Local Signal Reconstruction via Chromatic Differentiation Filter Banks

John Byrnes
Kromos Communications, Inc., Fremont, CA 94539
john@kromos.com

Abstract

We investigate the theory of chromatic differentiation introduced by Ignjatovic [1]. The theory has a filter bank representation in which the analysis filters consist of high order chromatic differentiators, which are orthogonal linear combinations of standard differentiators allowing for practical FIR implementation. One distinctive aspect of chromatic differentiation is its local signal approximation: each reconstructed point depends only on the values of the chromatic derivatives at the nearest sample point. When all infinitely many chromatic derivatives are available at any point, the entire signal can be perfectly reconstructed.

In this paper, we first consider the expected error of local reconstruction when only finitely many chromatic derivatives are available, and its impact on the design of the chromatic differention scheme. We next consider the problem of obtaining chromatic derivatives via FIR filtering and use the expected error term to justify a particular scheme proposed in [1]. Finally we consider the more general problem of obtaining higher order chromatic derivatives from available lower order chromatic derivatives. We conclude with an analysis of the ability of this technique to reduce the error of local signal reconstruction.

1 Introduction

The theory of chromatic differentiation as an alternative method of signal representation is introduced in [1]. Chromatic derivatives (CD's) of a signal are obtained by approximating the spectrum of the signal by complex-valued functions $W(\omega)P_k(\omega)$, where W is an arbitrary power spectral density and the P_k form a family of polynomials which are orthonormal with respect to W. A chromatic differentiator is an operator on signals $f(t)$ of the form $K_k[f](t) = a_0 f(t) + a_1 f'(t) + \cdots + a_k f^{(k)}(t)$, where $f^{(k)}(t)$ is the k^{th} ordinary derivative of $f(t)$, K_k has as its transfer function $P_k(\omega)$, and the a_m are complex constants. Let $\mathcal{F}[f](\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt$ denote the Fourier transform and $\mathcal{F}^{-1}[F](t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\omega t}\,d\omega$ its inverse. Defining the time-domain *basis functions* $B_k(t) = (-1)^n \mathcal{F}^{-1}[W(\omega)P_k(\omega)](t)$, most of the properties of chromatic differentiation can be derived from the following decomposition of the Fourier kernel: $e^{j\omega t} = \sum_{k=0}^{\infty} P_k(\omega)B_k(t)$. In particular, given a signal $f(t)$, we have for any fixed t_0:

$$f(t) = \sum_{k=0}^{\infty} K_k[f](t_0)B_k(t - t_0) \qquad (1)$$

A *local* approximation of a signal f is obtained by truncating 1 to only the first $M+1$ many CD's. A sharp bound on the resulting approximation error is given in [1] as a function of the total windowed energy of the signal f (*i.e.*, the energy of $W(\omega)F(\omega)$). As this error is small for a reasonably large time interval around t_0, we can approximate the entire signal f by taking local approximations at regular intervals.

As an example, let $W(\omega) = \frac{4}{\pi}\sqrt{1-(\omega/\pi)^2}$ for $|\omega| \leq \pi$ and be 0 otherwise. $P_k(\omega) = j^k U_k(\omega/\pi)$ for $k > 0$ where $U_k(x)$ is the k^{th} Chebyshev polynomial of the second kind (and $j = \sqrt{-1}$). $B_k(t) = J_k(\pi t) + J_{k+2}(\pi t)$ where $J_k(x)$ is the k^{th} Bessel function of the first kind. Figure 1 shows a few basis functions B_k. Also shown is the piecewise reconstruction of part of a signal, in which the approximations are intentionally placed far enough from each other in order to emphasize the locality of the approximation. Figure 2 graphs the average power of the approximation error as a function of distance from the center of approximation.

Thus, as in generalized sampling, or multirate filter bank theory, (see for example [2]), we sample multiple filtered versions of a signal at a sub-Nyquist rate. A unique characteristic of the CD representation of a signal, however, is that reconstruction consists of summing several *distinct* basis functions at a single sample point, each for a short time interval. The time intervals of expansion containing distinct sample points are non-overlapping. This is in contrast to standard sampling schemes like Nyquist-rate sampling, for which reconstruction consists of summing

0-7803-7147-X/01/$10.00©2001 IEEE

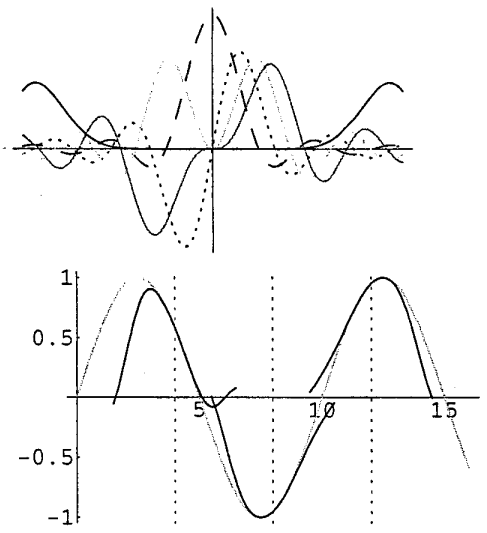

Figure 1: Top: time-domain basis functions $B_0(t)$ (dashed),$B_1(t)$ (dotted), $B_2(t)$ (lightest), $B_3(t)$ (medium), and $B_{12}(t)$ (darkest). Bottom: the function $f(t) = \sin(\frac{t}{5})$ (light) with approximations (dark) centered at $t = 4, 8, 12$ (indicated by the vertical dotted lines).

all of the (time-shifted) interpolatory functions defined for every sample point. Suppose that a signal f is sampled, and we wish to approximate the value of $f(t_1)$ using finitely many parameters, where t_1 is a fixed point not at a sample point. Choosing t_0 as the closest sample point to t_1, the CD approximation yields $f(t_1) \approx \sum_{k=0}^{M} K_k[f](t_0)B_k(t_1 - t_0)$, and the Nyquist approximation yields $f(t_1) \approx \sum_{k=-N}^{N} f(t_0 + kT)\text{sinc}(t_1 - t_0 - kT)$. As the CD approximation requires only parameters from the nearest sample point, we call it a *local* representation. The Nyquist representation is said to be *global*, as its approximation requires distant signal sample points.

The paper [3] presents chromatic differentiation in the context of a multirate filter bank, and points out that equation (1) above represents a limiting case of multiplexing and decimating: we get infinitely many parameters at a single time point. The *perfect reconstruction* property (equation (1)) is proved.

The current paper represents an attempt to move toward analysis of some of the practical issues in obtaining and using chromatic derivatives. The analyses given are preliminary and are intended to indicate some directions of inquiry rather than full investigations.

2 Relative Signal Spectrum

We restrict our attention to even weight functions $W(\omega)$ (i.e., those having real inverse Fourier transforms) for which $\int_{-\infty}^{\infty} W(\omega)\, d\omega = 2\pi$. If we always choose $P_0(\omega) = 1$ and $P_1(\omega) = aj\omega$ for $a > 0$, then W generates at most one orthonormal family of polynomials. I.e., if there exists a family of polynomials satisfying

$$\int_{-\infty}^{\infty} P_n(\omega)P_m^*(\omega)W(\omega)\, d\omega = \begin{cases} 2\pi & n = m \\ 0 & n \neq m \end{cases}$$

where each P_n is of degree n, then this family is unique; the associated family of basis functions $B_n(t)$ is of course also unique. Given such a weight, we let $\mathcal{S}(W)$ denote the class of all functions $f(t)$ such that $f(t) = \sum_{k=0}^{\infty} K_n[f](0)B_n(t)$; we implicity restrict our attention to those families W for which $\mathcal{S}(W)$ is nonempty.

Given any W and any wide-sense stationary (WSS) f in $\mathcal{S}(W)$, define $R_{km}(\tau) = E[K_k[f](t)K_m[f]^*(t + \tau)]$ where E is the statistical expectation operator. $R_{km}(\tau)$ is the cross-correlation function for the k^{th} and m^{th} chromatic derivatives of f; $R_{kk}(\tau)$ is the autocorrelation function of $K_k[f](t)$. Define $S_{km} = \mathcal{F}[R_{km}]$. S_{00} is the power spectral density of f. Note that $S_{km} = S_{k0}P_m^* = S_{00}P_kP_m^*$. The spectra and cross-spectra of the CD's of f therefore depend solely on the spectrum of f and the polynomials of the CD family. This indicates that correlations between the CD's of f at a given point t_0 can be used to estimate the spectrum of f relative to $W(\omega)$, much as the correlations between Nyquist-rate samples can be used to estimate the absolute spectrum. In particular, when $S_{00}(\omega) = W(\omega)$, we have

$$R_{km}(0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} W(\omega)P_k(\omega)P_m^*(\omega) = \begin{cases} 1 & k = m \\ 0 & k \neq m \end{cases}$$

Thus, the CD's of f are uncorrelated exactly when the power spectrum of f is $W(\omega)$. We say that such a signal is *white relative to* $W(\omega)$. This fact is applied to adapative filtering in [4].

3 Approximate Reconstruction

We wish to consider reconstruction of a signal given only finitely many of its CD's at each of the regularly spaced sample points. This is analagous to truncating the interpolation function in the Nyquist representation, which we now analyze for illustration.

The function $f(t) = \sum_{k=-\infty}^{\infty} a_k\text{sinc}(t - k)$ is approximated by $g_M(t) = \sum_{k=n_t-M}^{n_t+M} a_k\text{sinc}(t-k)$, where n_t is the integer nearest to t. When the variances of f and g_M are given by σ_f^2 and $\sigma_{g_M}^2$, respectively, and

both have zero mean, the error variance (the noise power) is given by $\varepsilon_M^{\text{sinc}}(t) = E[(f(t)-g_M(t))^2] = \sigma_f^2 + 2E[f(t)g_M(t)] + \sigma_{g_M}^2$. If the a_k are uncorrelated (i.e., if f is white), then $E[f(t)g_M(t)] = \sigma_{g_M}^2$, so $\varepsilon_M^{\text{sinc}}(t) = \sigma_f^2 - \sigma_{g_M}^2$. If the a_k have variance σ_A^2 then $\sigma_f^2 = \sigma_A^2\rho(0)$, where $\rho(t) = \text{sinc}(t) * \text{sinc}(-t)$. (A more detailed presentation is available, for example, in [5].) In order to analyze σ_{g_M} in a similar manner, we observe that $g_M(t)$ can be expressed as $\sum_{k=-\infty}^{\infty} a_k \text{sinc}_M(t-k)$ where $\text{sinc}_M(t)$ is $\text{sinc}(t)$ for $|t| \leq M$ and is 0 otherwise. Thus $\sigma_{g_M}^2 = \sigma_A^2\rho_1(0)$ where $\rho_1(t) = \text{sinc}_M(t) * \text{sinc}_M(-t) = \int_{-M}^{M} \text{sinc}^2(\tau)\, d\tau$. Thus $\varepsilon_M^{\text{sinc}}(t) = \sigma_A^2(1 - \int_{-M}^{M} \text{sinc}^2(\tau)\, d\tau$. The independence of $\varepsilon_M^{\text{sinc}}(t)$ on t indicates the convergent sequence of approximations obtained by using increasingly larger M is *uniformly convergent*, a property shared among global approximations.

Given $K_0[f](0), \ldots, K_M[f](0)$, the error $\varepsilon_M(t)$ in the synthesized signal is given by

$$f(t) - \sum_{k=0}^{M} K_k[f](0)B_k(t) = \sum_{k=M+1}^{\infty} K_k[f](0)B_k(t) \quad (2)$$

In order to avoid convergence issues, we consider consider $\varepsilon_{M,N}(t) = \sum_{k=M+1}^{N} K_k[f](0)B_k(t)$. Then $\varepsilon_M(t) = \lim_{N\to\infty} \varepsilon_{M,N}(t)$. Note that $K_k[f](0)$ is a random variable for each k; we denote its value by x_k. The error variance $E[\varepsilon_{M,N}^2(t)]$ is given by

$$\begin{aligned} & E\left[\left(\sum_{k=M+1}^{N} x_k B_k(t)\right)\left(\sum_{m=M+1}^{N} x_m B_m(t)\right)\right] \\ &= \sum_{k=M+1}^{N}\sum_{j=M+1}^{N} B_k(t)B_m(t)E[x_k x_m] \\ &= \sum_{k=M+1}^{N}\sum_{j=M+1}^{N} B_k(t)B_m(t)R_{km}(0) \qquad (3) \end{aligned}$$

When $S_{00}(\omega) = aW(\omega)$, the uncorrelatedness of the CD's gives us

$$\begin{aligned} E[\varepsilon_M(t)^2] &= \sigma^2 \sum_{k=M+1}^{\infty} B_k(t)^2 \\ &= \sigma^2\left(1 - \sum_{k=0}^{M} B_k(t)^2\right) \qquad (4) \end{aligned}$$

where (4) comes from the fact that $\sum_{k=0}^{\infty} B_k(t)^2 = 1$ [1, 6]. An example of this error is graphed in figure 2. For this case, the standard deviation of the error has the same form as the maximum absolute error bound

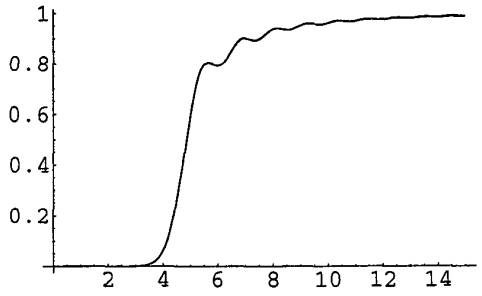

Figure 2: Average power of approximation error $\varepsilon_{12}(t)$ as a function of distance from the center of approximation. $W(\omega) = \frac{4}{\pi}\sqrt{1-(\omega/\pi)^2}$ (generating Chebyshev polynomials of the second kind). The signal being approximated is assumed to have average power $\sigma^2 = 1$.

established in [1], but with average power replacing total energy. If we know in advance the power spectra S of the class of signals we expect to analyze, we can attain this error variance by defining CD's based on $W(\omega) = S$. As a CD family is definable for any W corresponding to the power spectral density of a signal, we can choose to implement a given system relative to the spectra of the signals we expect to see.

4 Obtaining Chromatic Derivatives

We next turn to FIR implementation of the analysis filters. The transfer function of each operator $K_k[f](t)$ is given by $P_k(\omega)$, the k^{th} polynomial of the W-orthogonal family. Since the CD's are meant to be time-domain parameters, we optimize filters which will minimize the mean squared error in the time domain. Standard filter design techniques can be applied which yield reasonable results, but the only frequency-domain technique which minimizes the mean squared error in the time domain is minimization of the mean squared error in the frequency domain. Differentiators designed in this way usually require a frequency-domain weight, in which case the equivalence with time-domain mean squared error minimization is lost. For this reason, we consider projection (least-squares fit) in the time-domain, as developed in [1]. For a given signal $f(t)$ we represent each chromatic derivative $K_k[f](t_0)$ by the variable c_k. We then minimize the mean squared error under a given time domain weight $w(t)$ for a given order M of approximation and a given length $2s$ of signal to observe:

$$\int_{t_0-s}^{t_0+s} \left(w(t)\left(f(t) - \sum_{k=0}^{M} c_k B_k(t-t_0)\right)\right)^2 dt \quad (5)$$

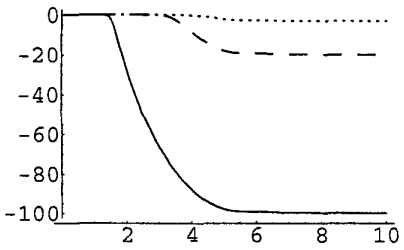

Figure 3: The weight function (equation 6) for $M = 12$ for the Chebyshev U CD's (centered at $t = 0$), for signal-to-noise ratios of 0dB (dotted), 20dB (dashed), 100dB (solid).

(For digital filtering, the integration must be approximated by a sum, so that oversampling of the signal becomes necessary.) The weight function $w(t)$ should be based on equation (4): we wish to maximize the fit where the approximation should be best. A natural criterion for a good weight might be that its square be inversely proportional to the variance of the error $\mathrm{err}(t)$:

$$\frac{w(t_1)^2}{w(t_2)^2} = \frac{E[\mathrm{err}(t_2)^2]}{E[\mathrm{err}(t_1)^2]}$$

If we assume that some (WSS) noise signal $n(t)$ is present with average power σ_n^2, and we assume that the signal to be analyzed has power spectral density $W(\omega)$ and average power σ^2, then $E[\mathrm{err}(t)^2] = E[(n(t) + \varepsilon_M(t))^2] = \sigma_n^2 + \sigma^2(1 - \sum_{k=0}^{M} B_k(t)^2)$, since the noise and the approximation error are uncorrelated and both have mean 0. If we normalize the weight function so that $w(0) = 1$ and substitute $w(t)^2/w(0)^2$ into the above criterion, we obtain

$$w(t)^2 = \frac{1}{1 + \frac{\sigma^2}{\sigma_n^2}(1 - \sum_{k=0}^{M} B_k(t)^2)} \qquad (6)$$

Note that $\frac{\sigma^2}{\sigma_n^2}$ is the signal-to-noise ratio (expressed linearly) for the conditions in which differentiation is to occur. Figure 3 graphs this weight function. A similar weight function appears in [1]. (The weight function arrived at here was first proposed by Matthew Cushman.)

5 Refined Signal Reconstruction

Suppose that we have obtained a number of CD's at various time points on a signal f. As shown above, this allows us to reconstruct $f(t)$ with a certain accuracy by using the approximation formula for each point t with respect to the CD's obtained at the sample point closest to t. Of course, there are a number of ways one might attempt to improve this reconstruction. Figure 1 suggests the most obvious approach: in the region of the signal between two approximations where neither is extremely accurate, use both approximations together in order to determine the signal values. This could be done most directly by a weighted average of the two signals using the weight arrived at in equation 6.

An alternative approach is to attempt to use the CD's from multiple sampling points to obtain higher order CD's, allowing for greater accuracy when each is expanded independently. Of course, the above approach is really a special case of this one: we estimate $K_0[f](t)$ for t between consecutive sample points t_0 and t_1 from $K_0[f](t_0), \ldots, K_M[f](t_0), K_0[f](t_1), \ldots, K_M[f](t_1)$. In fact, the discrete version of the problem described in the above section can also be looked at as a special case of this problem: given samples $K_0[f](t_0), K_0[f](t_1), \ldots, K_0[f](t_N)$, we wish to estimate $K_1[f](t_k), K_2[f](t_k), \ldots, K_M[f](t_k)$. This suggests an approach for obtaining higher CD's from lower ones: we expand the signal about several sample points, possibly with averaging to improve accuracy, and then we apply a least squares fit to the resulting sample points. A number of these techniques are developed in [1].

Using the statistical approach presented above for error estimation, one may also treat the problem of obtaining higher CD's from lower ones as one of statistical estimation. Using classical statistics, the higher CD's to be obtained may be thought of as parameters which determine the probability density functions of the lower CD's, which are sampled in the presence of noise. As we may have information about the spectrum of the higher order CD's (as derived above), however, treating the higher CD's as random variables and estimating them using Bayesian techniques seems much more appealing.

Additional knowledge of the signal (such as the modulation technique which generated it) can also be incorporated into the probability density functions for the higher CD's. Determination of the likelihood term depends on the derivation of CD's at remote points from the higher order CD's to be estimated. Methods for expressing $K_n[f](t_1)$ as a linear combination of $K_0[f](t_0), K_1[f](t_0), \ldots$ are described in [1, 8].

When the CD's to be estimated are expected to have Gaussian distributions, the problem fits the linear estimation model. Let $\mathbf{x}$ be the vector of observed CD's (at multiple time points, possibly including t_0), let $\Theta = \langle K_0[f](t_0), \ldots, K_M[f](t_0) \rangle$ be the vector of

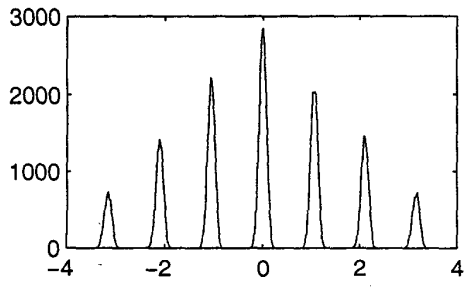

Figure 4: A very non-Gaussian distribution for the fourth Chebyshev U chromatic derivative of a 25% raised cosine signal. The Nyquist-rate sample points of the signal were uniformly distributed among $\{-3,-1,1,3\}$; approximately 100000 samples were used.

CD's to infer at $t = t_0$. The assumed noise in the samples can be based on the error term derived above; note that the term needs to be generalized to account for the error in higher order CD's. This is done easily given the expressions for remote CD's in terms of CD's at a given sample point. We then have $\mathbf{x} = \mathbf{H}\Theta + \mathbf{w}$ where $\mathbf{w}$ is the noise in the samples and $\mathbf{H}$ contains the linear combinations of estimated CD's which define sampled CD's. Note that some of the data may include CD's to be estimated. This allows us to re-estimate these CD's, in case the original estimates may contain noise. If the original estimates are not noisy, they can be subtracted out from the samples. (This includes appropriate modification of $\mathbf{H}$.) The covariance matrix, needed for the estimation and needed to analyze the error, is obtained from the functions $R_{km}(\tau)$ described above. See [7] for the details of Bayesian estimators.

When f is Gaussian, the CD's will be as well. Suppose instead that f is a modulated signal, say, $f(t) = \sum_{k=-\infty}^{\infty} a_k p(t-k)$ for some interpolation function p. Then $K_n[f](t) = \sum_{k=-\infty}^{\infty} a_k K_n[p](t-k)$. When the a_k are independent and identically distributed, it will be tempting to assume that the central limit theorem applies, but this must be done with care. The random variables being added are $a_k K_n[p](t-k)$, and these are not identically distributed. If the power of $p(t)$ is local, the central limit theorem will not apply. Figure 4 shows an experimentally obtained distribution of $K_4[\sum_{k=n_t-7}^{n_t+7} a_k rc(t-k)]$ where $rc(t)$ is a raised cosine interpolant with a 25% rolloff.

References

[1] Aleksandar Ignjatovic. Numerical differentiation and signal processing. Technical Report 1, Kromos Technology, Los Altos, CA, February 2001. `http://www.kromos.com`.

[2] P.P. Vaidyanathan. *Multirate systems and filter banks.* Prentice Hall, Inc., 1993.

[3] P.P. Vaidyanathan, A. Ignjatovic, and M.J. Narasimha. New sampling expansions for band limited signals based on chromatic derivatives. In *Proc. IEEE Asilomar Conf. Signals, Systems, and Computers*, Monterey, CA, November 2001.

[4] Tim Herron. Towards a new transform domain adaptive filtering process using differential operators and associated splines. In *International Symposium on Intelligent Signal Processing and Communication Systems*, Nashville, TN, November 2001.

[5] A. Papoulis. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill Book Company, New York, second edition, 1984.

[6] Matthew Cushman. The general theory of chromatic derivatives. Kromos Technology manuscript.

[7] S. M. Kay. *Fundamentals of statistical signal processing.* Prentice Hall, Inc., New Jersey, 1993.

[8] Timothy Herron and John Byrnes. Families of orthogonal differential operators for signal processing. Technical Report 2, Kromos Technology, Los Altos, CA, May 2001. `http://www.kromos.com`.