# SMOOTHING BY SAVITZKY-GOLAY AND LEGENDRE FILTERS

PER-OLOF PERSSON* AND GILBERT STRANG*

**1. Introduction.** This paper is about the definition and effectiveness and fast implementation of a particular family of smoothing filters. These "Savitzky-Golay filters" are popular in spectroscopy. But to filter experts in other areas they are virtually unknown! Since the filters are constructed in a very natural way, they allow analysis and explanation. They can give excellent results provided the filter length is correctly chosen, and they deserve to be understood.

Their construction is based on a least squares fit to each window of data by a polynomial of fixed degree $n$. The (smoothed) output value is taken at the center of the window. Already this may raise doubts. Starting from an ideal lowpass filter (with one-zero frequency response), its least squares approximation is not a successful favorite. We would be truncating the slowly convergent Fourier series of a step function (with Gibbs phenomenon at the jump). Minimizing the maximum error produces equiripple filters that generally perform better. On the other hand equiripple filters don't preserve moments of the input signal, which spectroscopists want. Savitzky-Golay fits the data by low-degree polynomials in the *time domain*, not high-degree polynomials in the frequency domain. The construction is robust for long filters, so the window can and should match the intrinsic scale of the input signal.

We will give explicit formulas for the filter coefficients (of course the formulas are well established for polynomials of low degree $n$). For all degrees, we show how the Savitzky-Golay filters come directly from Chebyshev's construction in 1854 of "discrete orthogonal polynomials". The most direct approach is to orthogonalize the $n+1$ vectors $(1, 1, \ldots, 1), (0, 1, \ldots, N-1), \ldots, (0^n, 1^n, \ldots, (N-1)^n)$, which are the columns of a rectangular Vandermonde matrix. Least squares is simplified, as always, by orthogonality. The polynomials satisfy a three-term recurrence, and a Christoffel-Darboux sum formula. All the classical properties extend to these polynomial vectors $t_n$, and lead to concise formulas for the filters.

The continuous analogue of Chebyshev's construction produces the Legendre polynomials. (It is Gram-Schmidt on the functions $1, x, x^2, \ldots, x^n$. The inner product is an integral instead of a sum.) Naturally the limit of the discrete polynomial vectors, with suitable scaling, is Legendre's family of continuous polynomials $P_n(x)$. If we sample these polynomials, we are extremely close to Savitzky-Golay. In fact, for reasons of simplicity

*Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139 (persson@math.mit.edu, gs@math.mit.edu). This research was supported by the Singapore-MIT Alliance.

and speed, we recommend the Legendre-based filters (you will see that they represent the leading terms in the Chebyshev-Savitzky-Golay formulas). Legendre has the extra advantage, in the case of irregularly spaced or missing data, that the polynomials stay the same and it is only the sampling points that change.

Our forthcoming paper [9] will provide software in pseudo-code to execute both fast transforms (Savitzky-Golay and Legendre-based). The normal implementation of a length $N$ filter involves $N$ multiplications in each window. But a polynomial of degree $n$ is determined by $n+1$ coefficients. Bromba and Ziegler in [1] and Scott and Scott in [11] have shown how each output value requires only $\mathcal{O}(n+1)$ steps, using the previous outputs recursively. For degree $n=0$, the filter is a simple average (mean filter) over the window. Shifting the window adds one new sample and drops one old sample. The average over that new window updates the previous average by using those two samples:

$$\text{new average} = \text{old average} + (x_{\text{newest}} - x_{\text{oldest}})/N. \tag{1}$$

For a higher $n$, the recursive correction involves polynomial degrees lower than $n$ (and stability is a significant problem). It is fast multiplication by Toeplitz matrices with "polynomial rows".

The important question, and the hardest to answer, is the effectiveness of these filters. We will report on experiments that largely justify their use (for a correctly chosen filter length!) in a significant class of applications. Our simplest model is a single Gaussian corrupted by white noise. In this case we can analyze the standard rule that the filter window should match the width of the Gaussian at half maximum. A correction term can reflect the degree $n$. By comparing Savitzky-Golay and Legendre with equiripple, the reader can see how to match the choice of filter with the application.

In summary, the four parts of our final paper [9] will attempt to provide:

1. Explicit formulas for Savitzky-Golay from Chebyshev's discrete orthogonal polynomials $q_n(x)$. The filter's impulse response (a multiple of $q_{n+1}(k)/k$) is the least squares approximation of a discrete delta function.
2. New Legendre-based polynomial filters from a continuous analogue of the same construction. Applying the Christoffel-Darboux sum formula leads us to propose symmetric filters of even degree $n$ formed by sampling a multiple of Legendre's $P_{n+1}(x)/x$.
3. Numerical experiments and analysis for "Gaussian signals plus noise" to determine the characteristics of these filters. We could study also the asymptotics, for large (and moderate) $N$ and $n$, in the time and frequency domains.
4. Fast implementation of both filters by a stabilized recursion with $\mathcal{O}(n+1)$ steps per output sample.

We want to say clearly: These are not tremendously powerful filters. Their great virtues are simplicity and speed. These properties can be preserved as the filters are improved.

We have just received a far-reaching paper by Bromba and Ziegler [5], which goes beyond our Chebyshev formulas in Section 2. The authors apply Christoffel-Darboux to obtain weighted least squares filters from several classical polynomials. By increasing the weighting parameter $L$, they approach the "maxflat filter" which is central to wavelet theory and is associated with Krawtchouk polynomials.

This paper [5] deserves the reader's attention (with excellent references, including what may be the first appearance of the maxflat filter). We hope that our further analysis of the Legendre-based filters (and our codes) will be helpful. In estimating the optimal filter length $N$, Section 4 proves the unexpected identity $\int_{-1}^{1}(P_{n+1}(x)/x)^2\,dx = 2$ for even $n$. This might or might not be new!

**2. Savitzky-Golay and Chebyshev.** The Savitzky-Golay filter $SG(N,n)$ is linear and shift-invariant. It acts on a vector of input samples $x(k)$ to produce a smoothed vector $y(k)$. On the window of $N = 2M+1$ samples $x(-M),\ldots,x(M)$, we find the best least squares fit by a polynomial vector $p(-M),\ldots,p(M)$ of specified even degree $n$. The output $y(0)$ from the filter is the value $p(0)$ at the center of the window. The same process applies to the samples $x(k-M),\ldots,x(k+M)$ when the window is shifted by $k$ time steps. The filter output $y(k)$ is the center value (at time $k$) of the degree $n$ least squares fit to the $2M+1$ samples.

An adjustment is required near the sample boundaries, when the window extends beyond the beginning or the end of the input vector $x(k)$. A simple option is "symmetric extension" of the signal at both ends. (This is usually superior to zero-padding, unless the signal amplitudes are negligible at the ends.) Assume now that $x(k)$ is infinite in both directions, $-\infty < k < \infty$. In practice the degree is $n = 2$ or $n = 4$ and certainly $n \ll N$. A "polynomial vector of degree $n$" is a vector of sample values $p(k)$ of an ordinary degree $n$ polynomial $p(x)$.

Because each Savitzky-Golay filter is linear and shift-invariant, it is enough to find its response to the unit impulse $x(k) = \delta(k)$. That response is certainly $SG(k) = 0$ for $|k| > M$, since for these $k$ the window of values $x(k-M),\ldots,x(k+M)$ will contain all zeros. *The problem is to find the (symmetric) degree $n$ polynomial vector $SG = (SG(-M),\ldots,SG(0),\ldots,SG(M))$ that best fits the discrete impulse $\delta_N = (0,\ldots,1,\ldots,0)$.* Then the filter acts on any input vector $x$ by convolution $y = SG * x$ with that impulse response:

$$y(k) = \sum_{j=-M}^{M} SG(j)x(k-j). \tag{2}$$

This reduction to an impulse input and its response $SG = SG * \delta_N$ is standard. The least squares problem for the best fit becomes $Vc = \delta_N$ with $N$ equations and $n+1$ unknowns $(c_0, \dots, c_n)$. Normally these are the coefficients of the best polynomial $SG(x) = c_0 + \dots + c_n x^n$. The rectangular matrix $V$ consists of the first $n+1$ columns of a Vandermonde matrix: $V_{ij} = i^j$ for $-M \leq i \leq M$ and $j = 0, \dots, n$ (with $0^0 = 1$). In *Numerical Recipes* [10] the coefficients $c$ are found from the normal equations $V^{\mathrm{T}}Vc = V^{\mathrm{T}}\delta_N$. In the MATLAB Signal Processing Toolbox, the code ***sgolay*** orthogonalizes the columns of $V$ (creating this matrix at extravagant expense). But this "QR factorization" is what Chebyshev did 150 years ago! His formulas lead to a straightforward expression for $SG(x)$, depending on $n$ and $N$.

It is understood that for each $n$ and $N$ this is a specific least squares problem, whose solutions can be tabulated (as chemists have done for small $n$). Lengths like $N = 51$ or $N = 101$ are not uncommon, following a reasonable rule of thumb for Gaussian inputs: $N$ should match the number of samples within the bump at half maximum. Thus the window width matches the scale of the input signal.

To repeat, Chebyshev orthogonalized the columns of the Vandermonde matrix $V$. The $j$th column of the new matrix $Q$ is still a polynomial vector of degree $j$, since Gram-Schmidt subtracts multiples of earlier columns (which are polynomials of lower degree). By changing to an orthogonal basis $q_0, \dots, q_n$, the projection $SG(N,n)$ of $\delta_N$ onto the column space of $V$ (the discrete polynomials of degree $n$) is a sum of one-dimensional projections. The projections are given in (8) below, and their sum $SG$ is in (15). This is our explicit formula.

EXAMPLE. *Consider the approximation by a parabola (degree $n = 2$) to the delta vector $\delta = (0,0,1,0,0)$ with $N = 5$. The coefficients of the best polynomial $SG(x) = c_0 + c_1 x + c_2 x^2$ are the least squares solution of*

$$
Vc = \begin{bmatrix} 1 & -2 & (-2)^2 \\ 1 & -1 & (-1)^2 \\ 1 & -0 & 0^2 \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ c_2 \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \delta. \tag{3}
$$

*The 5 by 3 matrix is Vandermonde's. The three normal equations $V^{\mathrm{T}}Vc = V^{\mathrm{T}}\delta$ give a direct least squares solution. The alternative (chosen by MATLAB and by Chebyshev) is to orthogonalize columns. The constant and linear columns are already orthogonal. The column $(4,1,0,1,4)$ of squares is orthogonal to the constants only after subtracting $(2,2,2,2,2)$ to leave $(2,-1,-2,-1,2)$. Thus $q_2(x) = x^2 - 2$ for $n = 2$, $N = 5$. The best $a_0 + a_1 x + a_2(x^2 - 2)$ can be computed a component at a time:*

$$(4)\quad Qa = \begin{bmatrix} 1 & -2 & 2 \\ 1 & -1 & -1 \\ 1 & 0 & -2 \\ 1 & 1 & -1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \textit{has} \quad a_i = \frac{(\text{column } i) \cdot \delta}{(\text{column } i) \cdot (\text{column } i)}.$$

*The key point is that* $Q^{\mathrm{T}}Q$ *is diagonal. Then* $a_0 = \frac{1}{5}$, $a_1 = \frac{0}{10}$, $a_2 = \frac{-2}{14}$. *The parabola closest to* $\delta$ *is* $SG(x) = a_0 + a_1 x + a_2(x^2 - 2) = -\frac{1}{7}x^2 + \frac{17}{35}$.

*Suppose we include the Vandermonde column of cubes:* $v_3 = (-8, -1, 0, 1, 8)$. *It is not orthogonal to the linear column* $q_1 = (-2, -1, 0, 1, 2)$. *The multiple of* $q_1$ *to be subtracted from* $v_3$ *is* $v_3 \cdot q_1 / q_1 \cdot q_1 = \frac{34}{10}$. *Then the orthogonalized column* $q_3 = v_3 - \frac{34}{10}q_1$ *gives the next Chebyshev polynomial* $q_3(x) = x^3 - \frac{34}{10}x$ *(with leading coefficient 1, a normalization to be changed). To conclude the example, notice especially that a multiple of* $q_3(x)/x$ *recovers the best parabola:*

$$(5)\qquad -\frac{1}{7}x^2 + \frac{17}{35} = -\frac{1}{7}\left(\frac{q_3(x)}{x}\right).$$

The best approximation to $\delta_N$ of degree $n$ is always a multiple of $(q_{n+1}(x)/x)$! This is $SG(x)$ in Equation (15) and in [5].

We confess to one small difficulty. Chebyshev (spelled Tchebitchef in earlier years) happened to choose the one-sided interval $[0, N-1]$ instead of the symmetric interval $[-M, M]$. A linear change of variables will center the problem. In other words, Chebyshev orthogonalized the column vectors $(0^j, 1^j, \ldots, (N-1)^j)$ of the usual Vandermonde matrix, to obtain the polynomial vectors given in Szegö's notation [12] by an $n$th forward difference $\Delta^n$:

$$(6)\qquad t_n(x) = n!\Delta^n \left[ \binom{x}{n} \binom{x-N}{n} \right].$$

The first three polynomials, orthogonal for sampling at $x = 0, \ldots, N-1$, are

$$\begin{aligned} t_0(x) &= 1 \\ t_1(x) &= 2x + 1 - N \\ t_2(x) &= 6x^2 + 6x + 2 - 6xN - 3N + N^2. \end{aligned}$$

For the symmetric Savitzky-Golay filters, we want to shift those $N$ sampling points to $x = -M, \ldots, M$. This shifts the polynomial to

$$(7)\qquad q_n(x) = n!\Delta^n \left[ \binom{x+M}{n} \binom{x-M-1}{n} \right].$$

The first four polynomials (whose samples give orthogonal columns) are now

$$\begin{aligned} q_0(x) &= 1 \\ q_1(x) &= 2x \\ q_2(x) &= 6x^2 - 2M(M+1) = 6x^2 - \frac{1}{2}(N^2-1) \\ q_3(x) &= 20x^3 - 4x(3M^2+3M-1) = 20x^3 - x(3N^2-7). \end{aligned}$$

The polynomial $q_j(x)$ is even or odd according as $j$ is even or odd. Our notation will be $q_j(x)$ for the polynomial, $q_j(k)$ for its samples at $k = -M, \dots, M$, and $q_j$ for the vector of those $N = 2M+1$ samples.

With these orthogonal vectors $q_j$ (discrete orthogonal polynomials) the projection of $\delta_N$ onto the discrete polynomials of degree $n$ is a sum of 1D projections:

$$SG(N,n) = \sum_0^n \frac{q_j^{\mathrm{T}} \delta_N}{q_j^{\mathrm{T}} q_j} q_j = \sum_0^N \frac{q_j(0)}{q_j^{\mathrm{T}} q_j} q_j. \tag{8}$$

The squared length of $q_j$ with $N = 2M+1$ components comes from shifting Szegö's formula to our centered interval:

$$L_j \equiv q_j^{\mathrm{T}} q_j = N(N^2-1^2)(N^2-2^2)\cdots(N^2-j^2)/(2j+1). \tag{9}$$

The same centering yields the crucial three-term recurrence that connects $q_{n-1}$, $q_n$, and $q_{n+1}$:

$$(n+1)q_{n+1}(x) - 2x(2n+1)q_n(x) + n(N^2-n^2)q_{n-1}(x) = 0. \tag{10}$$

Thus the coefficient $k_n$ of the leading term $x^n$ in $q_n(x)$ is multiplied at each step by $2(2n+1)/(n+1)$ to yield $k_{n+1}$. Starting from $k_0 = 1$ this gives

$$k_n = \binom{2n}{n}. \tag{11}$$

Now we can compute the sum in (8) that yields the Savitzky-Golay filter coefficients. As Alan Edelman pointed out, the key is the classical Christoffel-Darboux summation formula. This follows in the standard way (Szegö [12]) from the three-term recurrence:

$$\begin{aligned} &\frac{q_0(x)q_0(y)}{L_0} + \cdots + \frac{q_n(x)q_n(y)}{L_n} \\ &\qquad = \frac{n+1}{2(2n+1)} \frac{1}{L_n} \frac{q_{n+1}(x)q_n(y) - q_n(x)q_{n+1}(y)}{x-y}. \end{aligned} \tag{12}$$

For $SG(N,n)$ in (8) we take $y=0$ in (12), and recall $L_j = q_j^T q_j$ from (9). The polynomial whose samples give the vector $SG(N,n)$ of filter coefficients (the impulse response) is $SG(x)$:

$$SG(x) = \sum_{j=0}^{n} \frac{q_j(0)q_j(x)}{L_j} = \frac{n+1}{2(2n+1)} \frac{1}{L_n} \frac{q_n(0)q_{n+1}(x)}{x}. \tag{13}$$

Recall that $n$ is even, so that $q_{n+1}(x)$ is an odd polynomial and $q_{n+1}(0) = 0$. Finally we need the central value $q_n(0)$. The three-term formula for $q_n$, $q_{n-1}$, $q_{n-2}$ applied recursively gives

$$\begin{aligned} q_n(0) &= -\frac{n-1}{n}(N^2 - (n-1)^2)q_{n-2}(0) = \dots \\ &= \frac{(-1)^{n/2}}{2^n}\binom{n}{n/2}\prod_{k=1}^{n/2}(N^2 - (2k-1)^2). \end{aligned} \tag{14}$$

Inserting $q_n(0)$ into (13) and simplifying gives our explicit formula for the impulse response of the Savitzky-Golay filter $SG(N,n)$. The filter coefficients are the samples at $x = -M, \dots, M$ of the polynomial $SG(x)$:

$$SG(x) = \frac{n+1}{2^{n+1}}\binom{n}{n/2}\frac{(-1)^{n/2}}{N(N^2-2^2)(N^2-4^2)\cdots(N^2-n^2)}\frac{q_{n+1}(x)}{x}. \tag{15}$$

Since $q_{n+1}(x)$ is an odd polynomial, division by $x$ is permitted. The resulting $SG(x)$ is even and the filter is symmetric. With $n = 2$, the closest discrete parabola to the discrete delta vector $\delta_N$ is given by the $N$ samples of

$$\begin{aligned} n = 2:\ SG(x) &= \frac{3}{8} \cdot 2 \cdot \frac{(-1)}{N(N^2-2^2)} \frac{q_3(x)}{x} \\ &= \frac{3}{4} \frac{3N^2 - 20x^2 - 7}{N(N^2-4)}. \end{aligned} \tag{16}$$

Similarly the discrete quartic polynomial closest to $\delta_N$ is

$$\begin{aligned} n=4:\ SG(x) &= \frac{5}{32} \cdot 12 \cdot \frac{1}{N(N^2-2^2)(N^2-4^2)} \frac{q_5(x)}{x} \\ &= \frac{15}{64} \frac{1008\,x^4 - 280\,x^2N^2 + 1960\,x^2 + 15\,N^4 - 230\,N^2 + 407}{(N^2-16)\,(N^2-4)\,N}. \end{aligned} \tag{17}$$

These low-degree expressions are known and frequently used. They appear in *The Calculus of Observations* by Whittaker and Robinson, which anticipated Savitzky-Golay by many years. And we recognize now that Chebyshev prepared the way for everything!

**3. A Legendre-based Filter.** We propose in this section a new smoothing filter. Its construction is the "continuous analogue" of Savitzky-Golay. The polynomial $L(x)$, whose samples give the filter coefficients that form the impulse response, turns out to consist exactly of the *leading terms* *of the Savitzky-Golay polynomial* $SG(x)$. $L(x)$ depends on the degree $n$, and only in a trivial way on the filter length $N = 2M + 1$. The Legendre-based filter is simpler, and for moderate or large $N$ it is extremely close to $SG(x)$. The simplicity becomes especially valuable when the input no longer consists of uniformly spaced samples. The output from the new filter will be the natural "non-uniform" generalization of an ordinary convolution.

Before we describe the construction of $L(x)$, let us give the conclusion. In analogy with the polynomial $q_{n+1}(x)/x$ in Savitzky-Golay, the Legendre-based projection (using integrals over $[-1, 1]$ instead of discrete sums) produces an odd-degree Legendre polynomial $P_{n+1}(x)$, divided by $x$ and rescaled in (28) to stretch the interval:

$$L(x) = c_{N,n} \left( \frac{P_{n+1}(2x/N)}{x} \right). \tag{18}$$

These are the $SG$ polynomials with lower-order terms removed:

$$n = 0: \; L(x) = \frac{1}{N}$$

$$n = 2: \; L(x) = \frac{9}{4}\frac{1}{N} - 15\frac{x^2}{N^3} \qquad \text{(remove } -7 \text{ and } -4\text{)}$$

$$n = 4: \; L(x) = \frac{225}{64}\frac{1}{N} - \frac{525}{8}\frac{x^2}{N^3} + \frac{945}{4}\frac{x^4}{N^5}$$

$$n = 6: \; L(x) = \frac{1225}{256}\frac{1}{N} - \frac{11025}{64}\frac{x^2}{N^3} + \frac{24255}{16}\frac{x^4}{N^5} - \frac{15015}{4}\frac{x^6}{N^7}$$

$$n = 8: \; L(x) = \frac{99225}{16384}\frac{1}{N} - \frac{363825}{1024}\frac{x^2}{N^3} + \frac{2837835}{512}\frac{x^4}{N^5} - \frac{2027025}{64}\frac{x^6}{N^7} + \frac{3828825}{64}\frac{x^8}{N^9}.$$

Starting from $SG(x)$, the quickest approach to reach $L(x)$ would be simply to pick out the leading terms after rescaling. The Legendre polynomials are limits of Chebyshev:

$$P_n(x) = \lim_{N\to\infty} N^{-n} q_n(Mx). \tag{19}$$

This limit $N \to \infty$ turns discrete sums (scaled) into integrals. For Szegö ([12]) this is a limit relation between classical orthogonal polynomials. (He knew they had signal processing applications!) Szegö's one-line proof of (19) uses a subtle version of the mean value theorem. It may be helpful to reach the new filter by a "projection of $\delta$" construction that is completely parallel to Savitzky-Golay.

The Legendre polynomials $P_n(x)$ come from orthogonalizing $1, x, x^2, \ldots$ over the interval $[-1, 1]$:

$$\int_{-1}^{1} P_n(x) P_m(x) dx = 0 \quad \text{if } n \neq m. \tag{20}$$

The normalization is $P_n(1) = 1$. The explicit Rodrigues formula, corresponding to (7) but with $n$th derivatives in place of forward differences, is

$$P_n(x) = \frac{1}{2^n n!} \left(\frac{d}{dx}\right)^n (x^2 - 1)^n. \tag{21}$$

The length of $P_n(x)$ comes from

$$\int_{-1}^{1} (P_n(x))^2 dx = \frac{2}{2n+1}. \tag{22}$$

The three-term recurrence relation (which follows from (21)) is

$$(n+1)P_{n+1}(x) - (2n+1)xP_n(x) + nP_{n-1}(x) = 0. \tag{23}$$

The least squares filter fits the input function in each window by a polynomial of degree $n$. Then the filter output $y_{\text{cont}}(T)$ is the value of the best fit at the center point $T$ of the window. As the window shifts this creates the output function.

Suppose the window is $[-1, 1]$. When the input is a standard Dirac impulse $\delta(x)$, the output is clearly zero for $|x| > 1$. Within $[-1, 1]$ we get the impulse response by projecting $\delta(x)$ (although it is not in $L^2$) onto the polynomials of degree $n$. With Legendre's orthogonal basis $P_j(x)$, this is a sum of one-dimensional projections in analogy with (8):

$$L(x) = \sum_{0}^{n} \frac{\int P_j(x)\delta(x)dx}{\int (P_j(x))^2 dx} P_j(x) = \sum_{0}^{n} \frac{2j+1}{2} P_j(0) P_j(x) \quad \text{on } [-1, 1]. \tag{24}$$

This sum again fits the Christoffel-Darboux formula. From the three-term recurrence (23), the coefficient $K_n$ of $x^n$ in $P_n(x)$ is multiplied by $(2n+1)/(n+1)$ to give $K_{n+1}$. Then

$$K_n = \frac{1}{2^n} \binom{2n}{n}. \tag{25}$$

The Christoffel-Darboux formula with $y = 0$ applied to (24) yields

$$L(x) = \frac{n+1}{2} P_n(0) \frac{P_{n+1}(x)}{x} \quad \text{on } [-1, 1]. \tag{26}$$

For even $n$, the *central value* for the Legendre polynomial is

$$P_n(0) = \frac{(-1)^{n/2}}{2^n}\binom{n}{n/2}. \tag{27}$$

The impulse response for our Legendre-based filter on $[-M, M]$ is then

$$L(x) = (-1)^{n/2}\frac{n+1}{2^{n+1}}\binom{n}{n/2}\frac{P_{n+1}\left(2x/N\right)}{x}. \tag{28}$$

To create a discrete filter, we sample this response $L(x)$ at $N$ equally spaced points.

**4. Analysis of Filtered Gaussians.** In this section, we will analyze the output from the Savitzky-Golay filters operating on Gaussian functions. These are particularly interesting to study, because they very commonly arise in real-world problems. Our theoretical studies of the filtered signals confirm numerical experience: A careful choice of the length $N$ is essential to a good $SG$ (or Legendre-based) filter.

Assume that the undisturbed input signal $x(k)$ is a sampled Gaussian:

$$x(k) = e^{-\frac{d^2}{\beta^2}k^2}, \qquad k = \ldots, -1, 0, 1, \ldots \tag{29}$$

Here, $d$ is the spacing between the samples and $\beta$ is a measure of the width of the Gaussian. In a practical situation, the signal might be centered at a value other than zero, and have a maximum amplitude other than one. But because we are studying linear time-invariant filters, we can assume a function of the form (29) without loss of generality. Each sample $x(k)$ is disturbed by Gaussian noise $w(k)$, with zero mean and standard deviation $\sigma$ (the $w(k)$ are independent):

$$X(k) = x(k) + w(k). \tag{30}$$

We will study the result of filtering this disturbed signal using a Legendre-based filter of size $N = 2M + 1$:

$$y(k) = \sum_{j=-M}^{M} L(j)X(k-j). \tag{31}$$

A measure of the error in the filtered signal $y(k)$ is the deviation $r$ at $k = 0$:

$$r \equiv x(0) - y(0) = 1 - y(0). \tag{32}$$

This definition of the error represents an important (and convenient) output from the filtered signal, the height of the Gaussian. One could imagine a filter that gives a poor reconstruction of the original signal, but still has $y(0) = x(0)$, that is, $r = 0$. But when filtering Gaussians using least-square polynomials, the signal height gives a good indication of the accuracy over the whole signal.

The error separates into two parts, $r_{\text{noise}}$ and $r_{\text{signal}}$. The first is the error due to the noise $w(k)$, the second is the error due to the filtering of the pure Gaussian. The expected value of the square of $r_{\text{noise}}$ is

$$E\left[r_{\text{noise}}^2\right] = \sigma^2 \sum_{j=-M}^{M} L(j)^2 \approx \sigma^2 \int_{-N/2}^{N/2} L(x)^2\, dx. \tag{33}$$

This integral can be evaluated by noting the following property (not expected by us!) of the Legendre polynomials. Recall from (26) that $L(x)$ is proportional to $P_{n+1}(x)/x$:

THEOREM 4.1.

$$\int_{-1}^{1} \left(\frac{P_{n+1}(x)}{x}\right)^2 dx = 2 \quad \textit{for all even } n. \tag{34}$$

*Proof.* Divide both sides of the three-term recurrence (23) by $x$. Then square and integrate over $[-1, 1]$. Denoting $\int_{-1}^{1} (P_{n+1}(x)/x)^2\, dx$ by $c_{n+1}$, the result is

$$\begin{aligned} (n+1)^2 c_{n+1} \;\; &= \;\; (2n+1)^2 \int_{-1}^{1} P_n(x)^2\, dx \\ &\quad - 2n(2n+1) \int_{-1}^{1} P_n(x) \frac{P_{n-1}(x)}{x}\, dx + n^2 c_{n-1}. \end{aligned} \tag{35}$$

Since $P_{n-1}(x)/x$ has degree $n-2$ for even $n$, it is orthogonal to $P_n(x)$. The first integral on the right is known to equal $2/(2n+1)$. Therefore

$$(n+1)^2 c_{n+1} = 2(2n+1) + n^2 c_{n-1}. \tag{36}$$

Starting from $c_1 = \int_{-1}^{1} (x/x)^2\, dx = 2$, this gives $c_3 = 2$ and all $c_{n+1} = 2$. □

Now, let $a_n = (-1)^{n/2} \frac{n+1}{2^{n+1}} \binom{n}{n/2}$. The explicit expression for $L(x)$ then becomes $L(x) = a_n P_{n+1}\left(2x/N\right)/x$. By (33) and Theorem 4.1, the error $r_{\text{noise}}$ has variance

$$E\left[r_{\text{noise}}^2\right] \approx \sigma^2 \int_{-N/2}^{N/2} L(x)^2\, dx = \frac{4\sigma^2 a_n^2}{N}. \tag{37}$$

Turning to the error $r_{\text{signal}}$ of the filtered pure Gaussian, we can directly write the error at the center as

$$\begin{aligned} r_{\text{signal}} &= \sum_{j=-M}^{M} L(j) x(k-j) - 1 \approx \int_{-N/2}^{N/2} L(s) x(s)\, ds - 1 \\ &= \int_{-1}^{1} a_n \frac{P_{n+1}(s)}{s} x(Ns/2)\, ds - 1. \end{aligned} \tag{38}$$

The expected value of the squared error $r^2$ after filtering $X(k)$ is now (by independence of the noise)

$$
\begin{aligned}
E\left[r^2\right] &= E\left[(r_{\text{noise}} + r_{\text{signal}})^2\right] = E\left[r^2_{\text{noise}}\right] + r^2_{\text{signal}} \\
&= \frac{4\sigma^2 a_n^2}{N} + \left(1 - \int_{-1}^{1} a_n \frac{P_{n+1}(s)}{s} x(Ns/2)\, ds\right)^2 .
\end{aligned} \tag{39}
$$

Our goal is to choose a filter size $N_{\text{opt}}$ that minimizes the error. The derivative of $E[r^2]$ is

$$
\begin{aligned}
\frac{dE\left[r^2\right]}{dN} = &- \frac{4\sigma^2 a_n^2}{N^2} - \left(1 - \int_{-1}^{1} a_n \frac{P_{n+1}(s)}{s} x(Ns/2)\, ds\right) \\
&\times \left(\int_{-1}^{1} a_n P_{n+1}(s) x'(Ns/2)\, ds\right).
\end{aligned} \tag{40}
$$

Set this equal to zero, insert the expression (29) for the Gaussian $x(k)$ and its derivative $x'(k)$, and simplify:

$$
\begin{aligned}
\frac{4\sigma^2\beta^2}{d^2} = &N^3 \left(\int_{-1}^{1} s P_{n+1}(s) e^{-\left(\frac{Nd}{2\beta}\right)^2 s^2}\, ds\right) \\
&\times \left(1 - \int_{-1}^{1} a_n \frac{P_{n+1}(s)}{s} e^{-\left(\frac{Nd}{2\beta}\right)^2 s^2}\, ds\right).
\end{aligned} \tag{41}
$$

Solving this equation for $N = N_{\text{opt}}$ determines the optimal filter length for the given polynomial degree $n$, noise level $\sigma$, step length $d$, and variance $\beta^2$. Using the odd integer closest to $N_{\text{opt}}$ as the filter length should give good results. In the next section, we show numerical evidence of the correctness of (41), as well as experiments showing how important it is to choose a good $N$.

**5. Numerical Results.** First, we investigate how well the Legendre-based filter $L(x)$ in (28) approximates the discrete Savitzky-Golay filter $SG(x)$ in (13). Figure 1 shows the maximum absolute error in the filter coefficients for five different polynomial orders, as a function of the filter size $N$. Recall that $\sum_k SG(k) = 1$ and $\sum_k L(k) \approx \int_x L(x) = 1$, so there is no need to scale the errors by the magnitude of the filter coefficients. The slopes of the curves are close to $-3$ (slightly above), indicating that $|SG(k) - L(k)| = \mathcal{O}(N^{-3})$.

The difference is very small. We claim that the Legendre-based filter can be used in all cases, except perhaps for the smallest filter sizes. In fact we have no evidence that the discrete filter gives a better result! The continuous one could be just as good.

Next, we show that our expression for the error of a filtered Gaussian (39) agrees with numerical experiments. In Figure 2, the predicted errors using a number of filters are shown together with the actual errors from

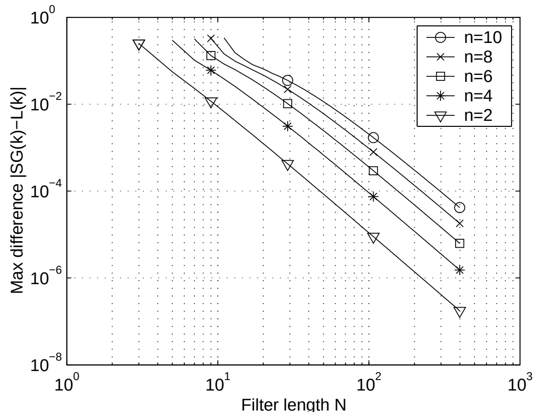


FIG. 1. *The difference between the Savitzky-Golay $SG(k)$ and the Legendre-based $L(k)$.*

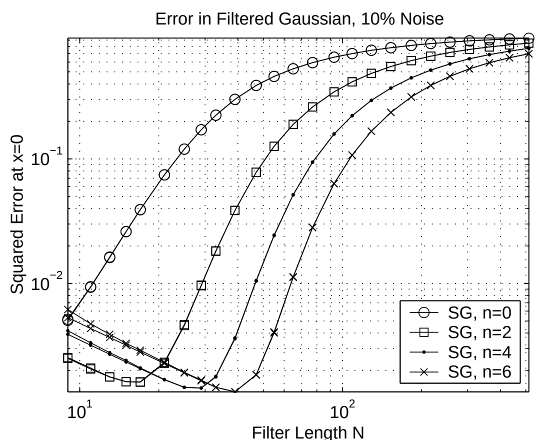


FIG. 2. *The error in a filtered Gaussian with width $\beta = 10$, $d = 1$, and white noise with standard deviation $\sigma = 0.1$, computed using the analytical expression (39) and using numerical simulation.*

a simulation using Savitzky-Golay. We are measuring an average error, so the actual errors have been produced by averaging a large number of simulations (100,000). The curves are amazingly close, even for small $N$. This is a little surprising since (39) is derived using Legendre-based filters, while the numerical simulations use Savitzky-Golay.

Figure 2 also answers another question. *How important is the filter length N?* For the fourth-order filter, the optimal filter size is $N = 25$, and this gives an error about $4 \times 10^{-4}$. For a filter twice as long, $N = 50$, the error increases above $10^{-2}$. Another factor of two in length, $N = 100$, gives

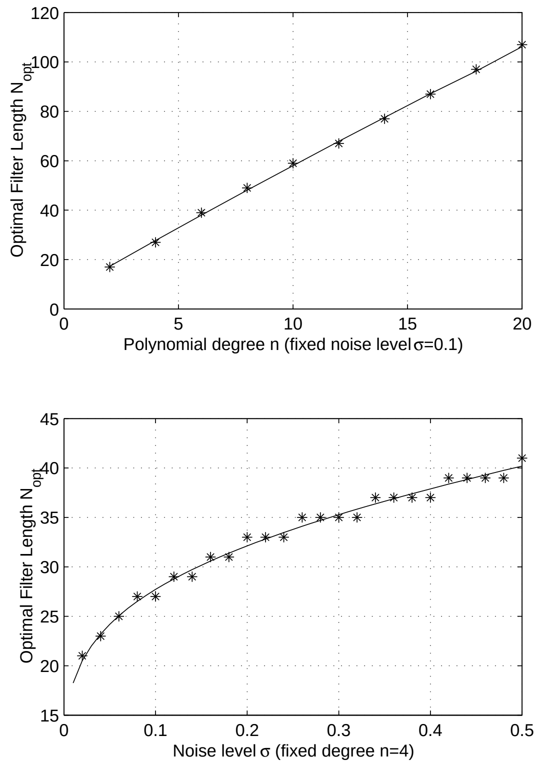


FIG. 3. *The optimal filter length $N_{\text{opt}}$ as a function of the polynomial degree $n$ (top) and the noise level $\sigma$ (bottom). The stars indicate the closest realizable filter, that is, the closest odd integer. The three lower points in the top plot and the $\sigma = 0.1$ point in the bottom plot can be verified with Figure 2.*

error $\approx 10^{-1}$. Furthermore, that optimal filter size changes significantly for different polynomial orders. It also depends on the noise level $\sigma$ (which can not be seen from this plot, because $\sigma = 0.1$ in all the curves).

This sensitivity to $N$ makes Equation (41) very important. Having confirmed our analytical expression (39) for the error, we can use (41) to compute an optimal $N$. In Figure 3, this has been done with a number of different parameters. In the top plot, $N_{\text{opt}}$ is shown as a function of the polynomial order $n$ for constant noise level $\sigma = 0.1$. We again use a Gaussian with $\beta = 10$ and $d = 1$. It is interesting that the optimal filter size increases almost linearly with $n$ in this range, something that is not obvious from the (rather complex) equation (41).

In the bottom plot, $N_{\text{opt}}$ is shown as a function of the noise $\sigma$, for fixed degree $n = 4$. Longer filters are better for higher noise levels. Our forthcoming paper [9] will extend this analysis, and provide a fast implementation for these simple, quick, but non-optimal filters.

## REFERENCES

[1] M.U.A. Bromba and H. Ziegler. Efficient computation of polynomial smoothing digital filters. *Analytical Chemistry*, **51**(11): 1760–1762, September 1979.

[2] M.U.A. Bromba and H. Ziegler. Application hints for Savitzky-Golay digital smoothing filters. *Analytical Chemistry*, **53**(11): 1583–1586, September 1981.

[3] M.U.A. Bromba and H. Ziegler. Digital filter for computationally efficient smoothing of noisy spectra. *Analytical Chemistry*, **55**(8): 1299–1302, July 1983.

[4] M.U.A. Bromba and H. Ziegler. Digital smoothing of noisy spectra. *Analytical Chemistry*, **55**(4): 648–653, April 1983.

[5] M.U.A. Bromba and H. Ziegler. On Hilbert space design of least-weighted-squares digital filters. *Int. J. of Circuit Theory and Applications*, **11**: 7–32, 1983.

[6] S.P. Davis, M.C. Abrams, and J.W. Brault. *Fourier Transform Spectrometry.* Academic Press, 2001.

[7] W. Gander and J. Hrebicek. *Solving Problems in Scientific Computing Using Maple and MATLAB.* Springer, 2002.

[8] A.F. Nikiforov, S.K. Suslov, and V.B. Uvarov. *Classical Orthogonal Polynomials of a Discrete Variable.* Springer, 1991.

[9] P. Persson and G. Strang. Smoothing by fast polynomial filters. In preparation.

[10] W. Press, S. Teukolsky, W. Vetterling, and B. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, 1993.

[11] L.B. Scott and L.R. Scott. Efficient methods for data smoothing. *SIAM J. Numer. Anal.*, **26**(3): 681–692, June 1989.

[12] G. Szegö. *Orthogonal Polynomials.* American Math. Society, 1939.