Low-Pass Filter Design using Locally Weighted Polynomial Regression and Discrete Prolate Spheroidal Sequences

Proietti, Tommaso and Luati, Alessandra

1 June 2009

Low-Pass Filter Design using Locally Weighted Polynomial Regression and Discrete Prolate Spheroidal Sequences

Tommaso Proietti
S.E.F. e ME. Q.
University of Rome "Tor Vergata"

Alessandra Luati
Department of Statistics
University of Bologna

1st June 2009

Abstract

The paper concerns the design of nonparametric low-pass filters that have the property of reproducing a polynomial of a given degree. Two approaches are considered. The first is locally weighted polynomial regression (LWPR), which leads to linear filters depending on three parameters: the bandwidth, the order of the fitting polynomial, and the kernel. We find a remarkable linear (hyperbolic) relationship between the cutoff period (frequency) and the bandwidth, conditional on the choices of the order and the kernel, upon which we build the design of a low-pass filter.

The second hinges on a generalization of the maximum concentration approach, leading to filters related to discrete prolate spheroidal sequences (DPSS). In particular, we propose a new class of low-pass filters that maximize the concentration over a specified frequency range, subject to polynomial reproducing constraints. The design of generalized DPSS filters depends on three parameters: the bandwidth, the polynomial order, and the concentration frequency. We discuss the properties of the corresponding filters in relation to the LWPR filters, and illustrate their use for the design of low-pass filters by investigating how the three parameters are related to the cutoff frequency.

Keywords Low-Pass filters; Kernels; Concentration; Filter Design.

1 Introduction

Trend filters that arise from fitting a locally weighted polynomial to a time series have a well established tradition in time series analysis and signal extraction; see Anderson (1971, ch. 3), Kendall (1973), Kendall, Stuart and Ord (1983), and the excellent historical review in Cleveland and Loader (1996). A thorough treatment is provided in books like Fan and Gijbels (1996), Loader (1999), Härdle (1989), Hastie and Tibshirani (1990) Ruppert, Wand and Carroll (2003), Fan and Yao (2005).

Locally weighted polynomial regression (LWPR) generates finite impulse response filters (LWPR filters, henceforth), that are widely applied in practice. See Ladiray and Quenneville (2001) for the application of the Henderson (1916) and Macaulay (1931) filters within the X-12 seasonal adjustment procedure. The seasonal-trend decomposition procedure STL, by Cleveland *et al.* (1990) is based on a particular LWPR filter called Loess.

The properties of LWPR filters are uniquely determined by three ingredients: the order of the approximating polynomial, the size of the neighborhood, also known as the bandwidth, and the weighting, or kernel, function. While on the one hand the choice of the kernel contributes only marginally to the estimation accuracy (see e.g. Härdle, 1989, sec. 4.5), and, for economic applications, the order of the polynomial never goes beyond the cubic, on the other hand the bandwidth is the most crucial component of the accuracy of the method, and thus its selection is a very sensitive issue. Usually, it is estimated from the observed data according to some criterion, e.g. by crossvalidation.

LWPR filters are trend extraction filter, and as such they somehow pass low frequency components and reduce the amplitude of high frequency ones. The first aim of the paper is to enforce the interpretation of LWPR filters as low-pass filters and to illustrate how the cutoff period (or frequency) is related to the above three key ingredients: bandwidth, polynomial order and kernel.

We define the cutoff period P_c as the smallest periodicity of the fluctuations that are passed i.e. that are preserved to a great extent, by the filter, whereas the fluctuations with smaller period are compressed. Conversely, the cutoff frequency in radians $\omega_c \in (0, \pi)$ is the maximum frequency in radians that is preserved by the filter, whereas higher frequencies are suppressed. The two quantities are inversely related, as $P_c = 2\pi/\omega_c$. An ideal low-pass filter has unit transfer function (TF) in the frequency range $(0, \omega_c)$, and zero TF elsewhere (see Percival and Walden, 1993).

For any LWPR filter, we derive the underlying cutoff frequency by least squares filter design principles, as the frequency at which a given LWPR filter provides the best approximation to an ideal low pass filter. Interestingly, we find out that given the kernel and polynomial order, the cutoff period is linearly related to the bandwidth; the intercept and slope of the linear relationship depend on the the order of the polynomial and the kernel.

This tight relationship is a mixed blessing: one the one hand, for any LWPR we can determine a reference period or frequency, which would enable us, say, to pick up the correct combination of bandwidth, order and kernel that produces a trend filter passing all the fluctuation with period greater than 8, or 10, years (the deviations of the observations from this component would be an high-pass component that is often identified as the business cycle). On the other hand, the relationship is also a most important limiting factor in the design of a LWPR low-pass filter. For instance, it will turn out that a local linear polynomial fit can produce a low-pass filter with a small cutoff period, such as one year and a half (e.g. 6 quarters), only when it uses a very small bandwidth, which is nevertheless detrimental to the reliability of the estimates (the accuracy of the trend estimator will be poor). If the bandwidth is increased then the cutoff period necessarily increases.

The second contribution of the paper is to propose a generalization of a well-known class of filters that are designed by using the discrete prolate spheroidal sequences (DPSS). The DPSS approach to filter design is based on the maximization of the concentration of a filter at a reference frequency, denoted ϖ. See by Tufts and Francis (1970), Papoulis and Bertran (1970), Eberhard (1973) and Slepian (1978). The concentration, introduced by Slepian and Pollak (1961) and Landau and Pollak (1961, 1962) for continuous signals, measures the proportion of the transfer function in the frequency range $[0, \varpi]$. The reader is referred to Slepian (1983), Mathews, Breakall and Karawas (1985) and Percival and Walden (1993, sec. 5.9) for further details on the development of the method. For the statistical properties of DPSS and their use in statistics, see Lii and Rosenblatt (2008).

A DPSS filter reproduces a constant or linear function of time. The filter that solves the traditional concentration problem is given by a zeroth order discrete prolate spheroidal sequence (Slepian, 1978; see also Thomson, 1982, and Xu and Chamaz, 1984). We generalize the DPSS approach by imposing the constraint that the filter reproduces a polynomial trend of any degree; we thus obtain a class of DPSS filters that depend on three parameters, the bandwidth, the order of the polynomial and the concentration frequency, ϖ.

We then discuss the interpretation of the latter. Our interpretation differs from the mainstream one: in fact, we show that the choice of the concentration frequency does not really differ from the choice of the kernel. Finally, we discuss the properties of the corresponding filters in relation to the LWPR filters, and illustrate their use for the design of low-pass filters by investigating how the three parameters are related to the cutoff frequency.

The plan of the paper is the following. Section 2 reviews trend estimation by local polynomial modelling and the essential properties of the corresponding LWPR filters. In Section 3 we discuss the interpretation of LWPR filters as low-pass filters and how the cutoff frequency is related to the the bandwidth, the order of the polynomial and the kernel. We also provide a comparison with a nonparametric low-pass filters widely applied in economics, due to Baxter and King (1999), and discuss the case when the reference series is difference stationary. The class of generalized DPSS filters is dealt with in section 4. In section 5 we draw our conclusions.

2 Trend estimation with locally weighted polynomial regression

Let y_t denote a time series measured at discrete and equally spaced time points. The series can be decomposed as

$$y_t = \mu_t + \varepsilon_t,$$

where μ_t is the underlying trend, and ε_t is the noise. We further assume that μ_t is a smooth but unknown deterministic function of time, which can be approximated in a neighborhood of time t by a polynomial of degree p of the time distance, j, between y_t and the neighboring observations $y_{t+j}, j = 0, \pm 1, 2, \ldots, h$. This enables to write $\mu_{t+j} \approx m_{t+j}$, with

$$m_{t+j} = \beta_0 + \beta_1 j + \cdots + \beta_p j^p, j = 0, \pm 1, \ldots, \pm h.$$

2.1 Specification: polynomial order and bandwidth

Replacing μ_{t+j} by its approximation gives the local polynomial model:

$$y_{t+j} = \sum_{k=0}^{p} \beta_k j^k + \epsilon_{t+j}, \quad j = 0, \pm 1, \ldots, \pm h. \tag{1}$$

If we assume that $\epsilon_{t+j} \sim \text{NID}(0, \sigma^2)$, then (1) is a linear Gaussian regression model with explanatory variables given by the powers of the time distance $j^k, k = 0, \ldots, p$ and unknown coefficients β_k, which are proportional to the k-th order derivatives of μ_t. Working with the linear Gaussian approximating model, we are faced with the problem of estimating $m_t = \beta_0$, i.e. the value of the approximating polynomial for $j = 0$, which is intercept β_0 of the approximating polynomial.

The model (1) can be rewritten in matrix notation as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \text{N}(\mathbf{0}, \sigma^2\mathbf{I}),$$

where $\mathbf{y} = [y_{t-h}, \cdots, y_t, \cdots, y_{t+h}]'$, $\boldsymbol{\epsilon} = [\epsilon_{t-h}, \cdots, \epsilon_t, \cdots, \epsilon_{t+h}]'$,

$$\mathbf{X} = \begin{bmatrix} 1 & -h & \cdots & (-h)^p \\ 1 & -(h-1) & \cdots & [-(h-1)]^p \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & h-1 & \cdots & (h-1)^p \\ 1 & h & \cdots & h^p \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

The degree of the polynomial is crucial in determining the accuracy of the approximation. Another essential quantity is the size h of the neighborhood around time t; in our particular setup the neighborhood consist of $H = 2h + 1$ consecutive and regularly spaced time points at which observations y_{t+j} are made. The parameter H is the *bandwidth*, but we shall also refer to h as the bandwidth parameter henceforth. We shall assume throughout that $p \leq 2h$.

2.2 Kernels

In the estimation of the unknown level, we would like to weight the observations differently according to their distance from time t. In particular, we may want to assign larger weight to the observations that are closer to t. For this purpose we introduce a kernel function $\kappa_j, j = 0, \pm 1, \ldots, \pm h$, which we assume known, such that $\kappa_j \geq 0$, and $\kappa_j = \kappa_{-j}$. Hence, the κ_j's are non negative and symmetric with respect to j.

There is a large literature on kernels and their properties. An important class, embedding several widely used kernels, is the class of Beta kernels:

$$\kappa(u) = k_{rs}\,(1-|u|^r)^s\, I(|u| \leq 1), \quad k_{rs} = \frac{r}{2B\left(s+1, \frac{1}{r}\right)}$$

with $r > 0$, $s \geq 0$, and

$$B\,(a,b) = \int_0^1 u^{a-1}\,(1-u)^{b-1}\,du$$

with $a, b > 0$ is the Beta function.

In the discrete case, up to a proportionality factor,

$$\kappa_j = \left(1 - \left|\frac{j}{h+1}\right|^r\right)^s, \quad j = 0, \pm 1, \ldots, \pm h.$$

The pair ($r = 1, s = 0$) gives the uniform kernel, $r = s = 1$ gives the triangle kernel, ($r = 2, s = 1$) the Epanechnikov kernel, $r = s = 2$ the biweight kernel, ($r = 2, s = 3$) the triweight kernel, $r = s = 3$ the tricube kernel. Other kernels are defined from parametric density functions, as in the case of the Gaussian kernel. For a comparative assessment of the various kernels see Wand and Jones (1995). The overall conclusion is that their choice is not very relevant in terms of efficiency.

2.3 Estimation and LWPR filters

We assume throughout that we are interested in the estimation of the trend at an interior point, and we will not be concerned with the treatment of end points, for which we refer to Proietti and Luati (2008).

Provided that $p \leq 2h$, the $p+1$ unknown coefficients $\beta_k, k = 0, \ldots, p$, can be estimated by the method of weighted least squares (WLS), which consists of minimising with respect to the β_k's the objective function:

$$S(\hat{\beta}_0, \ldots, \hat{\beta}_p) = \sum_{j=-h}^{h} \kappa_j \left(y_{t+j} - \hat{\beta}_0 - \hat{\beta}_1 j - \ldots - \hat{\beta}_p j^p\right)^2. \tag{2}$$

Defining $\mathbf{K} = \text{diag}(\kappa_h, \ldots, \kappa_1, \kappa_0, \kappa_1, \ldots, \kappa_h)$, the WLS estimate of the coefficients is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y}$. In order to obtain $\hat{m}_t = \hat{\beta}_0$, we need to select the first element of the vector $\hat{\boldsymbol{\beta}}$. Hence, denoting by $\mathbf{e}_1$ the $p+1$ vector $\mathbf{e}_1' = [1, 0, \ldots, 0]$,

$$\hat{m}_t = \mathbf{e}_1'\hat{\boldsymbol{\beta}} = \mathbf{e}_1'(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}\mathbf{y} = \mathbf{w}'\mathbf{y} = \sum_{j=-h}^{h} \mathrm{w}_j y_{t-j},$$

which expresses the estimate of the trend as a linear combination of the observations with coefficients

$$\mathbf{w}' = \mathbf{e}_1'(\mathbf{X}'\mathbf{K}\mathbf{X})^{-1}\mathbf{X}'\mathbf{K}. \tag{3}$$

The trend estimate is local since it depends only on the subset of the observations that belong to the neighbourhood of time t. The linear combination yielding our trend estimate is often termed a (linear) *filter*, and the weights w_j constitute its impulse responses. The latter are time invariant and carry essential information on the nature of the estimated signal; they enjoy two important properties, symmetry and reproduction of p-th degree polynomials. Symmetry ($\mathrm{w}_j = \mathrm{w}_{-j}$) follows from the symmetry of the kernel weights κ_j and the assumption that the available observations are equally spaced. As far as the second is concerned, from (3) we have that $\mathbf{X}'\mathbf{w} = \mathbf{e}_1$, or equivalently,

$$\sum_{j=-h}^{h} \mathrm{w}_j = 1, \qquad \sum_{j=-h}^{h} j^l \mathrm{w}_j = 0, \qquad l = 1, \dots, p.$$

As a consequence, the filter $\mathbf{w}$ is said to preserve a deterministic polynomial of order p, which means that if the series is $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ than the filter will reproduce it exactly, i.e. $\hat{m}_t = \beta_0 = y_t$.

2.4 Special cases

For illustrative purposes, in the sequel we shall concentrate on a few well known cases. As for the choice of the order p we shall focus on two cases: $p = 0, 1$ and $p = 2, 3$ which we shall refer to as the linear and the cubic case. The case $p = 0$ (local constant fit) yields the well-known Nadaraya-Watson estimator (Nadaraya, 1964, Watson, 1964).

Macaulay filters When the kernel is uniform, $\mathbf{K} = \mathbf{I}$, the weighting function $\mathbf{w} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_1$ is known as a *Macaulay's filter*. In the case of a local constant/linear polynomial, that is $p = 0, 1$, the filter performs the arithmetic moving average: $\mathrm{w}_j = \mathrm{w} = 1/(2h + 1), j = 0, \pm 1, \dots, \pm h$.

Epanechnikov Kernel The Epanechnikov (1969) kernel is such that

$$\kappa_j = \left[1 - \left(\frac{j}{h+1}\right)^2\right], j = 0, \pm 1, \dots, \pm h.$$

This kernel is often recommended (see Fan and Gijbels, sec. 3.2.6) as the optimal kernel minimizing the mean square estimation error in a local polynomial regression framework.

Loess Loess (Cleveland, 1979) is a LWPR method using a tricube kernel

$$\kappa_j = \left[1 - \left|\frac{j}{h+1}\right|^3\right]^3, j = 0, \pm 1, \dots, \pm h.$$

and either $p = 0, 1$, or $p = 2, 3$. The peculiar trait of Loess is that it uses a nearest neighbor bandwidth but for our purposes we will override this feature.

The Henderson filter An important class of local polynomial filters, proposed by Henderson (1916), arises as a particular case of the local cubic fit, $p = 3$ in (2). The relevance of Henderson's contribution to modern local regression is stressed in the first chapter of Loader (1999). The Henderson filters are employed for trend estimation in the X-12 nonparametric seasonal adjustment procedure. See Findley *et al.* (1998) for more details.

The problem faced by Henderson is to determine the weighting function (3) which, for a given bandwidth h and $p = 3$ provides the *smoothest* estimates of the trend. The smoothness criterion adopted by Henderson is based on the variance of the third differences of the estimates of the trend, in that the smaller the variance the greater the extent of smoothness, as the trend acceleration is subject to the least variation.

It can be shown (Kenny and Durbin, 1982) that the solution is (3) using

$$\kappa_j = \left[1 - \left(\frac{j}{h+1}\right)^2\right]\left[1 - \left(\frac{j}{h+2}\right)^2\right]\left[1 - \left(\frac{j}{h+3}\right)^2\right], j = 0, \pm 1, \dots, \pm h$$

This kernel can be seen as a discrete approximation of the triweight kernel.

3 LWPR filters as low-pass filters: the relationship between bandwidth and cutoff frequency

The LWPR filter are low-pass filters, i.e., they leave almost unchanged low frequency components, such as the trend, and attenuate high frequency fluctuations associated to the noise. This is evidenced by the transfer function of the filter

$$G_{h,p,\kappa}(\omega) = \sum_{j=-h}^{h} \mathrm{w}_j e^{-\imath\omega j}$$

where $\imath$ is the imaginary unit and $\omega \in (-\pi, \pi)$ is the frequency measured in radians. Due to the symmetry of the weights w_j, $G_{h,p,\kappa}(\omega)$ is real; $|G_{h,p,\kappa}(\omega)|$ is the gain of the filter.

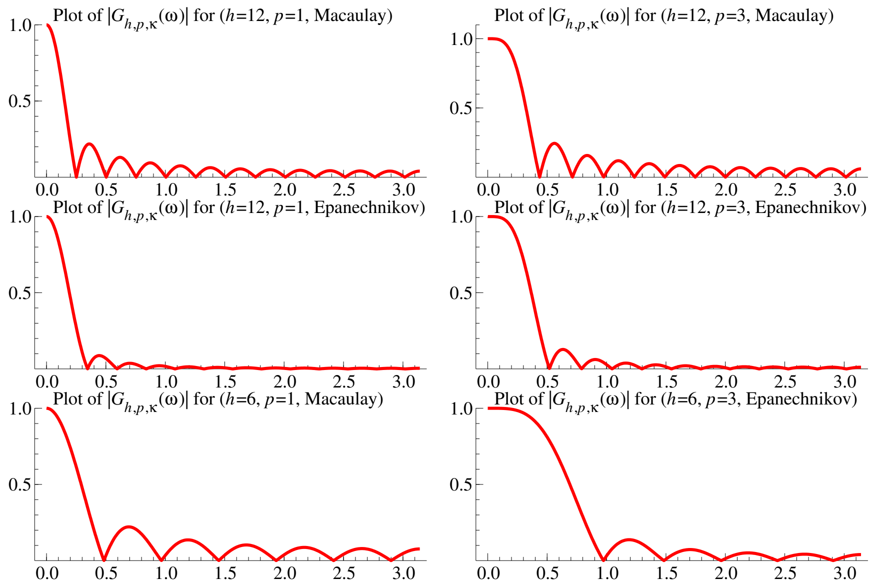
Figure 1 displays the gain of six filters arising from the following combinations:

($h = 12, p = 1$, Macaulay),	($h = 12, p = 3$, Macaulay)
($h = 12, p = 1$, Epanechnikov),	($h = 12, p = 3$, Epanechnikov),
($h = 6, p = 1$, Macaulay),	($h = 6, p = 3$, Epanechnikov).

It is clear that the filters behave differently, in the way certain components are passed. It is thus useful to understand what is the reference frequency for each filter. A common feature is $G_{h,p,\kappa}(0) = 1$, as it stems from the property $\sum_j \mathrm{w}_j = 1$. Also, the transfer function close to one at the low frequencies and close to zero at the high frequencies. Higher order kernels produce an attenuation of side lobes.

It should be recalled that an ideal low-pass filter retains only the low frequency fluctuations in the series and reduces the amplitude of fluctuations with frequencies higher than a cutoff frequency ω_c. Its

Figure 1: Gain of six LPWR filters.



transfer function takes the following form: for $\omega \in (0, \pi)$,

$$G_{lp}(\omega) = \begin{cases} 1 & \text{if } \omega \leq \omega_c \\ 0 & \text{otherwise} \end{cases}$$

As it is well known the ideal filter is available analytically, but unfeasible. Taking the inverse Fourier transform, yields the weights of the ideal filter:

$$\mathrm{w}_j = \begin{cases} \frac{\sin(\omega_c j)}{\pi j}, & j = \pm 1, 2, \ldots, \\ \frac{\omega_c}{\pi} & j = 0 \end{cases} \tag{4}$$

The filter can be approximated by truncating the weights at lead and lag h, and rescaling them so that their sum is one (see Percival and Walden, 1993, Baxter and King, 1999).

We now derive the cutoff frequency associated to any LWPR filter by a *least squares filter design* approach. The latter is usually applied for determining the filter which minimizes the squared modulus of the discrepancy between its transfer function, $G_{h,p,\kappa}(\omega)$ and that of the ideal low pass filter, $G_{lp}(\omega)$, for a given cutoff frequency ω_c. See Tufts, Rorabacher and Mosier, (1961), Tufts and Francis (1970) and Percival and Walden (1993, sec. 5.8). In our problem, instead, the filter weights are given and the cutoff is unknown. The latter will be determined as the frequency

$$\omega_c = \text{argmin}\left\{D(\omega_c; h, p, \kappa)\right\}, \quad D(\omega_c; h, p, \kappa) = \int_0^{\pi} \left|G_{h,p,\kappa}(\omega) - G_{lp}(\omega)\right|^2 d\omega.$$

for given h, p, and kernel κ.

The solution is straightforward: rewriting

$$D(\omega_c; h, p, \kappa) = \int_0^{\omega_c} |G_{h,p,\kappa}(\omega) - 1|^2 \, d\omega + \int_{\omega_c}^{\pi} |G_{h,p,\kappa}(\omega)|^2 \, d\omega,$$

and differentiating with respect to ω_c, the first order condition for the above problem is

$$\sum_{j=-h}^{h} \mathrm{w}_j e^{-\imath \omega_c j} = \frac{1}{2}, \quad (5)$$

i.e. we need to locate the frequency at which the transfer function is 1/2. The solution to equation (5) is obtained as the phase of the conjugate pair of roots with unit modulus of the polynomial $\sum_j w_j x^j$.

Figure 2 displays the result for the LWPR filters most commonly in use. The left panel shows the cutoff frequency (vertical axis) corresponding to a particular bandwidth h: as h increases, for given p and kernel function, the cutoff frequency is pushed towards zero, i.e we obtain a filter that produces a smoother trend. It is interesting to present the same results in terms of the cutoff period, $P_c = 2\pi/\omega_c$, since this enables to discover a remarkable linear relationship between cutoff period and bandwidth. For the Henderson filter, $P_c = 1.4 + 1.1h$, so that when $h = 12$, $P_c = 14.6$. The $(h = 12, p = 1$, Macaulay) filter (see first panel of figure 1) has P_c slightly greater than 40, whereas in the cubic case ($(h = 12, p = 3$, Macaulay), top right hand panel of figure 1) the period is less than 20. For the filter $(h = 6, p = 1$, Macaulay), see the bottom left panel of figure 1, the period is about 23.

It should be noticed that the intercept and slope depend on the order of the polynomial and the kernel. If we move from a linear to a cubic fit, cœteris paribus, i.e. using the same h and kernel, the cutoff period decreases. Secondly, higher order kernels yield lower cutoff periods for fixed h and p.

Figure 2 provides an effective tool for designing the LWPR filter that is appropriate for a particular cutoff frequency or period: drawing an horizontal straight line at a particular value and looking at the intersection with the hyperboles (left panel) or lines (right panel) would make immediately available the bandwidth, order and kernel of the relevant LWPR filter. We provide two applications in the next paragraph.

3.1 Comparison with the Baxter and King filter

Baxter and King (1999, BK henceforth) have popularized the use of the band-pass filter for cycle extraction in economic time series using a least squares approximation to the ideal filter that leads to a finite impulse response filter with a half bandwidth of 3 years, e.g. $h = 12$ when the data are quarterly. A band pass filter is obtained by subtracting two low-pass filters. Band-pass filtering is widely applied various fields, among which economics, to extract signals with given features. For instance, business cycle theory is concerned in the fluctuations with periodicity ranging from 1 year and half to eight years.

When the data are quarterly, the approximation of a low-pass filter with curoff frequency $\omega_c = 0.2$ ($P_c = 32$, i.e. eight years of quarterly data) as weights that are a scaled version of (4) for $h = 12$. Figure 3 plot this weights along the weights of the local linear LWPR using the Epanechikov filter and the same

Figure 2: Relationship between bandwidth (horizontal axis) and cutoff frequency ω_c (left panel), and cutoff period, $P_c = 2\pi/\omega_c$, (right panel) for different LWPR filters.

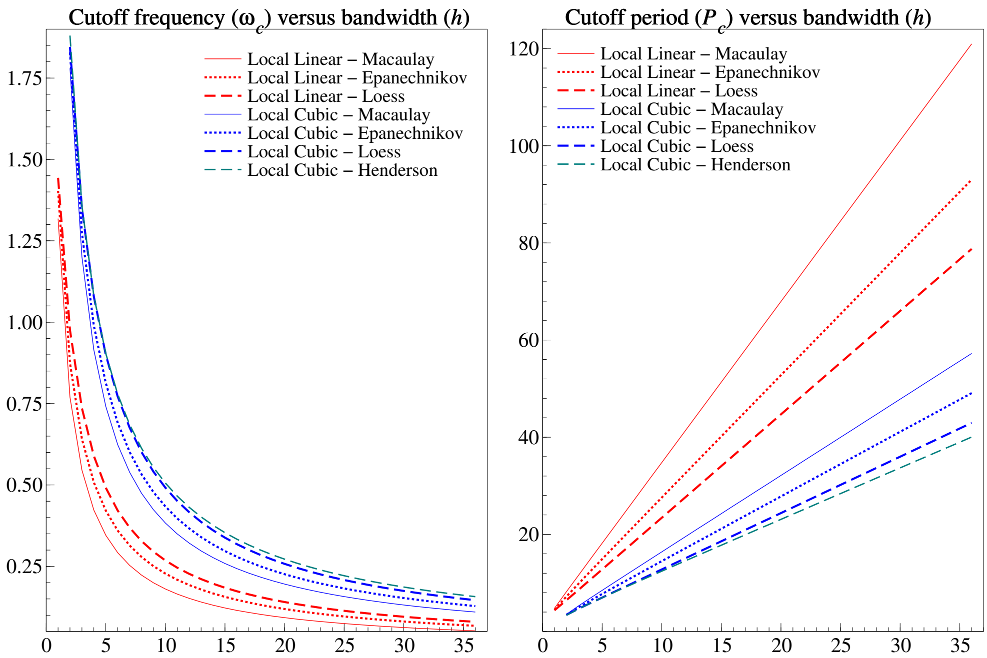
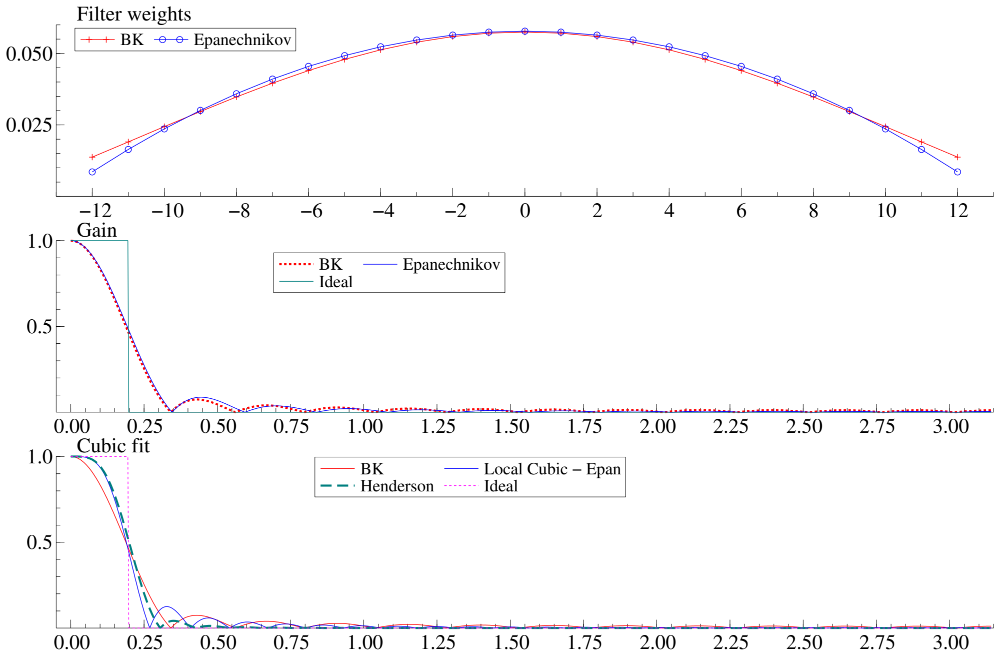


Figure 3: Comparison with Baxter and King low-pass filter with $P_c = 32$.



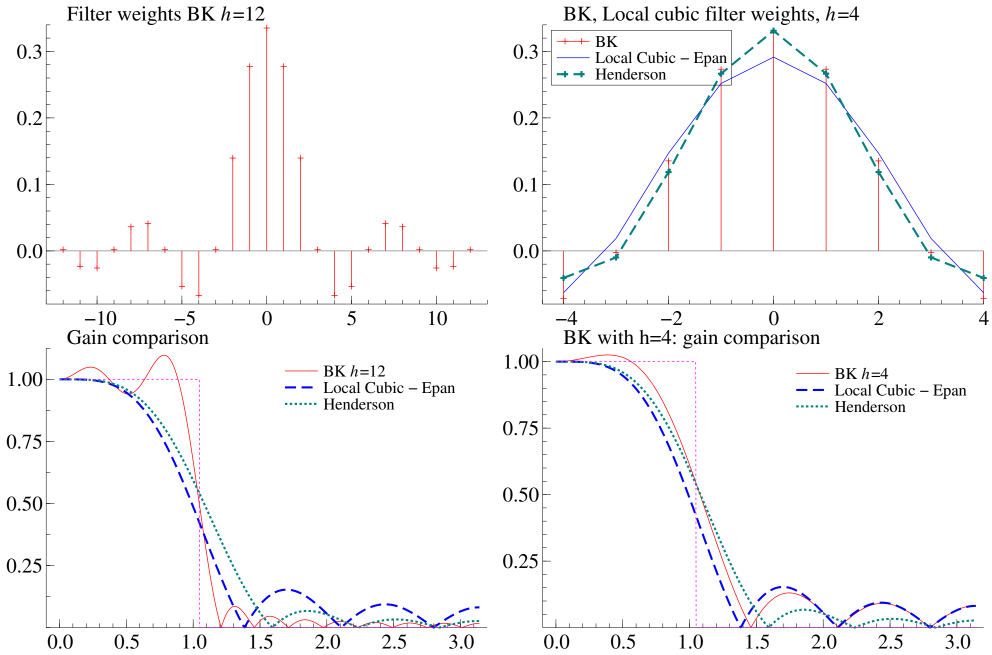
bandwidth. The latter is characterized by the same cutoff period, and thus the gains of the two filters, plotted in the central panel, are virtually indistinguishable.

Our results on the relationship between LWPR and the cutoff frequency also suggest that the local cubic polynomial regression filters ($h = 24, p = 3$, Epanechnikov) and ($h = 28, p = 3$, Henderson) are characterized by about the same cutoff frequency. The last panel of figure 3 compares their gain with that of the BK approximation to the ideal low-pass filter, showing that the two filters provide a better approximation. However, the higher polynomial order and bandwidths account for the attenuation of both leakage and the Gibbs phenomenon, i.e. the side lobes. If h is increased also for BK, then the filters behave almost identically.

The comparison with the BK filter is also useful to illustrate a possible limitation of LWPR filters, as far as the design of low-pass filters is concerned, which is a consequence of the stringent relationship between bandwidth and cutoff period. As mentioned above, BK construct a cycle filter by subtracting from the low-pass filter with cutoff period $P_c = 32$ quarters the low-pass filter with cutoff period equal to 6 quarters. The latter provides a trend estimate suppressing all the high frequency fluctuations, such as seasonal effect, with periodicity up to one year and half. Both BK low-pass filters have $h = 12$ (three years of quarterly data).

As we have shown, there exists an LWPR filter with $h = 12$ that has cutoff 32 quarters; however, we are unable to find a filter with $\bar{P} = 6$ and $h = 12$. From figure 2 it is clear that when the cutoff

Figure 4: Comparison with Baxter and King low-pass filter with $P_c = 6$.



period is small, such as $P_c = 6$, the local linear polynomial filters produce a too smooth trend estimate, even for small h. The only LWPR filters that get close to this cutoff are ($h = 4, p = 3$, Henderson) and ($h = 4, p = 3$, Epanechnikov). Figure 4 displays the filter weights and the gains of the two filters and compares them with the BK filter with $\omega_c = \pi/16$ and bandwidth 12 (left panels) and BK with bandwidth 4. Only in the second case we get filters with comparable properties, at the cost of lowering the bandwidth to a very small values.

A key consequence is that we cannot design a band-pass filter from two LWPR low-pass filters using the same h, p and kernel. At least one of the elements has to vary. As the bandwidth is the most important factor in determining the cutoff period, h qualifies as the key factor in band-pass filter design.

3.2 Modified least squares design

Christiano and Fitgerald (2003, CF henceforth) proposed a modified least squares criterion for deriving the optimal filter weights which minimizes the discrepancy function, for a given ω_c:

$$\int_0^\pi |\sum_j \mathrm{w}_j e^{-\imath\omega j} - 1|^2 \frac{d\omega}{2(1-\cos\omega)}.$$

With respect to the traditional least square problem, the discrepancy is defined with respect to a random walk process. This produces, for fixed h, a filter which differs slightly from the BK optimal approximation; see their figure 2, p. 447.

This is an interesting modification, since it brings under consideration the nature of the filtered series and its possible nonstationarity. It also makes clear that the assumption underlying the traditional least squares filter design is the assumption of trend stationarity around a white noise component.

It is thus sensible to ask whether the LWPR filter cutoff frequency changes when we take a modified least squares approach. The answer is no. The first order condition would be (5) again. The proof is very simple and follows from the straightforward the application of the fundamental theorem of calculus. This result does not contradict CF, as it uses the modified criterion for a different problem. In fact, the LWPR filter is available, whereas the cutoff frequency is unknown. In conclusion, the cutoff frequency that minimizes the discrepancy with the ideal filter is the same as that derived above in figure 2. The only change is in the curvature of the objective function, which is flatter.

4 The use of the DPSSs in low-pass filter design

We turn out our attention to a different approach to the design of a low-pass filter which can be adapted to the problem of fitting a local polynomial. This adaptation configures the main contribution of the paper to this strand of literature. The approach here considered is based on the idea of maximizing the concentration of the transfer function inside a particular frequency range, and leads to define the weights in terms of discrete prolate spheroidal sequences (DPSS).

Let $\mathbf{w}$ denote a filter characterized by the transfer function $G(\omega), \omega \in [0, \pi]$; then its concentration at frequency $\varpi \in (0, \pi)$ is defined as

$$\beta^2(\varpi) = \frac{\int_0^{\varpi} |G(\omega)|^2 d\omega}{\int_0^{\pi} |G(\omega)|^2 d\omega} \tag{6}$$

The above concentration measure was defined and analysed according to different perspectives by Tufts and Francis (1970), Papoulis and Bertran (1970), Eberhard (1973) and Slepian (1978), who also referred to (6) as an energy ratio.

The concentration can be expressed as a ratio of quadratic forms,

$$\beta^2(\varpi) = \frac{\mathbf{w}'\mathbf{A}(\varpi)\mathbf{w}}{\mathbf{w}'\mathbf{w}}, \tag{7}$$

where $\mathbf{A}(\varpi)$ is the symmetric and positive definite matrix whose generic ij-th element, $i, j = 1, \ldots, 2h+1$, is

$$a_{ij}(\varpi) = \begin{cases} \frac{\sin(\varpi(i-j))}{\pi(i-j)} & \text{for } i \neq j \\ \varpi/\pi & \text{for } i = j \end{cases}$$

as follows by replacing

$$|G(\omega)|^2 = G(\omega)G^*(\omega) = \sum_{i=-h}^{h} \mathrm{w}_i^2 + 2 \sum_{i=-h}^{h} \sum_{j=-h}^{i-1} \mathrm{w}_i \mathrm{w}_j \cos((i-j)\omega)$$

in (6) and then performing the integrations; $G^*(\omega)$ denotes the complex conjugate of $G(\omega)$.

A strategy for designing filters is, for a given bandwidth h, to choose the filter that maximizes the concentration at a given concentration frequency, under the constraint that the weights sum to one. Under this condition the filter capable of reproducing a polynomial of order 0, 1. Hence, given ϖ, the filter weights result from the solution of the following constrained maximization problem:

$$\max_{\mathrm{w}_j, j=-h,\dots,h} \beta^2(\varpi) \text{ subject to } \mathbf{w}'\mathbf{i} = 1, \tag{8}$$

where $\mathbf{i} = [1, \dots, 1]'$.

The quantity (7) is a Rayleigh quotient that reaches its maximum when it is equal to the largest eigenvalue of $\mathbf{A}(\varpi)$ (see Meyer, 2000, p. 549); let us call it $\lambda_0(\varpi)$. The corresponding eigenvector, denoted $\mathbf{v}_0$, is the filter that maximises the concentration (6) or (7) at the concentration frequency ϖ, when no restrictions are imposed to the weights except, possibly, to be of unitary norm. The filter weights arise from the normalization: $\mathbf{w} = \mathbf{v}_0/(\mathbf{v}_0'\mathbf{i})$.

The (unit norm) vector $\mathbf{v}_0$ is a subsequence of length $2h+1$ of a zeroth-order discrete prolate spheroidal sequence (DPSS). Its direct computation from the spectral decomposition $\mathbf{A}(\varpi)$ is computationally unattractive and unstable. Fortunately, it has been shown that it can be derived as the eigenvector associated to the largest eigenvalue of the symmetric tridiagonal matrix $\mathbf{C}$ whose nonzero elements are

$$c_{j+1,j+1} = \left(\frac{2h-2j}{2}\right)^2 \cos\omega, \quad j = 0, \dots, 2h$$

and

$$c_{j+1,j} = c_{j,j+1} = \frac{j(2h-j)}{2}, \quad j = 1, \dots, 2h,$$

see Slepian (1978) and the tutorial by Mathews, Breakall and Karawas (1985), where numerical solutions are also proposed. An explicit form for $\mathbf{v}_0$ cannot be obtained for all h but efficient algorithms for its computation are available also for large h (see Dhillon and Parlett, 2004).

4.1 Generalized local polynomial DPSS filter design and their relation with LWPR filters

The DPSS approach described above generates a filter that passes a local constant/linear polynomial, m_{t+j} for $p = 0, 1$; this is a simple consequence of the constraint $\mathbf{w}'\mathbf{i} = 1$. It seems therefore natural to propose a generalization by introducing further linear constraints on the weights that aim at the reproduction of higher order polynomials. We show below that there is a solution to this problem. The price to be paid is that there is no closed form solution to this problem and the solution has to be found iteratively.

For any bandwidth h, we aim at designing the filter that maximizes the concentration at a given frequency, under the constraint of reproducing a polynomial of degree p. This produces the following constrained maximization problem:

$$\max_{\mathrm{w}_j, j=-h,\dots,h} \beta^2(\varpi) \text{ subject to } \mathbf{w}'\mathbf{X} = \mathbf{e}_1', \tag{9}$$

Obviously, for $\mathbf{X} = \mathbf{i}$, we are back to the standard problem discussed above.

The objective function to be maximized is

$$\phi(\mathbf{w}, \boldsymbol{\mu}) = \frac{\mathbf{w}'\mathbf{A}(\varpi)\mathbf{w}}{\mathbf{w}'\mathbf{w}} - 2(\mathbf{w}'\mathbf{X} - \mathbf{e}_1')\boldsymbol{\mu},$$

where $\mathbf{e}_1' = [1, 0, \ldots, 0]$ and $\boldsymbol{\mu}$ is a vector of Lagrange multipliers. The first order conditions give $\mathbf{B}(\varpi)\mathbf{w}(\mathbf{w}'\mathbf{w})^{-1} = \mathbf{X}\boldsymbol{\mu}$ and $\mathbf{w}'\mathbf{X} = \mathbf{e}_1'$, where $\mathbf{B}(\varpi) = \mathbf{A}(\varpi) - \beta^2(\varpi)\mathbf{I}$. If $\mathbf{w}$ is an eigenvector of $\mathbf{A}(\varpi)$, associated to the eigenvalue $\beta^2(\varpi)$, then $\mathbf{B}(\varpi)$ is singular and the first order conditions imply that $\boldsymbol{\mu}$ is the null vector, i.e. the problem collapses to the unconstrained maximization discussed above. On the other hand, if $\mathbf{w}$ is not an eigenvector of $\mathbf{A}(\varpi)$, then $\mathbf{B}(\varpi)$ is nonsingular and its inverse $\mathbf{B}(\varpi)^{-1}$ exists. Rearranging, $\mathbf{w}(\mathbf{w}'\mathbf{w})^{-1} = \mathbf{B}(\varpi)^{-1}\mathbf{X}\boldsymbol{\mu}$ and $\boldsymbol{\mu}'\mathbf{X}'\mathbf{B}(\varpi)^{-1}\mathbf{X} = \mathbf{e}_1'(\mathbf{w}'\mathbf{w})^{-1}$, the solution of (8) is the filter

$$\mathbf{w}' = \mathbf{e}_1'(\mathbf{X}'\mathbf{B}(\varpi)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}(\varpi)^{-1}. \tag{10}$$

It is evident that this is an implicit solution, as $\mathbf{B}(\varpi)$ depends on $\mathbf{w}$ through $\beta^2(\varpi)$.

Hence, the solution can be obtained by the following iterative algorithm.

i. Start from a symmetric LWPR filter $\mathbf{w}_1$ that satisfies the polynomial reproducing constraints.

ii. For $r = 2, 3, \ldots,$

1. compute the concentration:
$$\beta_r^2(\varpi) = \frac{\mathbf{w}_r'\mathbf{A}(\varpi)\mathbf{w}_1}{\mathbf{w}_r'\mathbf{w}_r};$$
2. construct the matrix $\mathbf{B}_r(\varpi) = \mathbf{A}(\varpi) - \beta_r(\varpi)^2\mathbf{I}$;
3. update the solution
$$\mathbf{w}_r' = \mathbf{e}_1'(\mathbf{X}'\mathbf{B}_r(\varpi)^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{B}_r(\varpi)^{-1}. \tag{11}$$

iii. Repeat until convergence, i.e. for some r the discrepancy $\|\mathbf{w}_r - \mathbf{w}_{r-1}\|^2$ is negligible.

We would like to point out the correspondence of (11) with the expression (3) for the LWPR filters. The weights now arise from generalized least square regression using $\mathbf{B}_r(\varpi)^{-1}$ en lieu of the kernel. We shall refer to $\beta_r^2(\varpi)$ as the *restricted concentration* index. It is interpreted as the maximum concentration at the frequency ϖ that a filter reproducing a polynomial of a degree p can achieve.

The properties of the filter (11) essentially follow by those of the matrix $\mathbf{A}(\varpi)$, since $\mathbf{B}_r(\varpi)$ differs from $\mathbf{A}(\varpi)$ only for the diagonal elements. Denote $\sigma(\mathbf{A}(\varpi)) = \{\lambda_0(\varpi) > \lambda_1(\varpi) > \ldots > \lambda_{2h}(\varpi)\}$ the spectrum of $\mathbf{A}(\varpi)$, which satisfies $\sigma(\mathbf{A}(\varpi)) \subset (0, 1)$ (Eberhard, 1973); A first result $\mathbf{A}(\varpi)$ and $\mathbf{B}_r(\varpi)$ have the same eigenvectors: $\mathbf{B}(\varpi)\mathbf{v}_k = \mathbf{A}(\varpi)\mathbf{v}_k - \beta_r^2(\varpi)\mathbf{v}_k = \lambda_k(\varpi)\mathbf{v}_k - \beta_r^2(\varpi)\mathbf{v}_k = (\lambda_k(\varpi) - \beta_r^2(\varpi))\mathbf{v}_k = \xi_k(\varpi)\mathbf{v}_k$, where $\xi_k(\varpi) = \lambda_k(\varpi) - \beta_r^2(\varpi)$. Secondly, the restricted concentration $\beta_r^2(\varpi)$ can take values ranging from the minimum to the maximum eigenvalue of $\mathbf{A}(\varpi)$ (Meyer, 2000, p.549):

$$\lambda_0(\varpi) > \beta_r^2(\varpi) > \lambda_{2h}(\varpi). \tag{12}$$

We remind here that $\lambda_k(\varpi)$ is a measure of concentration,

$$\lambda_k(\varpi) = \frac{\mathbf{v}_k'\mathbf{A}(\varpi)\mathbf{v}_k}{\mathbf{v}_k'\mathbf{v}_k}$$

where $\mathbf{v}_k$ is the corresponding eigenvector of $\mathbf{A}(\varpi)$ and the unit norm vector $\mathbf{v}_0$ is the solution of (8). The eigenvalues of $\mathbf{B}_r(\varpi)$ are those of $\mathbf{A}(\varpi)$ shifted back of $\beta_r^2(\varpi)$, i.e. $\sigma(\mathbf{B}_r) = \{\xi_0(\varpi) > \xi_1(\varpi) > \ldots > \xi_{2h}(\varpi)\}$.

4.2 Filter design with generalized DPSS

For the design of a generalized DPSS filter, three parameters are crucial: the bandwidth, the order of the polynomial it can reproduce without bias, p, and the concentration frequency ϖ. Henceforth we shall write DPSS(h, p, ϖ). Such a filter has associated a value of the concentration index.

The concentration frequency has often been interpreted as a genuine cutoff frequency. We will argue that this interpretation, which has been responsible for the dismissal of this approach for low-pass filter design, is unwarranted. We are going to illustrate that perhaps ϖ is better interpreted as a kernel. Indeed, this interpretation is in line with the view of DPSS filters as *convergence factors*, supported for instance by Percival and Walden (1993, p. 182). Moreover, the comparison of expression (11) with the general expression of an LWPR filter (3) reinforces such interpretation: the only way ϖ affects the shape of the filter is via the matrix $\mathbf{B}(\varpi)$.

Figure 5 displays the value of percentage concentration, $100 \times \beta^2(\varpi)$, as a function of h and either $\bar{P}$ (top left panel) or $\varpi \in (0, \pi)$ (top right panel), for the standard, i.e. locally constant/linear, DPSS filters . The value of $\beta^2(\varpi)$ value is obtained from the first eigenvalue of the matrix $\mathbf{A}(\varpi)$. To enhance the interpretability of the underlying patterns, we present also the contour plots of $100 \times \beta^2(\varpi)$ on the $(h, \bar{P})$ plane (bottom left panel) and (h, ϖ) plane (bottom right panel). For given h, the concentration is a decreasing function of $\bar{P}$ and an increasing function of ϖ. For small periods (high concentration frequencies), and for $\bar{P} < h$, the DPSS filter is 100% concentrated in the desired frequency range. The concentration is poor, instead, when the concentration period (frequency) is high (small) and the bandwidth is small; the latter provides a limiting factor.

The concentration contours enable to assess the trade-off between h and the concentration frequency: to achieve a given concentration level we can decrease h, providing we increase $\bar{P}$ along a straight line, whose intercept and slope depend on the level of $\beta^2(\varpi)$, or equivalently decrease ϖ along an hyperbole. A remarkable feature is indeed that the combinations of $(h, \bar{P})$ giving the same concentration lie along straight lines, whereas the equi-concentration points describe hyperboles on the (h, ω_c) plane. The plot also informs us that if $h = 12$ (e.g. three years of quarterly data, so that 25 consecutive observations are used to estimate the trend), we can estimate with unit concentration fluctuations with period up to 22 time units. The concentration for $\bar{P} = 32$ (eight years of quarterly data) is about 94%. To get a concentration equal to 1 for this concentration period a bandwidth $h \geq 18$ would be needed.

The same considerations hold for the local quadratic/cubic DPSS filter given in (11). Figure 6 display the concentration for such filters. The major difference lies in the vertical shift of the concentration,

Figure 5: Concentration for standard DPSS filters, as a function of the concentration period and frequency.

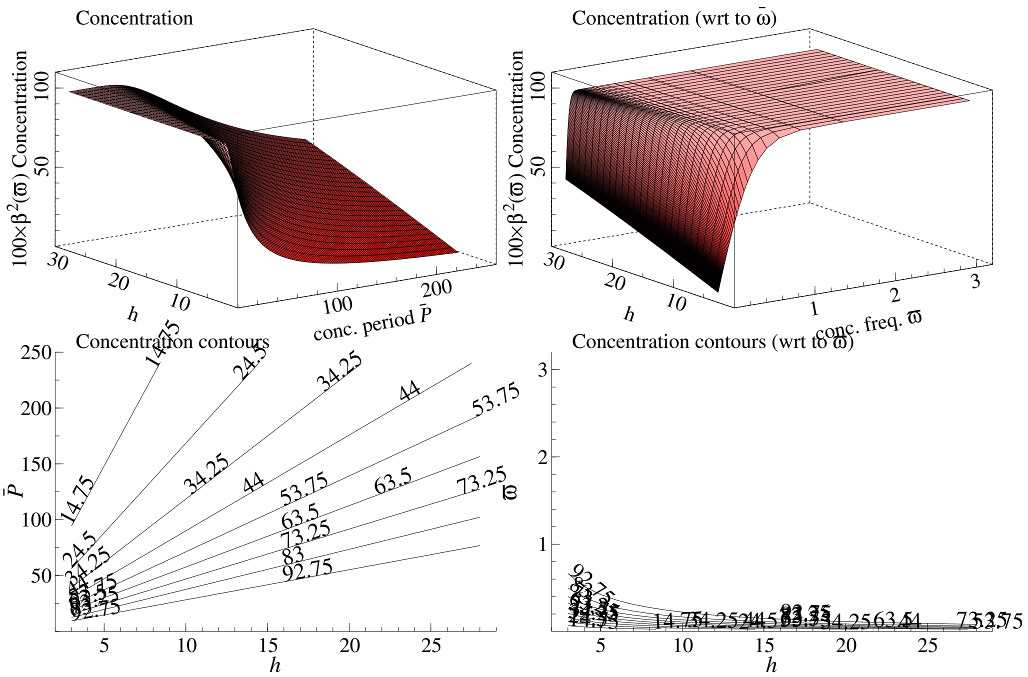
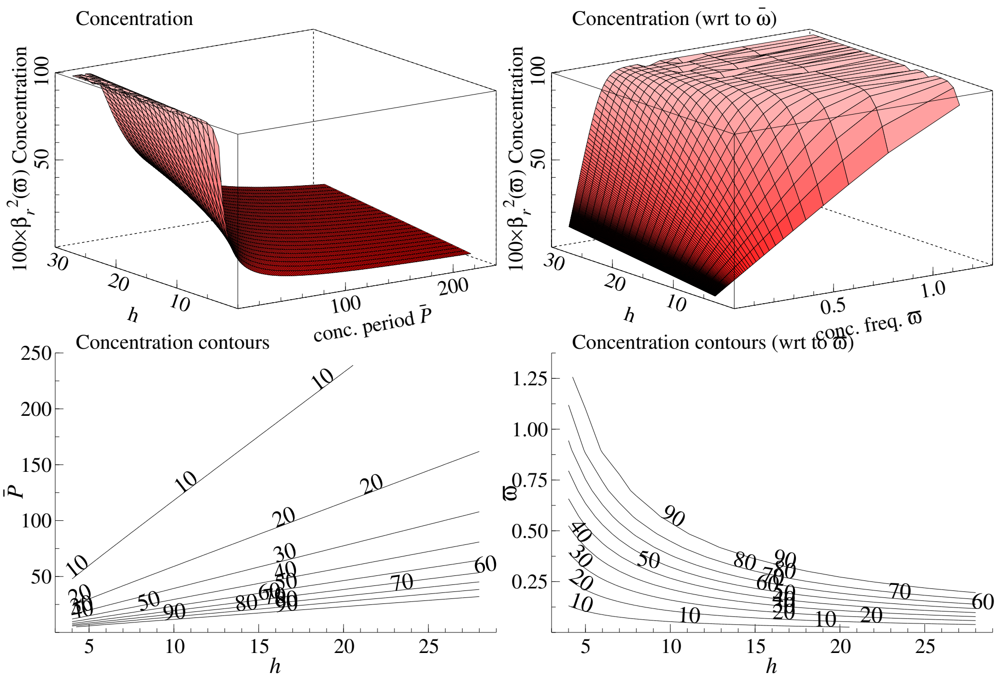


Figure 6: Concentration for cubic DPSS filters, as a function of the concentration period and frequency.



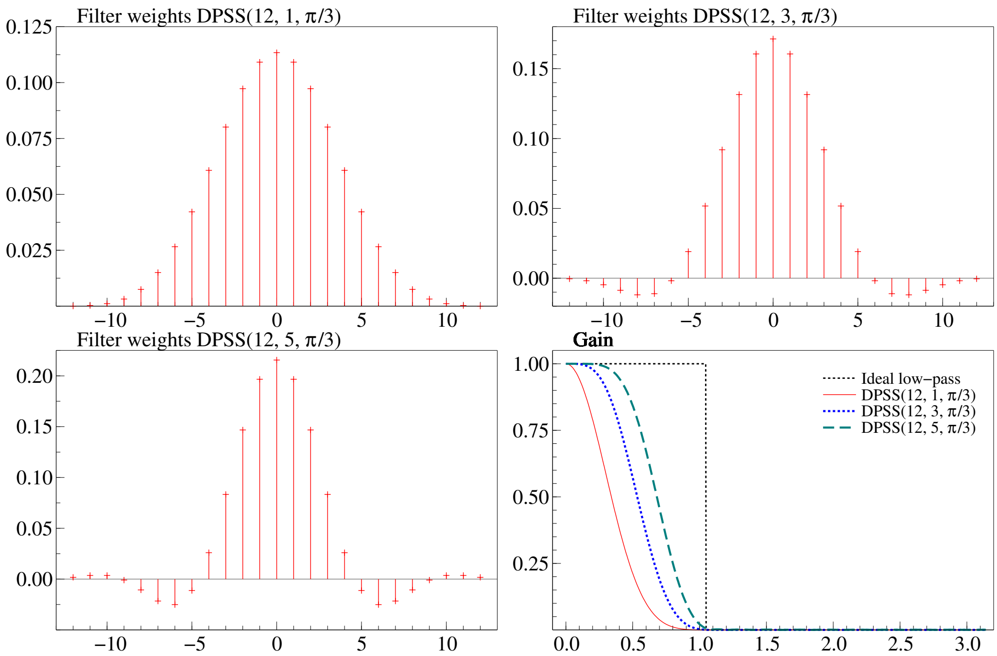
which results from the fact that, for given h and either $\bar{P}$ or ϖ, $\beta_r^2(\varpi) \leq \beta^2(\varpi)$. In other words, if we aim at designing a filter with the same concentration, but that passes a cubic, rather than linear, polynomial, we have to increase either h or $\bar{P}$, or both. The linear and hyperbolic patterns of the contours are present also in this case.

As we stated before, the concentration frequency is a different concept from the cutoff frequency: when h is high with respect to the concentration frequency, so that for that particular frequency we are able achieve a unit concentration, the gain of the DPSS filter will monotonically decrease from 1 to 0 in the interval $[0, \varpi]$. In general, the gain at the concentration frequency is smaller than 1/2.

Figure 7 illustrates this. It displays the filter weights and the gains of three DPSS$(12, p, \pi/3)$ filters. It is clear that the three filters, obtained for $p = 1$ (standard case) and $p = 3$ and 5, designed for a concentration frequency equal to $\pi/3$, are characterized by different cutoff frequencies, which are in turn smaller than the concentration frequency. In particular, if the order of the filter increases the localization of the weight pattern increases and the trend estimates are less smooth.

According to the same least squares design principle that was adopted for LWPR filters, the effective cutoff frequency ω_c of any DPSS(h, p, ϖ) filter is computed as the frequency at which the transfer function of the filter is equal to one half. Figure 8 presents the results of such computations; in particular, the right panel displays the contours of the cutoff period $P_c = 2\pi/\omega_c$, where ω_c is the frequency at which

Figure 7: Filter weights for generalized DPSS filters, with $h = 12$ and $\varpi = \pi/3$, as a function of the order of the polynomial p.



the gain of the filter equals 1/2, in the standard DPSS case ($p = 0, 1$), as a function of the bandwidth (horizontal axis) and the concentration period (vertical axis). For convenience, we overlay the contours of $100 \times \beta^2(\varpi)$, arising from the same $(h, \bar{P})$ combination (see also figure 5). The plot provides the pairs $(h, \bar{P})$ that produce the same cutoff period. For instance, to have a cutoff of 26 quarters (6 years and a half) we could choose a large $\bar{P}$ and a small bandwidth or a large h and a small $\bar{P}$. The first solution would however provide a poor concentration, e.g. not greater than 20%. The second solution would provide a filter with excellent properties (100% concentration), but more demanding in terms of data points required to estimate the signal (and more difficult to adapt at the extreme of the sample period).

The right panel displays the relationship between the cutoff period P_c and the bandwidth for the case $p = 0, 1$; the different curves correspond to different values of $\bar{P}$; the relationship is nonlinear, but we may appreciate the fact that as $\bar{P}$ increases the curves converge to a straight line that correspond to the Macaulay LWPR filter with the same bandwidth and $p = 0, 1$. In other words, for a fixed bandwidth, if we let $\bar{P}$ (ϖ) increase (decrease), the cutoff period (frequency) increases (decreases) up to an upper limit, which is the value of the cutoff period for the local linear Macaulay filter.

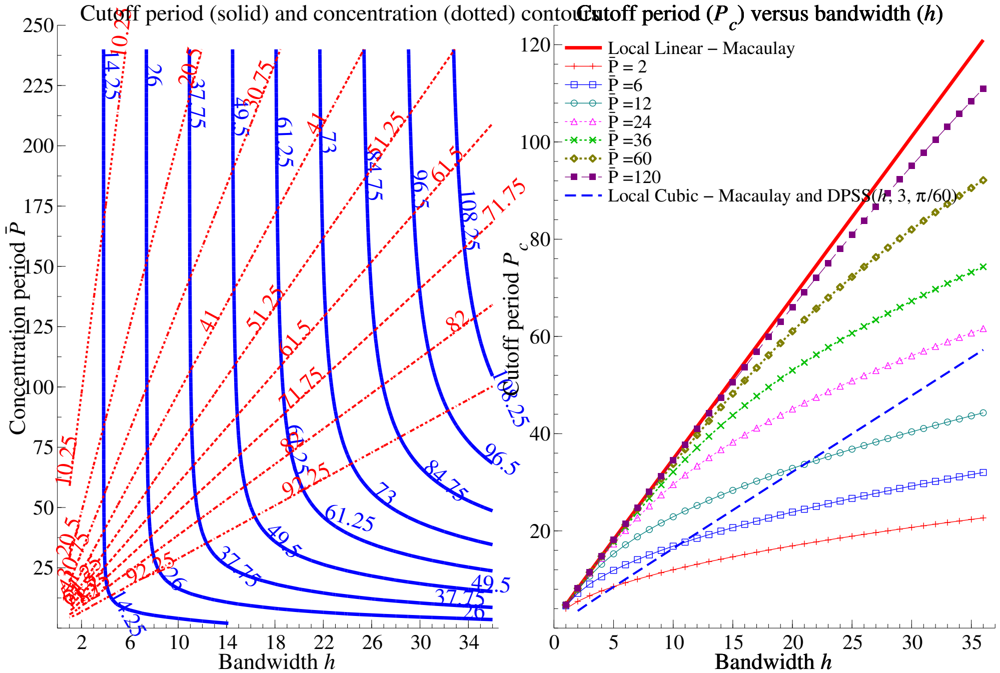
This finding actually generalizes to any order p. In general, for a given h the limiting DPSS(h, p, ϖ) filter as ϖ decreases, is the LWPR(h, p, Macaulay) filter. This is illustrated in the same plot by the dashed line which refers to the local cubic Maculay filter (see also figure 2) and is already coincident with the bandwidth-cutoff period relationship for the DPSS($h, 3, \pi/60$) filter. Another conclusion that we draw from the graph is that the generalized DPSS approach can reveal itself more flexible for the design of band-pass filters, with respect to the LWPR approach. In particular the concentration frequency, or period, can be seen as a device that is responsible for the continuous adaptation of the kernel of the local polynomial fit.

5 Concluding remarks

The paper has first aimed at characterizing local polynomial filters as low-pass filters, by investigating the relationship between the cutoff frequency and the three ingredients of the design: the bandwidth, the order of the polynomial, and the kernel. The conclusions of this investigation provide guidance over the choice of the three parameters for designing a low-pass filter with a particular cutoff frequency. They also point out that the strategy of designing a band-pass filter (e.g. for isolating the business cycle fluctuations) by subtracting two low-pass filters using the same bandwidth, kernel and polynomial order is unfeasible, as at least one of the elements has to change.

Secondly, the paper has proposed a generalized DPSS filter design approach that depends on three parameters: the bandwidth, the order of the polynomial that is reproduced by the filter, and the concentration frequency. The paper has investigated how this three parameters can be combined so as to design a low-pass filter with a preassigned cutoff frequency. The overall conclusion is that the DPSS filters have more flexibility and can provide a more coherent design for band-pass filtering. The preferred strategy would be to contrast two DPSS low-pass filters using the same bandwidth and polynomial order, but different concentration frequencies.

Figure 8: DPSS(h, p, ϖ) filters. Contour plots of the cutoff period $P_c = 2\pi/\omega_c$ as a function of the bandwidth h, and the concentration period (left panel, case $p = 0, 1$). Cutoff period as a function of bandwidth for different values of $\bar{P}$ (rigth panel).



References

Anderson T.W. (1971), *The Statistical Analysis of Time Series*, Wiley.

Baxter, M., King, R.G. (1999). Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series. *The Review of Economics and Statistics*, 81, 575593.

Behrens R.T. (1994), Signal Processing Applications of Oblique Projection Operators, *IEEE Transactions on Signal Processing*, 42, 6, 1413-1424.

Christiano L.J., Fitzgerald T.J., (2003), The band pass filter, *International Economic Review*, 44, 435-465.

Cleveland W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots, *Journal of the American Statistical Society*, 74, 368, 829-836.

Cleveland W.S., Cleveland R.B, McRae J.E., Terpenning I., (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess (with Discussion), *Journal of Official Statistics*, 6, 3-73.

Cleveland W.S. and Loader C.L. (1996). Smoothing by Local Regression: Principles and Methods. In W. Härdle and M. G. Schimek, editors, *Statistical Theory and Computational Aspects of Smoothing*, 1049. Springer, New York.

Dhillon I.S., Parlett B.N. (2004), Multiple Representations to Compute Orthogonal Eigenvectors of Symmetric Tridiagonal Matrices, *Linear Algebra and its Applications*, 387, 1-28.

Eberhard A. (1973), An Optimal Discrete Window for the Calculation of Power Spectra, *IEEE Transactions on Audio and Electroacoustics*, 21, 1, 37-43.

Epanechnikov V.A. (1969), Nonparametric Estimation of a Multivariate Probability Density, *Theory of Probability and Applications*, 14, 153-158.

Fan J. and Gjibels I. (1996). *Local Polynomial Modelling and its Applications*, Springer.

Fan J. and Yao Q. (2005). *Nonlinear Time Series. Nonparametric and Parametric Methods*, Chapman and Hall, New York.

Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen B. (1998). New Capabilities and Methods of the X12-ARIMA Seasonal Adjustment Program, *Journal of Business and Economic Statistics*, 16, 2.

Härdle, W. (1990), *Applied Nonparametric Regression*, Econometric Society Mopnographs, Cambridge University Press, Cambridge U.K.

Hastie T.J and Tibshirani R.J. (1990). *Generalized Additive Models*, Chapman and Hall, London.

Henderson R. (1916), Note on Graduation by Adjusted Average, *Transaction of the Actuarial Society of America*, 17, 43-48.

Kendall M. G. (1973). *Time Series*, Oxford University Press, Oxford.

Kendall M., Stuart, A., and Ord, J.K. (1983). *The Advanced Theory of Statistics*, Vol 3. C. Griffin.

Kenny P.B., and Durbin J. (1982). Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series, *Journal of the Royal Statistical Society A*, 145, I, 1-41.

Ladiray, D. and Quenneville, B. (2001). *Seasonal Adjustment with the X-11 Method* (Lecture Notes in Statistics), Springer-Verlag, New York.

Landau H.J., Pollak H.O. (1961), Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, II, *The Bell System Technical Journal*, 40, 65-84.

Landau H.J., Pollak H.O. (1962), Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, III, *The Bell System Technical Journal*, 41, 1295-1336.

Loader, C. (1999). *Local Regression and Likelihood.* Springer-Verlag, New York

Lii K.S., Rosenblatt M. (2008), Prolate spheroidal spectral estimates, *Statistics and Probability Letters*, 78, 11, 1339-1348.

Macaulay F.R. (1931), *The Smoothing of Time Series*, NBER.

Mathews J.D., Breakall J.K., Karawas G.D. (1994), The Discrete Prolate Spheroidal Filter as a Digital Signal Processing Tool, *IEEE Transactions on Acoustics, Speech and Signal Processing*, 33, 6, 1471-1478.

Meyer C.D., (2000), *Matrix Analysis and Applied Linear Algebra*, SIAM.

Musgrave J. (1964). A Set of End Weights to End All End Weights. Working paper. Census Bureau, Washington.

Nadaraya, E.A. (1964). On estimating regression, *Theory of Probability and its Applications*, 10, 186-190.

Papoulis A., Bertran M.S. (1970), Digital Filtering and Prolate Functions, *IEEE Transactions on Circuit Theory*, 19, 6, 674-681.

Percival D., Walden A. (1993). *Spectral Analysis for Physical Applications*. Cambridge University Press.

Proietti T., Luati A. (2008), Real Time Estimation in Local Polynomial Regression, *Annals of Applied Statistics*, 2, 4, 1523-1553.

Ruppert D., Wand, M.J. and Carroll R.J. (1989). *Semiparametric regression*, Cambridge University Press.

Slepian D., Pollak H.O. (1961), Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty, I, *The Bell System Technical Journal*, 40, 43-64.

Slepian D. (1978), Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty - V: The Discrete Case, *The Bell System Technical Journal*, 57, 1371-1430.

Slepian D. (1983), Some Comments on Fourier Analysis, Uncertainty and Modeling, *SIAM Review*, 25, 3, 379-393.

Thomson D.J. (1982), Spectrum Estimation and Harmonic Analysis, *Proceedings of the IEEE*, 70, 9, 1055-1096.

Tufts D.W., Francis J.T. (1970), Designing Digital Low-Pass Filters- Comparison of Some Methods and Criteria, *IEEE Transactions on Audio and Electroacoustics*, 18, 4, 487-494.

Tufts D.W., Rorabacher D.W., Mosier W.E. (1970), Designing Simple, Effective Digital Filters, *IEEE Transactions on Audio and Electroacoustics*, 18, 142-158.

Wand M.P. and Jones M.C. (1995). *Kernel Smoothing*, Monographs on Statistics and Applied Probability, 60, Chapman&Hall.

Watson, G.S. (1964). Smooth regression analysis, *Sankhyā* Series A, 26, 359-372.

Xu W.Y., Chamaz C. (1984), On the Periodic Discrete Prolate Spheroidal Sequences, *SIAM Journal on Applied Mathematics*, 44, 6, 1210-1217.