_1}{c_1})^{-K} = O(M_1^{n-K}) = O(M^{1-\frac{K}{n}}),$$

where Conditions C_1 and C_M were invoked to bound the above sum. In particular, the following argument was used and will be used again in the sequel: since $s \in A_1$, it follows that $s_1/M_1 = \max_k |s_k|/M_k$; but by Condition C_M, it is also true that $s_1/c_1 = \max_k |s_k|/c_k$, and hence the bound on $R(s)$ is in accordance with Condition C_1.

Similarly, note that

$$|\sum_{A_1^2}| = |\frac{1}{(2\pi)^n} \sum_{s \in A_1 \cap B_2} (\lambda_{M,m}(s) - 1) R(s) e^{-i(w \cdot s)}|$$

$$\leq \frac{1}{(2\pi)^n} \sum_{s \in A_1 \cap B_2} \frac{s_1 - m_1}{M_1 - m_1} |R(s)| \leq \frac{B}{(2\pi)^n} \sum_{s \in A_1 \cap B_2} \frac{s_1 - m_1}{M_1 - m_1} (\frac{s_1}{c_1})^{-K}$$

$$\leq \frac{B c_1^K}{(2\pi)^n} \left(\prod_{i=2}^n M_i \right) \sum_{s_1 = m_1}^{M_1} s_1^{-K} = O(M_1^{n-K}) = O(M^{1-\frac{K}{n}}).$$

Hence, $\sum_{A_1} = O(M^{1-\frac{K}{n}})$, uniformly in w; a similar analysis gives $\sum_{A_j} = O(M^{1-\frac{K}{n}})$, uniformly in w, for all $j = 1, 2, \ldots, n$. Therefore, $Bias(\hat{f}_c(w)) = O(M^{1-\frac{K}{n}})$, uniformly in w, as we were supposed to prove. Finally, note that by the discussion after equation (10), it follows that $Var(\hat{f}_c(w)) = O(M/N)$; hence, the asymptotic order of the $MSE(\hat{f}_c(w))$ is $O(N^{2(n-K)/(2K-n)})$ as stated in the theorem. **Q.E.D.**

PROOF OF THEOREM 2. The proof of Theorem 2 is based on the decomposition (12) analogously to the proof of Theorem 1. In a similar fashion, we write $\sum_{A_1} = \sum_{A_1^1} + \sum_{A_1^2} + \sum_{A_1^3}$, where $\sum_{A_1^1} = 0$ as before.

However, now we have

$$|\sum_{A_1^2}| = |\frac{1}{(2\pi)^n} \sum_{s\in A_1\cap B_2} (\lambda_{M,m}(s) - 1)R(s)e^{-i(w\cdot s)}|$$

$$\leq \frac{1}{(2\pi)^n} \sum_{s\in A_1\cap B_2} \frac{s_1 - m_1}{M_1 - m_1}|R(s)| \leq \frac{B}{(2\pi)^n} \sum_{s\in A_1\cap B_2} \frac{s_1 - m_1}{M_1 - m_1} e^{-K\frac{s_1}{c_1}} = O(M_1^{n-1}e^{-KM_1/c_1}),$$

and similarly

$$|\sum_{A_1^3}| = O(\sum_{s_1=M_1}^{\infty} s_1^{n-1}e^{-K\frac{s_1}{c_1}}) = O(M_1^{n-1}e^{-KM_1/c_1}).$$

Note that the bounds on both $\sum_{A_1^2}$ and $\sum_{A_1^3}$ are uniform in w. A similar analysis gives $\sum_{A_j} = O(M^{1-1/n}e^{-\frac{K}{c_j}M_j})$, uniformly in w, for all $j = 1, 2, \ldots, n$. Now letting $M_i \sim dc_i \log N_i$, for $i = 1, 2, \ldots, n$, it follows that

$$\sup_{w\in[-\pi,\pi]^n} |Bias(\hat{f}_c(w))| = O(\frac{(\log N)^{n-1}}{\sqrt[n]{N^{dK}}}) = o(\frac{1}{\sqrt{N}}).$$

Since the $Bias(\hat{f}_c(w))$ is of smaller order than the $Var(\hat{f}_c(w)) = O(M/N)$, it now follows that $MSE(\hat{f}_c(w)) = O(M/N) = O(\frac{(\log N)^n}{N})$ as claimed. **Q.E.D.**

PROOF OF THEOREM 3. The proof of Theorem 3 is again based on the decomposition (12) presented in the proof of Theorem 1. As before, $\sum_{A_j^1} = 0$, for any $j = 1, \ldots, n$; however, we now also have $\sum_{A_j^2} = \sum_{A_j^3} = 0$, because $R(s) = 0$ for s outside of B_1. Hence $Bias(\hat{f}_c(w)) \equiv 0$, and $MSE(\hat{f}_c(w)) = O(1/N)$ as claimed. **Q.E.D.**

References

[1] Bartlett,M.S. (1946), On the theoretical specification of sampling properties of autocorrelated time series, *J.Roy.Statist.Soc.Suppl.*, 8, 27-41.

[2] Butzer, P. and Nessel, R. (1971), *Fourier analysis and approximation*, Academic Press, New York.

[3] Katznelson, Y. (1968), *An Introduction to Harmonic Analysis*, Dover, New York.

[4] Minakshisundaram, S. and Szasz, O. (1947), On absolute convergence of multiple Fourier series, *Trans. Amer. Math. Soc.*, 61, pp. 36-53.

[5] Parzen, E. (1957a), On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.*, 28, 329-348.

[6] Parzen, E. (1957b), Choosing an estimate of the spectral density function of a stationary time series. *Ann. Math. Statist.*, 28, 921-932.

[7] Politis,D.N., and Romano,J.P. (1992), Bias-corrected nonparametric spectral estimation, Technical Report No. 92-50, Department of Statistics, Purdue University.

[8] Politis,D.N., and Romano,J.P. (1993), Nonparametric Resampling for Homogeneous Strong Mixing Random Fields, *J. Multivar. Anal.*, vol. 47, No. 2, pp. 301-328.

[9] Priestley, M.B. (1981), *Spectral Analysis and Time Series*, Academic Press.

[10] Rosenblatt, M. (1985), *Stationary sequences and random fields*, Birkhäuser, Boston.

[11] Samarov, A. (1977), Lower bound for risk of spectral density estimates, *Problems of Info. Transm.*, vol. 13, pp. 48-51.

[12] Stein, E.M., and Weiss, W. (1971), *Introduction to Fourier analysis on Euclidean spaces*, Princeton Univ. Press, Princeton, New Jersey.

[13] Wainger,S. (1965), Special trigonometric series in k dimensions, *Mem. Amer. Math. Soc.* vol. 59, Providence, Rhode Island.

[14] Woodroofe, M. (1970), On choosing a delta-sequence, *Ann. Math. Statist.*, vol. 41, no. 5, 1665-1671.

[15] Yuan, J. and Subba Rao, T. (1993), Spectral estimation for random fields with applications to Markov modelling and texture classification, in *Markov random fields: Theory and Applications*, R. Chellappa and A. Jain (eds.), Academic Press, Boston, pp. 179-210.

[16] Zhurbenko, I.G. (1986), *The Spectral Analysis of Time Series*, North-Holland, Amsterdam.

CAPTIONS FOR FIGURES

FIGURE 1. The Bartlett lag-window $\lambda_M^B(s)$ in the case $n = 2$, and with $M_1 = 30$, $M_2 = 20$.

FIGURE 2. The pyramidal lag-window $\lambda_M(s)$ in the case $n = 2$, and with $M_1 = 30$, $M_2 = 20$.

FIGURE 3. The flat-top lag-window $\lambda_{M,m}(s)$ in the case $n = 2$, with $M_1 = 30$, $M_2 = 20$, and $c = 1/2$.

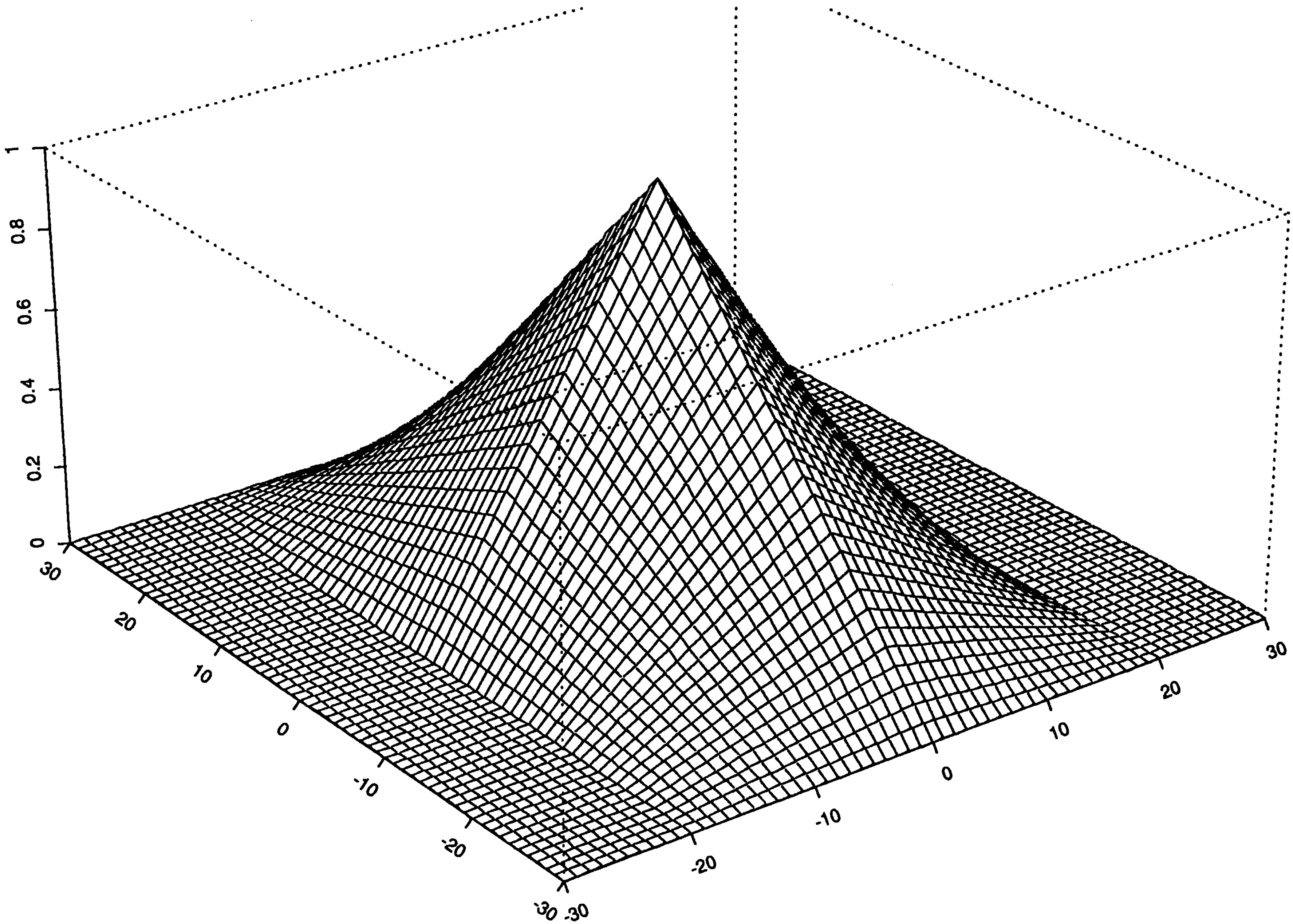

FIGURE 1

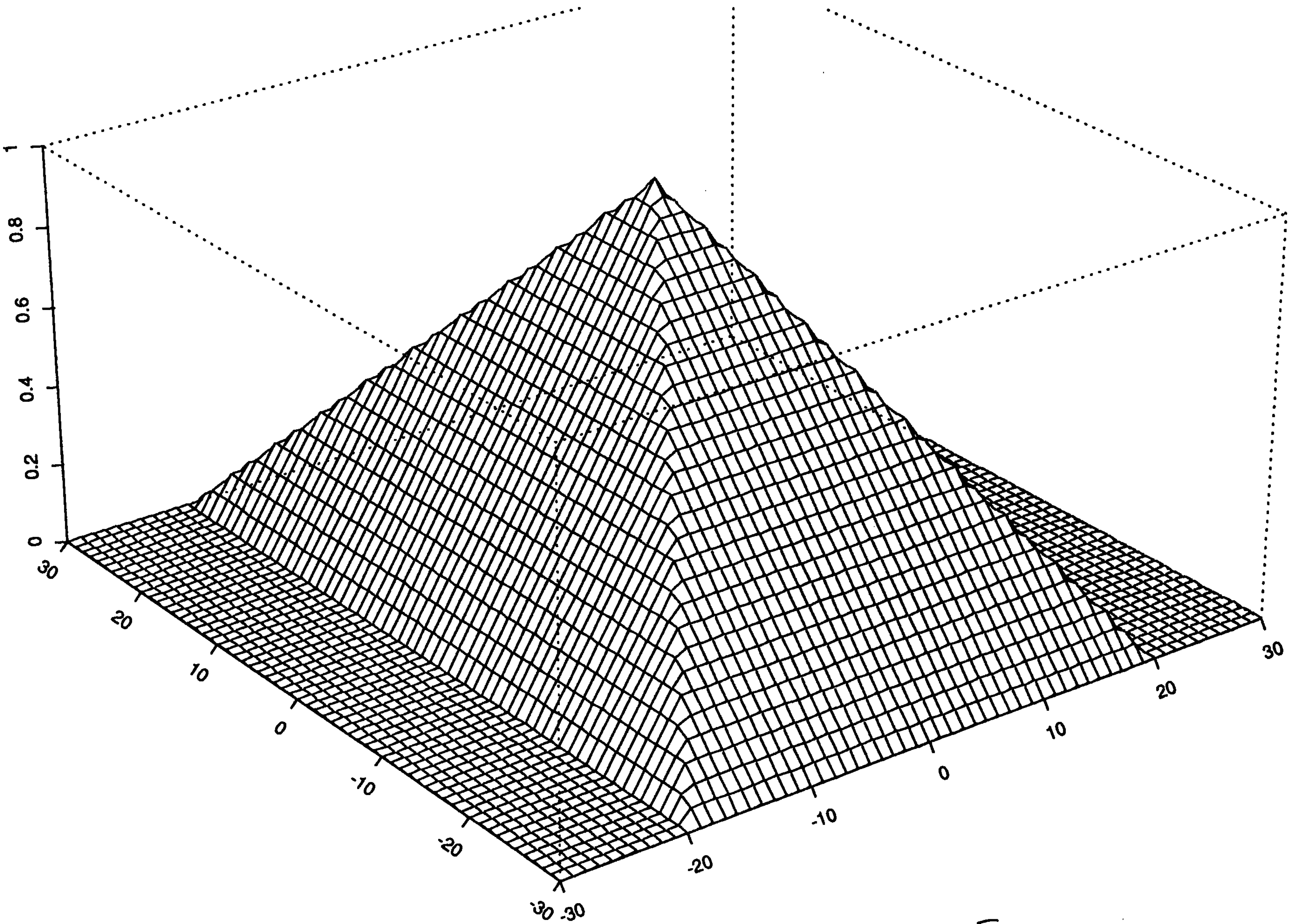

FIGURE 2

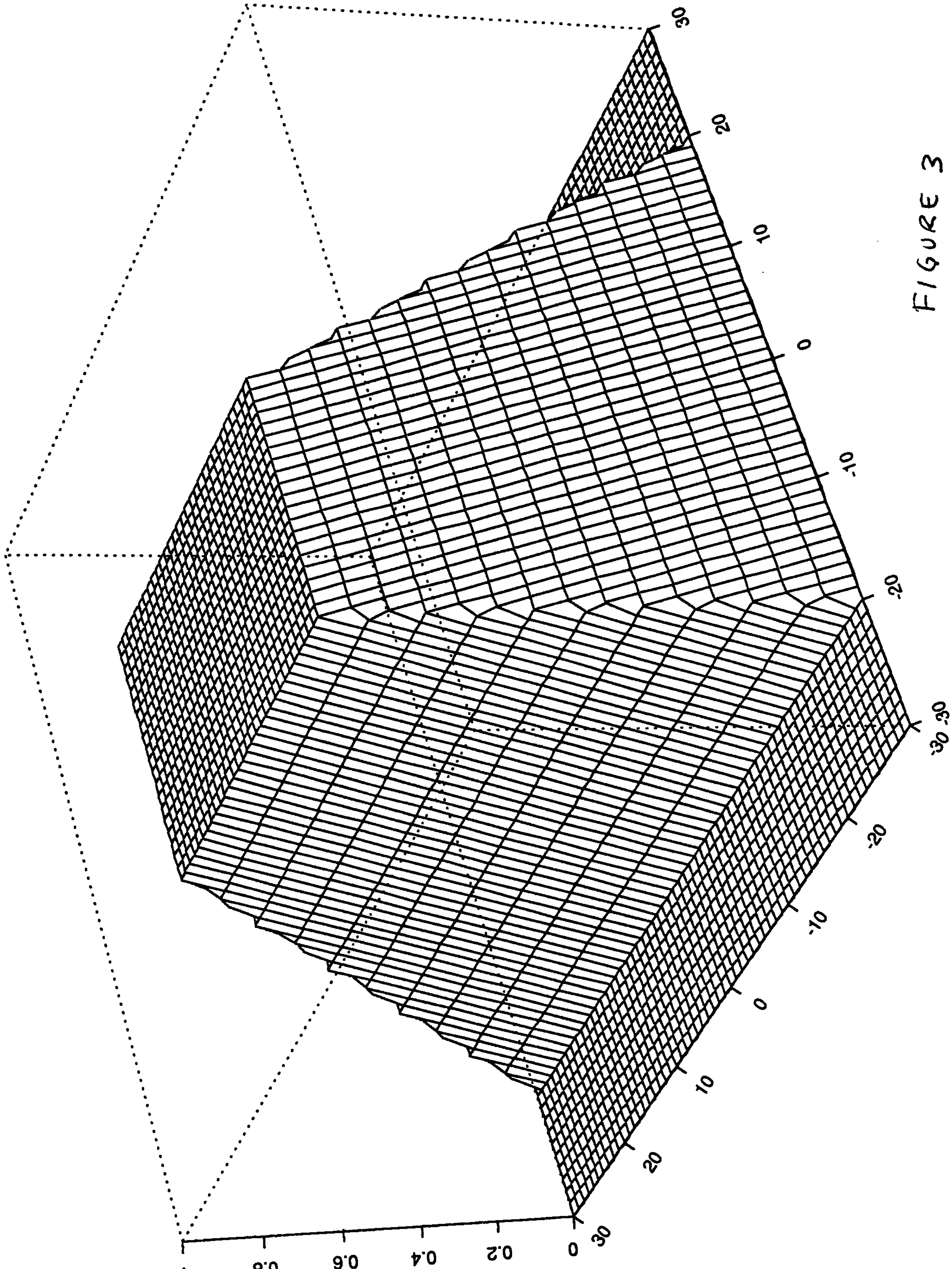

FIGURE 3