

Discrete Polynomial Moments and Savitzky-Golay Smoothing

Paul O'Leary, Matthew Harker

Abstract—This paper presents unified theory for local (Savitzky-Golay) and global polynomial smoothing. The algebraic framework can represent any polynomial approximation and is seamless from low degree local, to high degree global approximations. The representation of the smoothing operator as a projection onto orthonormal basis functions enables the computation of: the covariance matrix for noise propagation through the filter; the noise gain and; the frequency response of the polynomial filters. A virtually perfect Gram polynomial basis is synthesized, whereby polynomials of degree $d = 1000$ can be synthesized without significant errors. The perfect basis ensures that the filters are strictly polynomial preserving. Given n points and a support length $l_s = 2m + 1$ then the smoothing operator is strictly linear phase for the points $x_i, i = m+1 \dots n-m$. The method is demonstrated on geometric surfaces data lying on an invariant 2D lattice.

Keywords—Gram polynomials, Savitzky-Golay Smoothing, Discrete Polynomial Moments

I. INTRODUCTION

THE concept of moment invariants for pattern recognition was introduced in 1962 by Hu [1]. In his computations Hu used a geometric polynomial basis set $f(x) = \sum_{i=1}^n c_n x^n$ to determine the polynomial moments. The geometric polynomials — called the Vandermonde basis¹ here — have the following serious disadvantages: a) the Vandermonde matrix is poorly conditioned and quickly becomes degenerate as the degree of the polynomial increases. There is no unique inverse to a degenerate matrix and consequently these basis functions are not suited for the synthesis of a signal. The maximum degree of polynomial which can be used is approximately $d_{v,max} = 10$; b) the Vandermonde basis is not orthogonal in the discrete space, and consequently the basis is not polynomial preserving. This feature is important when doing polynomial filtering such as Savitzky-Golay smoothing [2].

In the years ensuing Hu's publication there was significant development in the understanding of Polynomial moments and their relationship to least square approximation [3]. The main step forward in polynomial moments was associated with the introduction of new polynomial bases [4]–[9], this enabled the computation of moments of higher degree on larger images; for example Zhu [10] was able to compute global moments for images with a size of 256×256 . The of generating virtually perfect bases sets of arbitrary size can be associated with Mukundan [11], his basis was limited to

The authors are at Institute for Automation, Universality of Leoben, A-8700 Leoben, Austria, email: paul.oleary@unileoben.ac.at

¹The name Vandermonde basis is used here, since a matrix B_v containing the geometric basis functions as columns is a Vandermonde matrix. The use of this name is advantageous since it leads to the theory and difficulties in inverting the matrix B_v .

regular grids and the errors — be it very small errors — are concentrated at lower degree polynomials. Then a new synthesis procedure was introduced for the generation of unitary discrete Gram polynomials on arbitrary nodes [12] and with a computation accuracy comparable with the Fourier basis. The availability of a unitary basis also enabled the development of understanding with respect to spectral propagation of gaussian noise and noise bandwidth of polynomial filters [13]. Discrete polynomial moments are global approximations of data where polynomials of high degree are used to model the details.

In 1964 Savitzky and Golay [2] introduced local polynomial approximation to smooth and evaluate the derivatives noisy spectrometer data. The smoothing was based on fitting a geometric polynomial (Vandermonde basis set) of low degree d to a data set of limited length, the original paper used a maximum degree of $d_{max} = 6$ and the data length of $l_s = 25$ also known as the support length or in filtering literature as the bandwidth — a clear misnomer. Modden [14] pointed out some errors in the original paper and extended the method to larger support lengths. The method was extended to image processing by a number of groups, e.g. by Rajagopalan [15]. All these applications are limited by the properties of the Vandermonde basis and limits there applicability to low degree approximations.

In 1990 two groups [16], [17], apparently independently, proposed the use of Gram polynomials [18] in place of the Vandermonde basis for Savitzky-Golay smoothing. Meer and Weiss [16] state that the Chebyshev and Gram polynomials are synonymous for the same set of basis functions; this is only partially correct. The modified discrete Chebyshev polynomials [19], [20] do not have a uniform scaling, where as the Gram polynomials do. Furthermore, the polynomials published by Meer and Weiss in their paper are not orthogonal. Nevertheless, the use of Gram polynomials is an important step in the improvement of the performance of orthogonal moments. Theoretically the Gram polynomials should form an ideal basis; however, their synthesis via the recurrence relationship introduces serious errors which limit the degree of polynomial which can be used: The recurrence relationship for the Gram polynomials is,

$$g_n(x) = 2 \alpha_{n-1} x g_{n-1}(x) - \frac{\alpha_{n-1}}{\alpha_{n-2}} g_{n-2}(x), \quad (1)$$

whereby,

$$\alpha_{n-1} = \frac{m}{n} \left(\frac{n^2 - 1/2}{m^2 - n^2} \right)^{1/2} \quad (2)$$

and

$$g_0(x) = 1, \quad g_{-1}(x) = 0 \quad \text{and} \quad \alpha_{-1} = 1, \quad (3)$$

x is computed on equidistant points,

$$x = -1 + \frac{(2k-1)}{m}, \quad 1 \leq k \leq m, \quad (4)$$

note these points do not span the full range $[-1, 1]$. The bases functions are scaled by \sqrt{m} yielding a unitary bases set. Now consider the matrix $G_n = [g_0(x), \dots, g_n(x)]$, whereby the i^{th} column correspond to the i^{th} basis function $g_i(x)$. If G contains an ideal unitary bases set then, $P = G^T G = I$, and consequently, the projection error $P_e = P - I = 0$ should yield a matrix which is uniformly zero. To demonstrate the limits associated with the generation of the Gram recurrence relationship a basis was computed for $m = 100$ and $n = 40$ the associated projection error is shown in Figure 1. The errors are in the range $\epsilon = 10^{-13}$ these are small but already significantly larger than the computational accuracy of MATLAB, indicating that the errors associated with the recurrence relationship are now dominant. The highest feasible degree for a Gram polynomial generated in this manner was $n = 20$ when computed for $m = 100$ points; at higher degrees the error exceeded the numerical limit in MATLAB of $eps = 10^{-16}$. With this, the Savitzky-Golay method remains a local polynomial approximation with limited degree, unable to reach the dimensions associated with global approximations of images.

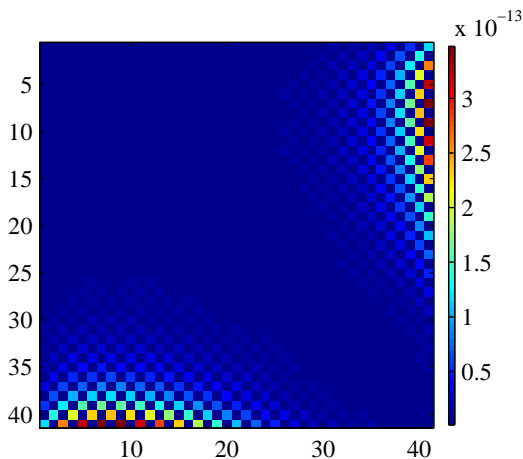


Fig. 1. The projection error P_e associated with the Gram basis set for $m = 100$ and $n = 40$. The errors are in the range $\epsilon = 10^{-13}$ these are small but already significantly larger than the computational accuracy available, indicating that the errors associated with the recurrence relationship are now dominant.

Mukundan [11] published a synthesis algorithm which produces a high quality basis with almost arbitrary degree. However, the errors are concentrated at lower degrees and the algorithm is limited to regular nodes. In 2008 O'Leary and Harker [12] published an algorithm which enabled the synthesis of a virtually perfect discrete orthogonal basis on arbitrary nodes – the nodes are truly arbitrary, the algorithm also functions for complex nodes. This algorithm is used in this paper for the generation of all orthogonal bases.

This paper presents a unified framework for local and global polynomial approximation, with or without weighting. The main contribution of this paper are:

- 1) a unified algebraic framework is presented for the synthesis and analysis of polynomial approximations;
- 2) the framework delivers an exact method of computing the frequency response of polynomial filters. It also shows the tendency of a polynomial approximation to exhibit oscillations at the border of their support;
- 3) the synthesis algorithm proposed by O'Leary and Harker [12] is used to generate a virtually perfect Gram polynomial basis. This makes polynomial approximation of arbitrary degree possible;
- 4) the frequency response of the polynomial approximation is combined with regularization theory to enable the computation of the noise gain associated with the filters;
- 5) the covariance propagation for gaussian noise is computed;
- 6) the polynomial approximations can be computed for irregular nodes and for data with missing samples, if required;
- 7) it is proved that the polynomial approximations presented are strictly polynomial preserving.

The theory is first presented for one dimensional data then extended to two dimensional data as encountered in image processing.

II. THEORETICAL FRAMEWORK

Past analysis of polynomial moments has focused on their application in pattern recognition and not on filtering [1], [5], [6], [8], [9], [21]–[23]. Polynomial moments have been proposed for filtering [24]; however, no formal analysis has been performed for such applications. Whereas, Savitzky-Golay smoothing focused on filtering but paid no attention to the global characteristics of the polynomial approximations. The aim in this section is to develop a theoretical framework which unifies both aspects. At this point no assumptions are made with respect to the basis function set being used².

The most general formulation of a discrete equivalent of an integral-transform is,

$$s = B_a^+ y. \quad (5)$$

The spectrum s of the data y is determined by computing the discrete equivalent of an integral transform. The formulation with the Moore-Penrose pseudo inverse B_a^+ has been chosen since it is the least squares solution for non-orthogonal bases. Discrete polynomial moments belong to a class of basis functions whose spectrum can be calculated as above. The synthesis of a signal, i.e. the inverse transform is computed as,

$$\hat{y} = B_s s \quad (6)$$

whereby, B_s contains the synthesis functions.

A. Completeness and invertibility

Equations 5 and 6 can be combined to yield,

$$\hat{y} = B_s s = B_s B_a^+ y. \quad (7)$$

²Some examples of suitable basis functions are: the Fourier basis; discrete orthogonal polynomials; the Bessel functions and the Haar functions. The most appropriate basis functions depend on the nature of the application at hand.

In general the basis function B_a and B_s are related via permutation matrices, i.e. they are fundamentally the same type of basis functions, but possibly evaluated at different nodes or of different degrees. The reconstruction error $\mathbf{r} = \mathbf{y} - \hat{\mathbf{y}}$ is,

$$\begin{aligned} \mathbf{r} &= \mathbf{y} - B_s B_a^+ \mathbf{y} \\ &= (I - B_s B_a^+) \mathbf{y} \end{aligned} \quad (8)$$

Consequently, perfect reconstruction is given if the projection onto the orthogonal complement $I - B_s B_a^+ = 0$. This can be achieved using a unitary³ and complete⁴ basis for analysis and reconstruction; i.e. $B = B_a = B_s$ such that,

$$\mathbf{r} = (I - B B^T) \mathbf{y} = 0. \quad (9)$$

In general, the complex conjugate transpose⁵ of any complete unitary basis is its inverse.

B. A formalism for filtering

Spectral filtering can be modelled in three steps:

- 1) computing the signal polynomial spectrum,

$$\mathbf{s}^\delta = B^T \mathbf{y}^\delta; \quad (10)$$

- 2) filtering the spectrum with the filter F

$$\mathbf{s}_F^\delta = F B^T \mathbf{y}^\delta; \quad (11)$$

- 3) synthesizing the filtered signal, \mathbf{y}_F^δ

$$\mathbf{y}_F^\delta = B \mathbf{s}_F^\delta = B F B^T \mathbf{y}^\delta. \quad (12)$$

If the filter function is factorable we can define $F \triangleq G G^T$ and $D \triangleq B G$. Consequently,

$$\mathbf{y}_F^\delta = B G G^T B^T \mathbf{y}^\delta = D D^T \mathbf{y}^\delta. \quad (13)$$

Note: $P \triangleq D D^T$ is an orthogonal projection onto the filtering basis functions. Given a set of n points the projection P has the dimension $n \times n$, i.e. each filtered point $y_{F,i}^\delta$ is a linear combination of all input values. Consequently, the rows of the projection matrix can be regarded as the coefficients of an FIR filter; this enables the direct computation of the frequency response [25]. Knowing that the gaussian input noise is spread evenly over the complete spectrum yields the noise gain $g_n = |F|_2^2$. This is a very important new result, enabling for the first time analytic computation on the noise behaviour of polynomial filters. In polynomial preserving filters an unmodified subset of the complete basis functions are used. In this case an orthogonal but incomplete basis set is used. In this manner both low pass and band pass filters can be implemented.

³Unitary implies $B^T B = I$, i.e. an orthogonal matrix.

⁴A unitary and complete matrix fulfills the condition $B^T B = B B^T = I$.

⁵The transpose $(.)^T$ always refers to the complex conjugate transpose in this paper.

C. Linear transformations and covariance propagation

In general measurements are perturbed by noise, i.e. the observations \mathbf{y}^δ are noisy,

$$\mathbf{y}^\delta = \mathbf{y} + \delta \mathbf{u}_n, \quad (14)$$

consisting of the ideal signal \mathbf{y} with additive white noise $\delta \mathbf{u}_n$, where: \mathbf{u}_n is a vector of gaussian noise with zero mean and covariance matrix $\Lambda_{\mathbf{y}}$.

There are two cases of interest with respect to the application of discrete orthogonal basis functions and the propagation of covariance:

- 1) propagation from the spatial (temporal) to the spectral domain during the computation $\mathbf{s} = B^T \mathbf{y}^\delta$;
- 2) propagation from input to output of a filter, $\mathbf{y}_f = P \mathbf{y}^\delta$.

Both these cases involve the covariance propagation through linear transformations. Given a general linear transformation represented as a matrix L , and the input covariance matrix $\Lambda_{\mathbf{y}}$, the covariance propagation is given by,

$$\Lambda_f = L \Lambda_{\mathbf{y}} L^T. \quad (15)$$

If the signal \mathbf{y} is perturbed by independent and identically distributed Gaussian noise, then the covariance of \mathbf{y} is,

$$\begin{aligned} \Lambda_{\mathbf{y}} &= \text{diag} \{ \sigma^2, \dots, \sigma^2 \} \\ &= \sigma^2 I_m \end{aligned} \quad (16)$$

where I_m is an $m \times m$ identity matrix. Consequently,

$$\Lambda_f = \sigma^2 L L^T. \quad (17)$$

In the special case of computing the spectrum with a unitary and complete set of basis functions, $L = B$ and $L L^T = I$, consequently the covariance of the spectrum is,

$$\Lambda_s = \sigma^2 I_m. \quad (18)$$

This proof states that gaussian noise has a flat power spectral density for all unitary and complete basis function sets, independent of their nature, i.e., Parseval's theorem is true for all unitary and complete bases. This fact is well known from regularization theory for unitary matrices [26] and has been applied in the past to the Fourier and Cosine Bases. The new algebraic approach to analyzing polynomial basis functions proposed here enables the extension of this theory to polynomials and generalized finite impulse response (FIR) filters.

III. LOCAL AND GLOBAL POLYNOMIAL SMOOTHING

Consider performing smoothing on a set of n data points, a support length of l_s and of degree d : note $l_s = n$ is a special case and corresponds to global polynomial approximation. To support understanding the example with $n = 10$, $l_s = 5$ and $d = 3$ are presented graphically in figures 2 and 3. The algorithm to generate the complete linear transformation P_C for all n points is:

Step 1: compute a unitary basis B for l_s and d using the procedure proposed by O'Leary and Harker?? [12]. This is a unitary but incomplete basis, since $d < l_s - 1$;

Step 2: compute the local projection matrix $P = B B^T$, this is an $l_s \times l_s$ matrix, see Figure 2 for the example $l_s = 5$ and $d = 3$. The center row of P corresponds to computing the projection at the center of the support $p_{x=0}$. These coefficients are symmetric implying that the frequency response is strictly linear phase⁶. The rows above and below $p_{x=0}$ correspond to the projections onto the basis functions at the start and end of the data respectively, i.e. where the end of the support is being approached. The coefficients in this region are asymmetric, the corresponding non-linear phase is responsible for the fact that polynomial approximations tend to oscillate the end of their support.

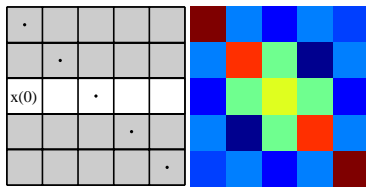


Fig. 2. Left: structure of the projection matrix. Right: example projection for a basis with support length $l_s = 5$ and of degree $d = 3$.

Step 3: generate the global complete matrix P_C . By placing the top and bottom of the projection matrix P at the start and end of P_C . The core of the matrix, $n_c = n - l_s + 1$ points, is filled diagonally with $p_{x=0}$, see Figure 3 for an example. Consequently, the core of the smoothing is strictly linear phase and produces no oscillatory behavior, unless of course there is a source of significant Gibbs error [20]. The matrix P_C is a linear transformation, but not a projection matrix; note it is not symmetric. This algorithm enables the generation of the

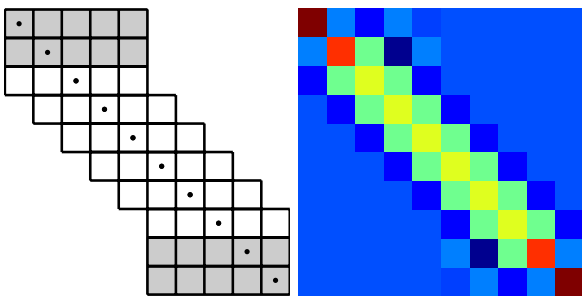


Fig. 3. Left: structure of the linear transformation for Savitzky Golay Smoothing. Right: example projection for support length $l_s = 5$, of degree $d = 3$, and $n = 10$ points.

linear transformation matrix required for the complete region from local to global polynomial filtering. To demonstrate this, the transformation for $l_s = 5$, of degree $d = 3$, and $n = 10$ points is shown in Figure 4, note the peaking at the end of the support. The corresponding frequency response is shown in Figure 5, it exhibits strong resonant peaking at the borders of the support. The core frequency response is shown in Figure 6.

⁶All FIR filters with symmetric coefficients are strictly linear phase [25].

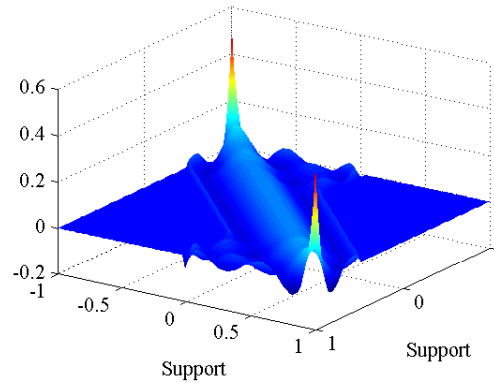


Fig. 4. Example of the transformation matrix with the parameters $l_s = 51$, of degree $d = 5$, and $n = 100$ points.

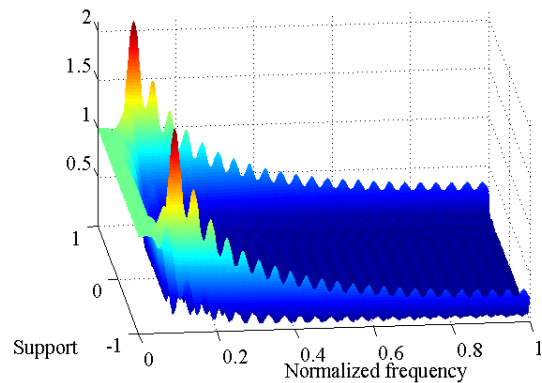


Fig. 5. Frequency response of the transformation from Figure 4.

IV. EXTENSION TO 2D LATTICES

Consider a set of 2D data Z lying on a rectangular separable lattice, this is generally the case when working with images or measurement data scanned on a regular interval. The filtered data \hat{Z} is computed as:

$$\hat{Z} = P_y Z P_x^T, \quad (19)$$

where, P_x and P_y are the transformation matrices applied in the x and y directions. Different transformations may be used in each direction as required. These then form anisotropic polynomial approximations. An example geometric surface data is shown in Figure 7. The local anomalies Z_A are extracted via local polynomial approximation,

$$Z_A = Z_G - P_y Z_G P_x^T \quad (20)$$

where, $l_{s,x} = 61$, $d_x = 3$, $n_x = 1215$ and $l_{s,y} = 31$, $d_y = 3$, $n_y = 640$. Anisotropic parameters have been chosen for this demonstration.

V. CONCLUSIONS

A new theoretical framework has been presented for polynomial approximations. A virtually perfect Gram polynomial basis set is synthesized: this ensures that the approximations

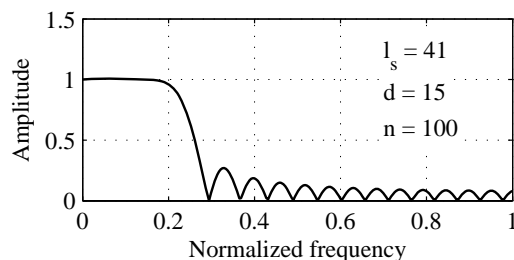


Fig. 6. Core frequency response of the transformation from Figure 4.

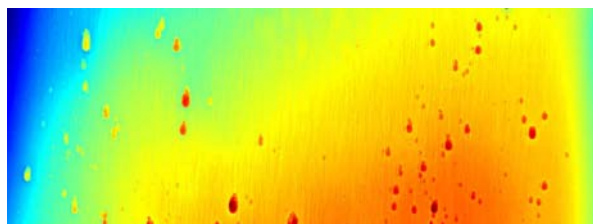


Fig. 7. Surface geometry Z_G scanned on a 640×1215 data grid.

are strictly polynomial preserving; and polynomials of degree $d = 1000$ can be generated without significant errors. Consequently, it is possible to have a seamless transition from Savitzky-Golay smoothing (low degree local approximation) to high degree global approximations. The unitary Gram basis also enables the computation of covariance propagation and noise gain for the polynomial filters.

The representation of polynomial smoothing as an orthogonal projection onto a basis function set, delivers directly a method of computing the frequency response of the filter. Furthermore, the projection matrix and the corresponding frequency response show the tendency of polynomial approximations to oscillate at the borders of the support. Given n points and a support length $l_s = 2m + 1$ then the smoothing operator is strictly linear phase for the points $x_i, i = m + 1 \dots n - m$.

REFERENCES

[1] M.-K. Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, pp. 179–187, 1962.
 [2] A. Savitzky and M. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36 (8), p. 1627..1639, 1964.
 [3] M. Eden, M. Unser, and R. Leonardi, "Polynomial representation of pictures," *Signal Processing*, vol. 10, pp. 385–393, 1986.
 [4] G. Yang, H. Shu, G. C. Han, and L. Luo, "Efficient Legendre moment computation for grey level images," *Pattern Recognition*, vol. 39, pp. 74–80, 2006.
 [5] P.-T. Yap and P. Raveendren, "Image analysis by Krawtchouk moments," *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1367–1377, 2003.
 [6] —, "An efficient method for the computation of Legendre moments," *IEEE Transactions on Pattern Analysis and Maschine Intelligence*, vol. 27, no. 12, pp. 1996–2002, 2005.
 [7] K. Hosny, "Exact Legendre moment computation for gray level images," *Pattern Recognition*, vol. doi:10.1016/j.patcog.2007.04.014, 2007.
 [8] R. Mukundan, S. Ong, and P. Lee, "Image analysis by Tchebichef moments," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1357–1363, 2001.
 [9] H. Zhu, H. Shu, J. Liang, L. Luo, and J.-L. Coatrieux, "Image analysis by discrete orthogonal Racah moments," *Signal Processing*, vol. 87, pp. 687–708, 2007.

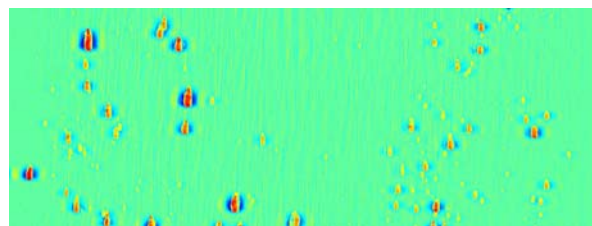


Fig. 8. Local anomalies extracted from the surface geometry data via local polynomial approximation.

[10] H. Zhu, H. Shu, J. Zhou, L. Luo, and J.-L. Coatrieux, "Image analysis by discrete orthogonal Racah moments," *Pattern Recognition Letters*, vol. doi:10.1016/j.patrec.2007.04.013, 2007.
 [11] R. Mukundan, "Some computational aspects of discrete orthogonal moments," *IEEE Transactions on Image Processing*, vol. 13, no. 8, pp. 1055–1059, 2004.
 [12] P. O'Leary and M. Harker, "An algebraic framework for discrete basis functions in computer vision," in *IEEE Indian Conference on Computer Vision, Graphics and Image Processing*, Bhubaneswar, Dec, 2008.
 [13] P. O'Leary, B. Mörtl, and M. Harker, "Discrete polynomial moments and the extraction of 3d embossed digits for recognition," *Submitted to Journal of Electronic Imaging*, 2009.
 [14] H. Madden, "Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data," *Analytical Chemistry*, vol. 50 (9), p. 13831386, 1978.
 [15] S. Rajagopalan and R. Robb, "Image smoothing with Savitzky-Golay filters," in *Medical Imaging 2003: Visualization, Image-Guided Procedures, and Display*, vol. Vol. 5029, May 2003, 2003, p. 773..781.
 [16] P. Meer and I. Weiss, "Smoothed differentiation filters for images," in *IEEE 110th International Conference on Pattern Recognition*, vol. 2, June 1990, Atlantic City, NJ, USA, 1990, p. 121..126.
 [17] P. Gorry, "General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method," *Analytical Chemistry*, vol. 62, p. 570..573, 1990.
 [18] R. Barnard, G. Dahlquist, K. Pearce, L. Reichel, and K. Richards, "Gram polynomials and the kummer function," *Journal of Approximation Theory*, vol. 94, no. 1, pp. 128 – 143, 1998. [Online]. Available: <http://www.sciencedirect.com/science/article/B6WH7-45K18JY-HI2/c92d2ff697d7631c99abe5d80f281d24>
 [19] J. Boyd, *Chebyshev and Fourier Spectral Methods*. Mineola, New York: Dover Publications Inc., 2001.
 [20] A. Jerri, *The Gibbs Phenomenon in Fourier Analysis, Splines and Wavelet Approximations*. Dordrecht, Netherlands: Kluwer Academic Publishers, 1998.
 [21] S. Ong and P. Raveendren, "Image feature analysis by Hahn orthogonal moments," *Lecture Notes in Computer Science*, vol. 3656, pp. 524–531, 2005.
 [22] Z. Ping, H. Ren, J. Zou, Y. Sheng, and W. Bo, "Generic orthogonal moments: Jackobi-Fourier moments for invariant image description," *Pattern Recognition*, vol. 40, pp. 1245–1254, 2005.
 [23] B. Bayraktar, T. Bernas, P. Robinson, and B. Rajwa, "A numerical recipe for accurate image reconstruction from discrete orthogonal moments," *Pattern Recognition*, vol. 40, pp. 659–669, 2007.
 [24] J. Thurston and J. Brawn, "The filtering characteristics of least-squares polynomial approximation for regional/residual separation," *Canadian Journal of Exploration Physics*, vol. 28(2), pp. 71–80, 1992.
 [25] A. V. Oppenheim and R. W. Schaffer, *Discrete-Time Signal Processing*. Englewood Cliffs: Prentice Hall, 1989.
 [26] H. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*, 1st ed. Dordrecht: Kluwer Academic Publishers, 2000.