TOWARDS A NEW TRANSFORM DOMAIN ADAPTIVE FILTERING PROCESS USING DIFFERENTIAL OPERATORS AND ASSOCIATED SPLINES

T. Herron

Kromos Technology
4984 El Camino Real, Suite 210
Los Altos, CA 94022, USA
Phone: 1-650-575-2196, Fax: 1-650-210-0066
Email: tim@kromos.com

ABSTRACT

We describe a new system for representing signals using differential operators and how they can be used to apply linear operators to bandlimited signals. The representation, what we call the chromatic derivative expansion, requires more coefficients than ordinary sampling theory dictates, but the representation has attractive features which make it suitable for use in adaptive filtering. We study equalization for the purpose of introducing to the reader the use of chromatic derivatives in adaptive filtering.

1. INTRODUCTION

We present a new system of representing bandlimited signals [1,2,3], called the chromatic expansion, which should find important applications in adaptive filtering. The representation is similar to that provided by Taylor's formula for an infinitely differentiable function in that the chromatic expansion is a local, numerically non-uniform approximation to the function in terms of derivatives. But the chromatic expansion is superior to Taylor's expansion for signal processing because a finite chromatic approximation is a bandlimited signal, the approximation is more accurate than Taylor's approximation, and, most importantly, the expansion's coefficients are numerically obtainable from noisy signal samples. Finally, the representation is associated with an interesting method of applying filters to signals; one which applies a novel transform to the raw signal samples and can be used in adaptive filtering to speed convergence of filter coefficients and/or promote convergence in difficult conditions for adaptation.

In this paper, we first explain how the chromatic representation works as well as listing its most important properties. Then, we see how filtering operates when we represent a signal locally as a chromatic approximation. We discuss some situations where building filters using chromatic approximations might be worth the extra expense which the chromatic representation requires in acquiring and processing its parameters. Finally, we present some empirical results in adaptive filtering obtained through the use of chromatic derivatives as applied to the problem of channel equalization by performing a computer simulation.

2. THE CHROMATIC REPRESENTATION

We introduce the chromatic representation by presenting a schematic derivation of the main equation. We start with an interesting mathematical formula and use it to decompose the Fourier kernel. But first, we list some notation. We denote linear, shift invariant operators L when applied to signals f as $L[f]$. The transfer function describing L we denote by $L(\omega)$. We write the Fourier transform as $\mathscr{F}$, and the signal f's Fourier transform of course will be F.

2.1 Decomposing the Fourier Kernel

The formula we state here is one that is well-known in physics and in the applied mathematical community, and we state what its components are in the sequel ($j=\sqrt{-1}$):

$$e^{j\omega t}=\sum_{k=0}^{\infty} j^{k}(2k+1)P_{k}\left(\frac{\omega}{\pi}\right) j_{k}(t) \tag{1}$$

The j_k are the spherical Bessel functions of the first kind, being defined most easily from the Bessel function of the first kind, J_k, as:

$$j_k(t)=\frac{J_{k+\frac{1}{2}}(\pi t)}{\sqrt{2t}} \tag{2}$$

It is instructive to note at this point that $j_0(t)=\sin(\pi t)/(\pi t)$, the cardinal sine function which appears prominently in the Nyquist sampling representation equation. To continue, the P_k are the Legendre polynomials, which are recursively defined polynomials for which the following orthogonality property holds [4]:

$$\int_{-1}^{1} P_k(\omega)P_m(\omega)d\omega=\begin{cases} 0, & \text{if } k\neq m \\ \frac{2}{2m+1}, & \text{if } k=m \end{cases} \tag{3}$$

We then can perform the following schematic derivation for any finite energy, π-bandlimited function f at time t_0:

$$f(t) = \mathcal{F}^{-1}(F)(t) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(\omega)e^{j\omega t_0}e^{j\omega(t-t_0)}d\omega \tag{4}$$
$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} F(\omega)e^{j\omega t_0}\Big(\sum_{m=0}^{\infty} j^{m}(2m+1)P_m\Big(\frac{\omega}{\pi}\Big)\, j_m(t-t_0)\Big)d\omega$$
$$= \sum_{m=0}^{\infty} j_m(t-t_0)(2m+1)\mathcal{F}^{-1}\Big(F(\omega)j^{m}P_m\Big(\frac{\omega}{\pi}\Big)\Big)(t_0)$$
$$= \sum_{m=0}^{\infty} K_m[f](t_0)\sqrt{2m+1}\; j_m(t-t_0)$$

where $K_m[f](t_0) \equiv \mathcal{F}^{-1}(F(\omega)j^{m}\sqrt{2m+1}P_m(\omega/\pi))(t_0)$. The above is just suggestive, and is not a true derivation because we have ignored the issue of convergence in commuting the integral and summation signs. Although the mathematical formula (1)'s summation converges essentially non-uniformly, the problem can be dealt with adequately [2], giving us an exact representation of f as:

$$f(t) = \sum_{m=0}^{\infty} K_m[f](t_0)B_m(t-t_0) \tag{5}$$

for the normalized spherical Bessel functions $B_m(t-t_0) = \sqrt{2m+1}\; j_m(t-t_0)$. The equality (5) we call the *chromatic expansion*.

2.2 Chromatic Derivatives

We call operator K_m the *chromatic derivative* operator, and from the definition as applied to F above, it is seen to consist of applying the filter

$$K_m(\omega) = \sqrt{2m+1}\; j^{m}P_m(\omega/\pi) \tag{6}$$

to the function f. $K_m(\omega)$ is properly called a derivative filter because the filter is a polynomial in terms of ω; $K_m(\omega)$ is a linear combination of ordinary derivative filters up to order m, where the latter have the form, when normalized to the π-bandlimit:

$$\left(\frac{\partial}{\partial(t/\pi)}\right)^{m}(\omega) = \left(\frac{j\omega}{\pi}\right)^{m} \tag{7}$$

when applied to bandlimited signal f. However, the chromatic expansion (5) provides a superior representation of a bandlimited signal when compared to the Taylor expansion associated with ordinary derivatives. First, the functions B_m are themselves bandlimited and of finite energy unlike the monomials that Taylor's formula relies on. See Figure 1 (thick lines) for the first eight normalized spherical Bessel functions.

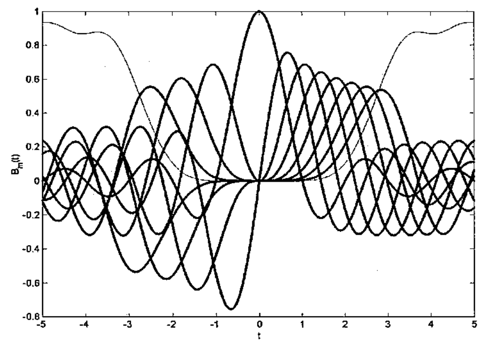

Figure 1

Also, numerical chromatic derivatives can be acquired from noisy data more reliably than ordinary derivatives; a fact that can be seen by comparing their respective transfer functions in Figure 2.

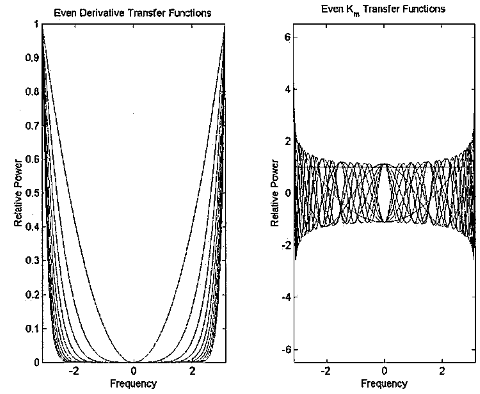

Figure 2

The transfer functions of the K_m use the entire spectrum in an equitable way and are orthogonal as a set whereas the transfer functions for ordinary derivatives effectively use only the part of the spectrum near the bandlimit [1].

2.3 Time Domain Functions and Perfect Reconstruction

The normalized spherical Bessel functions B_m that appear in the chromatic expansion can, in fact, be defined from the very same polynomials that define the chromatic derivative operators [1,2,3]:

$$B_m(t)=(-1)^m\sqrt{2m+1}\,\mathcal{F}^{-1}[j^m P_m(\omega/\pi)](t) \quad \textbf{(8)}$$

By recalling the orthogonality relation in equation (3), we see that the transfer functions $K_m(\omega)$ and $\mathcal{F}[B_k](\omega)$ are bi-orthogonal as two infinite sequences of transfer functions:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} K_k(\omega)\mathcal{F}[B_m](\omega)d\omega=\begin{cases}1, & k=m\\ 0, & k\neq m\end{cases} \quad \textbf{(9)}$$

But equation (9) is the condition for two infinite filter banks to be able to attain perfect reconstruction of a signal broken down into chromatic derivatives, sampled at one point t', and built up again using the time domain basis functions (see Figure 3)[5,6].

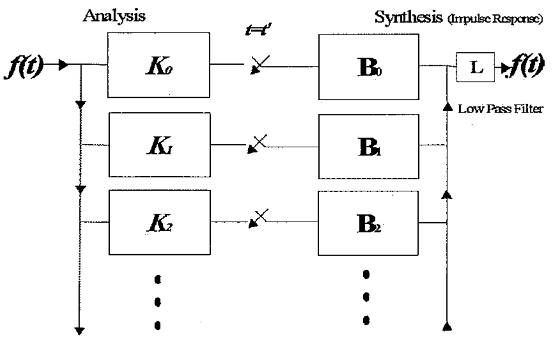

Figure 3

This fact provides another way of seeing that the chromatic expansion, equation (5), is true.

The two sets of biorthogonal filters that are implicit in the chromatic expansion are quite interesting for the theoretical implications they might hold, but how can they help one in adaptive filtering which approaches practicality? For that we need to the see how *finite* approximations of the chromatic expansion behave, which is the case where we restrict the above analysis and synthesis filter banks to be finite, resulting in imperfect reconstruction of a signal, even locally around t'.

2.4 Chromatic Approximation and Energy

The following tight upper bound inequality for a finite chromatic expansion, called a *chromatic approximation*, can be proven [1,2], and it is the reason we compare the chromatic expansion to Taylor's expansion:

$$\left|f(t)-\sum_{m=0}^{N} K_m[f](t_0)B_m(t-t_0)\right|\le E(f)\sqrt{1-\sum_{m=0}^{N} B_m^2(t-t_0)} \quad \textbf{(10)}$$

where $E(f)$ is the energy of the signal f. The square root term provides an accurate look at the relative (to energy) local approximation that a finite chromatic expansion provides for a bandlimited signal. Figure 1 shows the error curve, the thin line, for a finite approximation given the 0-th through 7-th chromatic derivatives (CDs) of f.

Equation (10) is what will distinguish our use of chromatic derivatives from ordinary filtering theory based on harmonic analysis because it provides a locally converging approximation of a signal given a finite number of CDs as its characteristic parameters. Such an approximation stands in contrast to those provided by basic Fourier analysis with its Nyquist interpolation [1] or analogous representations provided by wavelet theory [6]. Specifically, the approximation is focused at one point, t', and contains an infinite number of time domain representation functions instead of being a global sum of just one "mother" function φ shifted over many time points, as in:

$$f(t)=\sum_{k=-\infty}^{+\infty} c_k(f)\varphi(t-k) \quad \textbf{(11)}$$

Also, the chromatic approximation is generally much more accurate near the center than what is obtained from the *uniform* approximations given by Fourier theory or wavelet theory [1] when the latter are approximated by using signal samples.

A final equation to note, one which suggests that the chromatic derivatives provide a reasonable transform of a bandlimited signal f (as well as determining how we normalize the B_m) is

$$E(f)=\sqrt{\int_{-\infty}^{\infty}|F(\omega)|^2 d\omega} \propto \sqrt{\sum_{m=0}^{\infty}(K_m[f](t_0))^2} \quad \textbf{(12)}$$

(12) follows from Parseval's equality as do most other frequency-domain transforms' analogues (see [2,6]).

2.5 Chromatic Derivatives, Splines, and Sampling

With the chromatic approximation and its sharp error bound in hand, we propose to analyze bandlimited signals in a piecewise fashion, and adapt filters to those signals, by using spherical Bessel functions as splines. In particular, for the purpose of filter adaptation, we will acquire the first N+1 CDs at a rate dictated by a chosen level of accuracy that we can obtain via the error bound in (10). The rate specified by (10) - see [1,3] for some helpful formulas for computing it - will always be higher than that which is permitted under generalized Nyquist theorems which state that one can sample a signal and its N derivatives only every N+1 Nyquist intervals and still reconstruct the signal perfectly [1,7]. However, the sampling of *CDs* that one needs to perform in order to reconstruct the signal piecewise with accuracy using approximation (10) is less than twice the generalized Nyquist rate. Thus, the extra burden imposed by trying to attain (10) is not large.

Also, if we are to convert digital signal samples into accurate CDs using a finite filter bank of K_m FIR filters of practical length, then it appears certain that we need to sample at a rate higher than the Nyquist rate determined by the bandlimited signal f. This unfortunate fact is not surprising to those who represent signals using various spline or wavelet techniques for the purposes of good signal interpolation, such as is common in video processing. One reason why oversampling is needed to obtain accurate chromatic derivatives can be seen by looking at the speed of convergence of Lagrange polynomials approximating bandlimited functions f under various sampling rates and comparing them to the infinite Taylor formulas which are equivalent to the spherical Bessel functions - see [8,1,3]. It appears that oversampling at a rate of at least two times the Nyquist rate is required to obtain CDs which have the precision needed to take advantage of the local accuracy that the chromatic approximation's equation (10) claims.

One simple method of designing rather short oversample-rate FIR filters to implement the K_m transfer functions is by computing filter coefficients which minimize the approximation error in equation (10) through finding the best linear combination of B_m functions which approximate f locally, using the error curve $\sqrt{1-\sum_{m=0}^{N} B_m^2(t-t_0)}$ to build a window which defines locality [1]. In other words, we perform a multi-B_m spline using a least squares fit. Such filters are designed jointly and have impulse responses not much larger than the distance for which the CDs provide a good approximation. However, these filters are not particularly resistant to noise that is not far out-of-band, and so we assume that the filter bank computing the chromatic derivatives is preceded by an anti-aliasing filter in order that CDs can be acquired accurately from a noisy signal.

However, taking the unusual approach to signal representation, approximation (as in (10)), and oversampling that we have outlined in this subsection is not necessary for use of many of the results in this paper, especially those consequences following directly from using filters generated by orthogonal polynomials. However, we believe that maximal benefit from the CD representation (5) will accrue only when the unique convergence implicit in equation (10) is used.

2.6 Generalizations of Chromatic Representation

Before moving on to filtering, we want to briefly introduce a few useful generalizations of the chromatic representation [2,3]. If we replace the orthogonality relation (3) with the following weighted orthogonality relationship:

$$\int_{-\infty}^{\infty} P_k'(\omega)P_m'(\omega)\rho(\omega)d\omega = \begin{cases} 0, & \text{if } k\neq m \\ 2\pi, & \text{if } k=m \end{cases} \tag{13}$$

for (suitably vanishing, integrable) frequency weight ρ, and polynomials P'_m, then we can extend the chromatic expansion theory in two ways. First, we can acquire CDs which encode derivatives of non-bandlimited, square-integrable signals, though we will not purse this possibility in the present paper [2,3]. Second, we can make the resulting chromatic derivatives of a signal f "better prepared" numerically for signals with spectrum ρ - we will explain later precisely what this means. The weight associated with the Legendre polynomials is a special case of the above formula: ρ there has a flat, bandlimited spectrum.

One final important generalization that can be proven to apply to the CD framework is that for many important instances of weight ρ, the general chromatic expansion extends to tempered distributions (e.g. sine, $e^{j\omega t}$), and not restricted to finite-energy, square-integrable functions [2].

After selecting the weighted orthogonal polynomials, we can then define the chromatic derivatives and associated spline functions which obey the chromatic representation (5) as:

$$\begin{aligned} K_m(\omega) &\equiv j^{\,m}P_m'(\omega) \\ B_m(t) &\equiv \mathcal{F}^{-1}[(-j)^m P_m'(\omega)\rho(\omega)](t) \end{aligned} \tag{14}$$

Note that the spline functions B_m contain the weight ρ in their defining transfer functions - this is the reason for referring to the perfect reconstruction (which works for weighted CDs) spline and CD pairs (9) as a *bi*orthogonal pair of sequences. We also obtain the approximation error formula (10) with the one modification being that $E(f)$ is now a weighted energy term:

$$E(f) \equiv \sqrt{\int_{-\infty}^{\infty} \frac{|F(\omega)|^2}{\rho(\omega)} d\omega} \propto \sqrt{\sum_{m=0}^{\infty} (K_m[f](t_0))^2} \tag{15}$$

which is proportional to the sum of the squares of the CDs as in (12).

2.7 CD Recursion and Iteration

Polynomials orthonormal with respect to a weight, equation (13), can be computed by using a simple recursive algorithm as follows; a fact that is guaranteed by the Gram-Schmidt orthogonalization procedure [9,3,4].

$$P'_0(\omega)=1$$
$$P'_1(\omega)=a_0\omega-c_0 \qquad \textbf{(16)}$$
$$P'_{n+1}(\omega)=(a_n\omega+c_n)P'_n(\omega)-b_nP'_{n-1}(\omega), \quad n\geq 1$$

The constants a_n, b_n, and c_n are fully determined by (13) given the requirement that each P'_m is a polynomial of order m. The most interesting weights ρ are even, $\rho(\omega)=\rho(-\omega)$, which implies that $c_n=0$, and we will restrict ourselves to such cases in the sequel since they handle the case where signal f is real valued.

For the purposes of filtering, there is one more important relationship that the recursive equations of (16) imply, the so-called *linearization* formulas for orthogonal polynomials. The constants in the following set of linearization equations can be recursively computed by recursion starting from the constants a_n and b_n in (16) [3]:

$$P'_m(\omega)P'_k(\omega)=\sum_{i=0}^{\min\{m,k\}}\alpha_{m,k,i}P'_{m+k-2i}(\omega) \qquad \textbf{(17)}$$

(17) immediately implies an important equation, used when iterating applications of the chromatic derivative operator:

$$K_m[K_k[f]]=\sum_{i=0}^{\min\{m,k\}}(-1)^i\alpha_{m,k,i}K_{m+k-2i}[f] \qquad \textbf{(18)}$$

Equation (18) provides an important link between different CDs and will allow us to use them in filtering to over specify a filter's action upon a signal. Finally, we record the equation:

$$B_m(t)=(-1)^mK_m[B_0](t) \qquad \textbf{(19)}$$

which is very useful and follows immediately from (14).

3. FILTERING USING CHROMATIC DERIVATIVES

In this section we describe the local effects of linear, time-invariant operators on the CD representation of a signal (10). First, we describe the statistical effects that the CD transform (14) has upon signals to which it is applied. We then describe how such transfer function filters act upon a finite chromatic approximation.

3.1 CDs of Wide Sense Stationary Signals

We first note a few facts about the statistics of filtering when using CDs. These facts will provide guidance to the uses that the chromatic representation (5) and approximation (10) can be put to in adaptive filtering.

If X and Y are wide-sense stationary mean 0 random variables on the real number line (thought of as describing bandlimited signals), then we define the crosscorrelation function as $R_{XY}(t_0) = E_t(X(t)Y^*(t-t_0))$ and the autocorrelation function when $X=Y$, where E is the statistical expectation operator. A well-known theorem [10] dealing with applying filters to random variables tells us, when applied to the CD filters K_m, that:

$$\mathcal{F}[R_{K_m[X],K_n[X]}(t)]=\mathcal{F}[R_{X,X}(t)]\cdot K_m(\omega)\cdot K_n^*(\omega) \qquad \textbf{(20)}$$

In other words, the cross-spectrum, the Fourier transform of the crosscorrelation function, of random variables $K_m[X]$ and $K_n[X]$ is the spectrum of the process X multiplied by the transfer functions of the two chromatic filters. Notice then what transpires if X happens to have the spectrum $\rho(\omega)$:

$$\begin{aligned} R_{K_m[X(\rho)],K_n[X(\rho)]}(0) &= \mathcal{F}^{-1}[\rho(\omega)K_n(\omega)K_m^*(\omega)] \\ &= \mathcal{F}^{-1}[\rho(\omega)P'_n(\omega)j^nP'_m(\omega)(-j)^m] \\ &= \begin{cases} 0, & m\neq n \\ 1, & m=n \end{cases} \end{aligned} \qquad \textbf{(21)}$$

by using equation (13). We shall call, for $1\leq n,m\leq Q$ given any fixed integer Q, $R_{K_m[x],K_n[x]}(t)$ the *CD-autocorrelation matrix* of signal x at point t.

The interpretation that we give to equations (21) and (15) is the following. The CDs that are based upon the orthogonality relation of equation (13) provide a new method of transform-domain filtering [11,12]; a comparative one in which signals are filtered in comparison to a signal with spectrum $\rho(\omega)$, the latter signal the CDs treat as statistically uncorrelated signals. And the weighted power estimate that the sums of squares of CDs provide is a comparative power estimate - one comparing the power of the signal to the power of $\rho(\omega)$ within each frequency band. It is immediate how useful (21) should be in adaptive filtering given that convergence of filter coefficients to their desired values depends upon how far from diagonal is the autocorrelation matrix of the filter's input signal [11,13].

3.2 CD Filtering

Let us apply a shift-invariant bounded linear operator $L[\cdot]$, with transfer function $L(\omega)$, to a finite chromatic approximation of bandlimited, power-limited signal f centered at the point $t'=0$. First, however, we need to know two facts about applying L to any B_m [1,2]:

- $L[B_0]$ is a function which can be represented by an infinite chromatic expansion $L[B_0](t)=\sum_{k=0}^{\infty}K_k[L[B_0]](0)B_k(t)$.
- A good finite chromatic approximation of $L[B_0]$ depends upon the filter L; in particular the order M can be set by using the $\|L(\omega)\|_\infty$ norm, the maximal action of the filter on some frequency [1]. A more reasonable and workable M can be

obtained, however, by using $\|L(\omega)\|_2$, the mean square metric applied to the filter.

Then, by using (18) and (19) and commuting the filters L and K_m, we can derive the following approximations to $L[f]$, by starting with the finite approximation of f:

$$\begin{aligned} L[f](t) &\approx \sum_{i=0}^{N} K_i[f](0)L[B_i](t) \\ &\approx \sum_{i=0}^{N} K_i[f](0)(-1)^i K_i[L[B_0]](t) \\ &\approx \sum_{i=0}^{N} (-1)^i K_i[f](0) K_i\left[\sum_{j=0}^{M} K_j[L[B_0]](0)B_j(t)\right] \end{aligned} \tag{22}$$

We now compute, using (18) once and (19) twice:

$$\begin{aligned} K_i[B_j] &= (-1)^j K_i[K_j[B_0]] = (-1)^j \sum_{k=0}^{\min\{i,j\}} (-1)^k a_{i,j,k} K_{i+j-2k}[B_0] \\ &= (-1)^i \sum_{k=0}^{\min\{i,j\}} (-1)^k a_{i,j,k} B_{i+j-2k} \end{aligned} \tag{23}$$

which allows us to conclude from (22) that near the point 0:

$$L[f](t) \approx \sum_{j=0}^{M}\sum_{i=0}^{N} K_i[f](0)\cdot K_j[L[B_0]](0)\cdot \sum_{k=0}^{\min\{i,j\}} (-1)^k a_{i,j,k} B_{i+j-2k}(t) \tag{24}$$

By substituting index $p=i+j-2k$ for the index k and rearranging sums, we obtain the following local approximation (near 0), where *mod2* is the modulo 2 operation:

$$L[f](t) \approx \sum_{p=0}^{M+N} B_p(t) \sum_{i=\max\{0,p-M\}}^{\min\{N,M+p\}} K_i[f](0)\cdot \sum_{j=\max\{p-i,mod2\{p-i\}\}}^{p-i+2*\min\{i,M\}} (-1)^{(i+j-p)/2} a_{i,j,(i+j-p)/2} K_j[L[B_0]](0) \tag{25}$$

Note that (25) has the form of a finite chromatic expansion. We wish to do a few more transformations and simplifications to the above equation to reduce it to its final form. First, we substitute the index $m=(i+p-j)/2$ into (25) for the index j and then use the fact that $a_{u,u+v,u-w}=a_{u,v+2w,w}$, which is true by recursion from (16) and (17). Finally, we will simplify the lower and upper indices of the resulting three summations by defining $p_{\max}\equiv M+N$, and defining analogous terms for the other sums (along with performing some simplifications). Thus, we can derive from (25) the following fundamental filtering equation:

$$\begin{aligned} L[f](t) &\approx \sum_{p=0}^{p_{\max}} B_p(t)\cdot K_p[L[f]](0), \quad \text{where} \\ K_p[L[f]](0) &\approx \sum_{i=0}^{i_{\max}} K_i[f](0)\cdot \sum_{m=0}^{\min\{i,p\}} (-1)^{(i+m)} a_{i,p,m} K_{i+p-2m}[L[B_0]](0) \end{aligned} \tag{26}$$

Equation (26) is an approximation of the application of filter L to signal f, and it says three important things [1]:

- Linear filtering can be thought of as a local operation of linearly combining CDs of f at 0 to produce CDs of $L[f]$ at 0:
- The filter L's action on any signal f at 0 is determined by the action of the filter on B_0 at 0: this is much like how a filter's impulse response determines how the filter acts upon other signals when using ordinary convolution filtering.
- The filtering coefficients which use the CDs approximating f near a central point and also produce CDs approximating $L[f]$ near the point form a matrix. However, note that this *matrix* of values, as mentioned above, are determined by only one *vector* of values: the CDs of $L[B_0]$ at the central point 0.

By elementary Fourier analysis, the orthogonality relation (13), along with the definition of K_m, implies that $K_m[L[B_0]](0)=\mathcal{F}^{-1}[P_m'(\omega)\rho(\omega)L(\omega)](0)$ are coefficients which approximate the spectrum of the transfer function $L(\omega)$:

$$L(\omega) \approx \sum_{m=0}^{M} K_m[L[B_0]](0)\cdot P_m'(\omega) \tag{27}$$

Thus, the order M of approximation mentioned previously can also be determined by knowing how many polynomials it takes to approximate the transfer function $L(\omega)$ of the filter to implement. Equation (27) suggests that we call the coefficients $K_m[L[B_0]](0)$ the *projection coefficients* of the filter $L(\omega)$, since they project the transfer function of the filter onto the space of orthogonal polynomials that we use to build the CDs. The projection coefficients thus provide an FIR-like model of the linear filter that we wish to approximate in the sense that the transfer functions that we can build using a finite number of CDs are polynomials (though not polynomials in the z-transform as true FIR filters).

4 ADAPTIVE FILTERING USING CDS

We now turn to the main goal of this paper and discuss the prospects of using chromatic derivatives in adaptive filtering. First, we gather together and elaborate previously stated properties of the CD-transform that are related to adaptation, especially those properties which contrast our approach with standard methods [11,12,13]. Then, we walk through an application of CDs to equalization and report on some numerical simulations of a simple model of equalization.

4.1 CD-Transform Domain Filtering Properties

Let us repeat here the following basic properties that we have noticed about the chromatic derivative transform:

- CDs can be designed for a power spectral density $\rho(\omega)$ which determines how signals are processed using CD filters K_m : the resulting CD-autocorrelation matrices and sums of squares of CDs compare a processed signal's statistics and energy, respectively, to those of the "ideal" signal with PSD $\rho(\omega)$.
- The application of the CD filters K_m obey important iteration relationships (18) which translate into a CD filtering structure which has significant redundancies (26): Columns of the CD transformation matrix, which implements a linear filter when applied to a signal's CDs, share coefficients with one other.
- Accepting the point of view of subsections 2.4 and 2.5, we view the chromatic approximation and its error bound as a new way of performing fractionally-spaced filtering. CDs are acquired digitally using fractionally-spaced filters, but filtering is performed using densely encoded CDs, which only encode inband information, thus avoiding out-of-band nulls.

Let us take a look at each point in more detail.

4.1.1. Custom Transform Domains

The first property above allows us to view our approach to adaptive filtering as an extension of the usual methods of transform domain filtering. The extension is that we make available a whole new class of transforms. These transforms stand in contrast to standard transforms, for example the DCT [11,12,14], which are also used in filtering in order to diagonalize the autocorrelation matrix used in solving for a linear filter. But their use is simply based upon the *similarity* between the transform domain's basis vectors and the eigenvectors of the autocorrelation matrix computed from a filter's incoming signal (see [6] for an especially good discussion of basis vectors of standard transforms), whereas the CD transform can be designed to fit a particular application. It remains to be seen whether fast ρ-CD transforms can be designed to process the incoming signals. If so that would negate a point currently in favor of the standard transforms like the DCT when doing transform domain adaptive filtering.

4.1.2 Overdetermination of the Projection Coefficients

The second property above distinguishes our CD transform filters from those transforms, such as the DFT, which break up a signal into separate frequency components and filter them either jointly or separately [11,12,15]. Because subband transform adaptive filters are very effective in reducing the eigenvalue spreads of the autocorrelation matrices by isolating separate subband component of the spectrum, there will be no cross- dependance between the transform's different values as there is with CDs which can be used jointly in adaptation. To make this point clear, we give two examples of how this cross-dependance, or redundancy, can be used.

First, if while training a filter we have access to the values of $K_m[L[f]](0)$ for several $m>0$ in addition to knowing signal values (i.e. for $m=0$), then we can use this additional information to help determine the projection coefficients $K_m[L[B_0]](0)$ for the filter L. In fact, we will see this idea used in the next subsection in our example equalization simulation, so we postpone further discussion until then.

Second, one can do the opposite of the first example's procedure - adaptively solve for a filter knowing nothing of the higher CDs that the filter should output, but then filter *to produce* higher CDs using the converged projection coefficients. For example, in equalization it can be useful to produce equalized higher derivatives of shaped symbols for the purpose of extracting timing information or perhaps as extra information for the symbol slicer. The facts (apparent from (16) and (17)):

$$a_{0,j,k}=a_{i,0,k}=1 \quad \text{for all } i,j,k\geq 0 \qquad \textbf{(28)}$$

imply, when used with equation (26) in this case, $p=0$, that solving for the projection coefficients, $K_m[L[B_0]](0)$, is easy. The vector of projection coefficients (every other one negated) times the vector $K_m[f](0)$ equals $K_0[L[f]](0)$. Thus, to adaptively solve for the projection coefficients by minimizing the mean squared error of $K_0[L[f]](0)$ involves solving the system of equations in matrix form:

$$\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & (-1)^n R_{K_m[f],K_n[f]} & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \cdot \begin{pmatrix} \cdot \\ K_m[L[B_0]] \\ \cdot \end{pmatrix} = \begin{pmatrix} \cdot \\ R_{K_n[f],K_0[L[f]]} \\ \cdot \end{pmatrix} \qquad \textbf{(29)}$$

at the center point 0. Solving for the projection coefficients via equation (29) will allow us to calculate $K_i[L[f]](0)$ for $i\geq 0$. And any ordinary adaptive method (LMS, RLS, etc.) can be used to solve the CD-autocorrelation matrix equation of (29).

4.1.3 CD Filtering as Fractionally-Spaced Filtering

The third property of CD filtering mentioned above is based upon a rather simple idea. The chromatic approximation provides a way of representing a bandlimited signal in a precise way locally, and that precision requires oversampled signal values in order to produce the signal's CDs accurately. In particular, this dense representation uses the fractionally-spaced samples to encode signal phase information very accurately, as the chromatic error bound (10) should indicate. However, the fact that bandlimited CDs encode only in-band information gives rise to a potential advantage for CD filtering over ordinary fractionally-spaced adaptive equalization [16,17], for example.

Namely, CD filtering does not suffer from the following trade-off: that of using fractionally-spaced samples to deal with phase timing and band-edge power situations which can give rise to inband nulls on the one hand (such nulls increase autocorrelation eigenvalue spread), with the resulting inclusion of very low power out-of-band frequency components into the filter due to the oversampling on the other hand. To repeat, CDs do not encode low-power, out-of-band frequencies while still benefitting from the phase information they contain.

Of course, including "empty" higher frequencies into the structure of a filter for the sake of better phase information by eliminating aliasing is not always a problem: e.g. out-of-band noise is often sufficient to keep the eigenvectors of the autocorrelation matrix far enough away from 0. Also, it is well-known theoretically how to solve for the excess eigenvectors that fractional spacing introduces into an autocorrelation matrix [18,16], and there are special methods of solution for the Weiner filter by which these out-of-band eigenvectors can be avoided [17]. Our point is simply that CDs avoid this potential problem by their very construction.

It will be interesting to perform extensive simulation testing of CD filtering to see if the above claims can be consistently born-out in practical settings. This is especially true given its use of fractionally spaced samples and its use of redundant information in light of other recently developed equalization methods that make similar claims [19,20].

4.2 Example: Equalization Using The CD-Transform

In an attempt to provide some empirical confirmation of the usefulness of CDs in linear adaptive filtering, we simulate equalizing a distorting channel which is being used to transmit an ordinary PAM signal shaped by a raised cosine filter. In the spirit of thinking of using CDs as a way of doing transform domain filtering, we compare our results to those obtained by filtering using the DCT transformation. We restrict ourselves to using the transform-domain, power-normalized LMS adaptation method [12] in both cases when solving for filter coefficients because the method is simple, stable, and practical [11,13].

4.2.1 Training on the Signal Alone

The term "CD-LMS" is what we will call the method of solving equations (29) using a transform-domain version of the most basic stochastic gradient method [11,12]. In comparing the DCT-LMS method to the CD-LMS adaptation method, we select a channel that is sensible to want to equalize and also is favorable to the DCT transform [6,11,14]: a single-pole, low pass ("AR(1)") filter, namely $C=(1/40)(1-(39/40)z)^{-1}$ in z-transform notation where we assume the upper bandlimit of C to be 6π. We normalize the PAM signal, before pulse shaping it, as being π-bandlimited, and let r be the raised cosine rolloff factor, using $r=1.2$ to mean that we use a 20% roll-off, which is what we will use in the simulation.

On the side of the CD-LMS adaptation method, we choose the "ideal" spectrum response to be $r\pi$-bandlimited with spectrum response $\rho(\omega)=(1-(\omega/(r\pi))^2)^{3/2}$. We choose this particular spectrum $\rho(\omega)$ so that the CD-LMS process "expects" higher frequencies to be attenuated more severely by the channel while at the same time avoiding too good of a match to C. Because filtering with CDs is filtering with respect to the ideal signal with PSD $\rho(\omega)$, the sums of squares of a CD equalizer will not reflect the energy of the inverse channel, but the weighted energy of the inverse channel; as in equation (15). In the present case this implies that had we chosen $\rho(\omega)=C$, we would be boosting any white noise coming into the equalizer twice - once because the CDs we use do not expect high frequencies which are present in the white noise, and once because the filter coefficients are large in order to produce the original PAM signal which is mostly white. Also, the CD filter functions K_m and B_m are easy to compute in this case [3]. Finally, we elect to capture nearly all of the bandwidth of the shaped PAM signal by using an $r\pi$-bandlimit instead of a π-bandlimit for $\rho(\omega)$.

We performed a Matlab simulation for a PAM-4 signal sent over the channel C with additive noise applied to give a 30 dB signal at the receiver's front end. Figure 4, a plot of each filter's errors during initial convergence (for 100 averaged runs) indicates that the convergence of the CD-LMS method, the black line, is not quite as fast as the convergence given by the DCT-LMS method, the gray line, but is faster (and ultimately more accurate) than plain LMS, the dotted line.

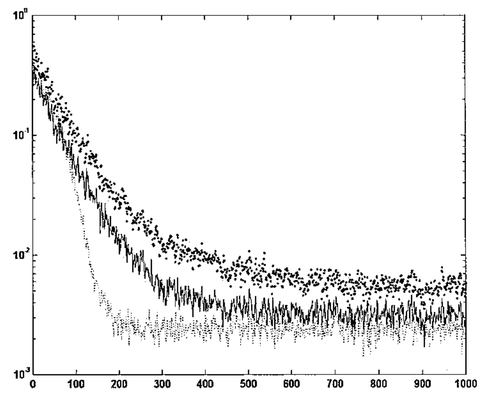

Figure 4

A few qualitative observations and notes on the simulations:

- The CD-LMS routine uses CD filters that are operating at an oversample-rate with a 3π-bandlimit. Using a 2.5π-bandlimit CD filter oversampling results in about the same performance as the LMS routine. The DCT transform operates at 1.5π, 2π

and 3π-bandlimit sampling rates with similar performance in each case, while the performance degrades at bandlimits of 1.2π and below. Ditto for the LMS method.

- The signal is pre-filtered at the receiver before being transformed by the DCT or CD transforms, which is neccesary for the CD-LMS algorithm to outperform the LMS algorithm, by having a sharper and/or more severe low-pass pre-filter emptying the frequencies beyond $r\pi$. In particular, a weak low-pass filter results in similar performance for both algorithms. As expected, the LMS and DCT-LMS algorithms are relatively insensitive to noise in frequencies beyond $r\pi$ but within its (oversampled) bandwidth.
- The impulse response of the channel C is approximately 20+ Nyquist intervals, measured using a π-bandlimit. Although sampling was done at the 3π-bandlimit rate, we only required 22 CD taps for the CD-LMS algorithm for the performance recorded above. Using a 1.5π-bandlimit sampling rate for the DCT-LMS algorithm, we used 37 DCT taps to get about as good a performance as we could get from the algorithm.
- As expected, all the above fractionally spaced equalizers are phase-insensitive (along with having no need for a matched filter).

4.2.2 Training on Several CDs

Now we take up the interesting suggestion of subsection 4.1.2 in which we use information about the higher-order derivatives associated with the training sequences the adaptation routine receives. We stay with the situation in 4.2.1 of adaptively equalizing a PAM signal. By incurring a small delay corresponding to the size of the raised cosine filter on the front end, we can in fact use the PAM training symbols to easily compute the first few derivatives associated with a raised cosine-shaped PAM symbol - a small table in the receiver can accomplish this task. The remaining challenge, of using this higher-order CD information to solve for the projection coefficients $K_m[L[B_0]](0)$, is met by using equation (26) again.

We propose to enhance the CD-LMS routine by adding to it a few parallel LMS routines which compare the values of the higher-order CDs from training to those values that the projection coefficients compute. In particular, using equation (26) we minimize the mean squared error between $K_p[L[f]](0)$ and the values computed by the computed CDs

$$\sum_i K_i[f](0)\cdot\sum_m (-1)^{(i+m)} a_{i,p,m} K_{i+p-2m}[L[B_0]](0) \qquad \textbf{(30)}$$

for each $p>0$ that we wish to use. Unlike the $p=0$ case, it takes some effort to rearrange the minimization equations to yield an MMSE solution (for each fixed p). However, if one does so, then one can then express the resulting set of equations in the matrix form:

$$\begin{pmatrix} \cdot & \cdot & \cdot \\ \cdot & R_{T_m[f],T_n[f]}(0) & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \cdot \begin{pmatrix} \cdot \\ K_m[L[B_0]](0) \\ \cdot \end{pmatrix} = \begin{pmatrix} \cdot \\ R_{T_n[f],K_p[L[f]]}(0) \\ \cdot \end{pmatrix} \qquad \textbf{(31)}$$

where the T_m are linear transformations of the filter *L's* input CDs computed from the signal *f*:

$$T_m[f] \equiv T_m[K_0[f], K_1[f], \ldots, K_Q[f]] \qquad \textbf{(32)}$$

Such linear transformations T_m of the CDs are computed from the $a_{i,j,k}$ coefficients of (17) and can thus be known in advance. Each such sequence of transformations is unique to a fixed $p>0$, but the transformations are usually easy to compute - for each $p>0$ they jointly form a sparse matrix. We can store the transformations as tables in the receiver, and if we solve for each $p>0$ the equations (31) using, e.g., the LMS method, then we can update the coefficients $K_m[L[B_0]](0)$ during every received symbol as many times as the number of p's (higher-order CDs of the training signal) we choose to use.

This method of using higher-order CDs obtained from the training signal we call the *multi-CD-LMS* algorithm. Use of each higher CD in a separate LMS routine provides a more-or-less independent gradient by which the algorithm converges towards the projection coefficients that are used to equalize the PAM symbols. Figure 5 demonstrates the improved convergence that can be obtained, as compared with the DCT-LMS and LMS methods again, when equalizing the same channel C as used in 4.2.1, when we use K_0, K_1, and K_2 in solving for the projection coefficients.

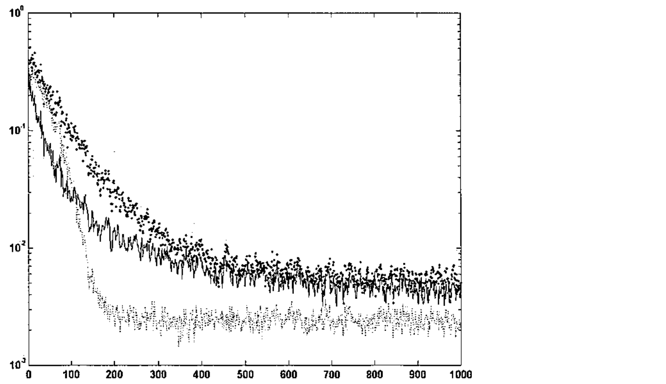

Figure 5

A few qualitative observations and notes on the multi-CD-LMS algorithm simulation used to produce the above chart:

- The CDs that we used in this case are those with the "ideal" spectrum being flat: $\rho(\omega)=1$ within the bandlimit $r\pi$. This case thus shows the performance of multi-CD-LMS when we have little idea of the type of channel that we want to equalize.
- The results above were obtained using a raised cosine filter of length 14 Nyquist intervals (NI in terms of the PAM signal's bandlimit, π). Filter lengths shorter than length 12 tend to degrade the ultimate convergence of the algorithm (though not the general speediness) for a rather simple reason: early termination of the ideal raised cosine's impulse response causes the shaped PAM signals to have significantly discontinuous higher derivatives, and thus discontinuous CDs.
- Each of the CD-LMS routines, one per higher-order training CD, can be turned on and off independently of the others. Thus, this provides another option for altering adaptation speed much like the technique of gear-shifting allows a designer to alter the convergence speed/convergence error properties of the adaptation process by altering the gradient step size.
- Faster initial convergence of the multi-CD-LMS algorithm, even in comparison to DCT-LMS, comes at a price: multi-CD-LMS increases the noise which hurts the ultimate convergence of the tap coefficients. This occurs because the transformations T_m which are used to convert CDs into coefficients are not unitary: on average they increase the norm of vectors they operate on (and thus boost noise).

In summary, there is no clear cut advantage for using CD transforms in equalization, but they do have some interesting properties which call for further study.

5. ACKNOWLEDGMENTS

The most important thanks for the results recorded in this paper go to Aleks Ignjatovic, the person responsible for inventing chromatic derivatives, discovering their associated spline functions, and working out the basics of the use of CDs in linear filtering. I would also like to thank the other members of the Kromos research team, John Byrnes and Matt Cushman, many of whose ideas, formulations, corrections, and suggestions are used herein and make this paper very much a joint effort, not counting all the errors and over generalizations for which I claim sole credit. Finally, Richard van Dorn's work on CD filtering simulations provided practical guidance in how to use CDs for adaptive filtering.

6. REFERENCES

[1] A. Ignjatovic, "Numerical Differentiation and Signal Processing", Technical Report #1, Kromos Technology, http://www.kromos.com

[2] M. Cushman, "The General Theory of Chromatic Derivatives", Manuscript, Kromos Technology, http://www.kromos.com

[3] T. Herron and J. Byrnes, "Families of Orthogonal Differential Operators for Signal Processing", Technical Report #2, Kromos Technology, http://www.kromos.com

[4] Abromowitz, Milton and Irene A. Stegun, eds., *Handbook of Mathematical Functions,* Dover, 1972.

[5] Vaidyanathan, P.P., *Multirate Systems and Filter Banks*, Prentice Hall, Englewood Cliffs, New Jersey, 1993.

[6] Akansu, Ali N. and Richard A. Haddad, *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*, 2nd Edition, Academic Press, San Diego, 2001.

[7] Unser, Michael, "Sampling - 50 Years After Shannon", Proceedings of the IEEE, vol. 88, no. 4, April 2000.

[8] Ignjatovic, A. and N. Carlin, "Signal Processing with Local Signal Behavior", Provisional Patent Application No. 60/143,074, U.S. Patent Office, July 9, 1999.

[9] Stoer, J. and R. Bulirsch, *Introduction to Numerical Analysis*, 2nd Ed., Springer Verlag, Berlin, 1993.

[10] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 1984

[11] S. Haykin, *Adaptive Filter Theory*, 3rd Edition, Prentice Hall, Upper Saddle River, New Jersey, 1996.

[12] Glentis, G-O., Berberidas, K., and S. Theodoris, "Efficient Least Squares Adaptive Algorithms for FIR Transversal Filtering", IEEE Signal Processing Magazine, July 1999.

[13] Treichler, John R., C. Richard Johnson Jr., and Michael Larimore, *Theory and Design of Adaptive Filters*, Prentice Hall, Upper Saddle River, New Jersey, 2001.

[14] Beaufays, F., "Transform Domain Adaptive Filters: An Analytical Approach", IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-42, pp. 422-431, Feb. 1995.

[15] M. de Courville and P. Duhamel, "Adaptive Filtering in Subbands Using A Weighted Criterion", IEEE Trans. Signal Processing, SP-46, no. 9, pp. 2359-2371, Sept. 1998.

[16] Qureshi, S.U.H., "Adaptive Equalization", in Advanced Digital Communication Systems and Signal Processing Techniques, ed. K. Feher, Prentice Hall, Englewood Cliffs, New Jersey, 1987.

[17] Treichler, J.R., I. Fijalkow, and C.R. Johnson, "Fractionally Spaced Equalizers: How Long Should They Really Be?", IEEE Signal Processing Magazine, vol. 13, pp. 65-81, May 1996.

[18] Qureshi, S.U.H. and G.D. Forney, "Performance and Properties of a T/2 Equalizer", NTC '77, Proceedings.

[19] Vrcelj, B. and P.P. Vaidyanathan, "Theory of MIMO Biorthogonal Partners and Their Application in Channel Equalization", Proceedings of the ICC, Helsinki, June 2001.

[20] Vaidyanathan, P.P. and B. Vrcelj, "Biorthogonal Partners and Applications", IEEE Trans. Signal Proc., May 2001.