2020 59th IEEE Conference on Decision and Control (CDC)
Jeju Island, Republic of Korea, December 14-18, 2020

Nonlinear Matched Filters via Chen–Fliess Series

W. Steven Gray[†]

Abstract— **A class of nonlinear matched filters is introduced suitable for detection problems using Chen–Fliess functional series. Such series can be viewed as a noncommutative analogue of Taylor series. They are written in terms of a weighted sum of iterated integrals indexed by words over a noncommuting set of symbols. The primary goal is to identify within this class the set of filters which maximizes the signal-to-noise ratio at a given time instant in order to provide a statistic for detecting a known signal.**

I. Introduction

Detection of a known signal in additive noise is a fundamental problem in communications, radar, biomedical applications, and most recently in gravitational-wave astronomy [2]. For wide-sense stationary noise, the optimal linear detector has an impulse response satisfying a certain Fredholm type integral equation so as to maximize the signal-to-noise ratio (SNR) at a given time instant. This provides a test statistic suitable for hypothesis testing. For the case of additive white noise, the impulse response *matches* in some sense the signal to be detected [34], [36]. The solution is independent of the noise distribution, but in the Gaussian case it coincides with the correlation receiver derived from likelihood functions, which is known to be the optimal detector over the class of all linear and nonlinear filters. When the assumption of linearity is relaxed, there is the potential for improved detection performance in the non-Gaussian case, but the complexity of the filter becomes an issue. A number of authors have considered using Volterra series to develop filters [5]–[7], [20], [27], [30]–[32]. But determining the optimal kernel functions beyond the second-order case is a largely intractable problem. Others have taken numerical approaches to provide detectors that are locally optimal in some information theoretic sense [10], [25], [26] or approach the optimal filter asymptotically [3], [4]. Despite an extensive literature on the subject, the general problem appears far from settled.

The objective of this paper is to propose a matched filter approach to the nonlinear detection problem using Chen–Fliess series. Such series can be viewed as a noncommutative analogue of Taylor series for functional maps. They are written in terms of a weighted sum of iterated integrals indexed by words over a noncommuting set of symbols [11], [12]. Chen–Fliess series are widely used in control theory to describe the input-output map of a dynamical system with real analytic vector fields and an affine dependence on the input [19], [28]. They contain as a special case all Volterra series with real analytic kernel functions. Therefore, any linear or bilinear time-invariant system can be written in terms of a Chen–Fliess series. It should also be noted that these series are distinct from the eigenfunction series used to provide formal solutions to integral equations [18], [36]. The specific goal here is to identify within the class of all input-output maps having Chen–Fliess series representations, the set of filters which maximizes the SNR at a given time instant in order to provide a statistic for detecting a known signal. This does *not* automatically imply that any corresponding detector will necessarily be optimal in any sense, however, the intuitive appeal is obvious. Intrinsically this approach is for the continuous-time case, but the proposed filter can be approximated and implemented in discrete-time to any desired accuracy modulo some numerical limitations. Finally, it should be stated that the approach taken here has some aspects in common with a class of pattern classification techniques based on the *signature* of a path [17], [23], [24]. Analogous to Chen–Fliess series, the signature is written in terms of iterated integrals of the coordinates functions of the path. However, a path is translation invariant and involves a normalization of its *speed*, features not available in the present context. In addition, the concept of SNR does not play a central role in typical applications like text recognition [37].

The paper is organized as follows. In the next section, the mathematical preliminaries are presented in a concise manner to make the presentation self-contained. The main results are given in Section III, and a discrete-time implementation of the proposed filter is developed in the subsequent section. The conclusions and directions for future results are summarized in the final section.

II. Preliminaries

Let $X = \{x_0, x_1, \ldots, x_m\}$ be a nonempty finite set of noncommuting symbols. X_0 corresponds to the case where $m = 0$. Each element of X is called a *letter*, and any finite sequence of letters from X, $\eta = x_{i_1} \cdots x_{i_k}$, is called a *word* over X. Its *length* is $|\eta| = k$. In particular, $|\eta|_{x_i}$ is the number of times the letter $x_i \in X$ appears in η. The set of all words of length k is denoted by X^k, while the set of all words including the empty word, $\emptyset$, is written as X^*. The latter forms a monoid under catenation. The set of all words with prefix or suffix $\eta \in X^*$ is written as ηX^* and $X^*\eta$, respectively. Any mapping $c : X^* \to \mathbb{R}^\ell$ is called a *formal power series*. The value of c at $\eta \in X^*$ is written as (c, η) and called the *coefficient* of η in c. Typically, c is represented as the formal sum $c = \sum_{\eta \in X^*}(c, \eta)\eta$. The subset of X^* defined by $\mathrm{supp}(c) = \{\eta : (c, \eta) \neq 0\}$ is called the *support* of c. The set of all formal power series over X taking coefficients in $\mathbb{R}^\ell$ is denoted by $\mathbb{R}^\ell\langle\langle X\rangle\rangle$.

[†]Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, Virginia 23529 USA.

A. Chen–Fliess series: continuous-time case

Given a $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$ one can associate a causal m-input, ℓ-output system, F_c, in the following manner. Let $\mathfrak{p} \geq 1$ and $t_0 < t_1$ be given. For a Lebesgue measurable function $u : [t_0, t_1] \to \mathbb{R}^m$, define $\|u\|_\mathfrak{p} = \max\{\|u_i\|_\mathfrak{p} : 1 \leq i \leq m\}$, where $\|u_i\|_\mathfrak{p}$ is the usual $L_\mathfrak{p}$-norm for a measurable real-valued function, u_i, defined on $[t_0, t_1]$. Let $L^m_\mathfrak{p}[t_0, t_1]$ denote the set of all measurable functions defined on $[t_0, t_1]$ having a finite $\|\cdot\|_\mathfrak{p}$ norm and $B^m_\mathfrak{p}(R_u)[t_0, t_1] := \{u \in L^m_\mathfrak{p}[t_0, t_1] : \|u\|_\mathfrak{p} \leq R_u\}$. Assume $C[t_0, t_1]$ is the subset of continuous functions in $L^m_1[t_0, t_1]$. Define inductively for each $\eta \in X^*$ the map $E_\eta : L^m_1[t_0, t_1] \to C[t_0, t_1]$ by setting $E_\emptyset[u] = 1$ and letting

$$E_{x_i\bar{\eta}}[u](t, t_0) = \int_{t_0}^{t} u_i(\tau) E_{\bar{\eta}}[u](\tau, t_0)\, d\tau,$$

where $x_i \in X$, $\bar{\eta} \in X^*$, and $u_0 = 1$. The *Chen–Fliess series* corresponding to c is

$$F_c[u](t) = \sum_{\eta \in X^*} (c, \eta)\, E_\eta[u](t, t_0). \tag{1}$$

It can be shown that if there exists $K, M \geq 0$ such that

$$|(c, \eta)| \leq K M^{|\eta|}\, |\eta|!, \quad \forall \eta \in X^*, \tag{2}$$

then the series defining F_c converges absolutely and uniformly for sufficient small $R, T > 0$ and constitutes a well defined *Fliess operator* mapping $B^m_\mathfrak{p}(R)[t_0,\ t_0 + T]$ into $B^\ell_\mathfrak{q}(S)[t_0,\ t_0 + T]$, where the numbers $\mathfrak{p}, \mathfrak{q} \in [1, \infty]$ are conjugate exponents, i.e., $1/\mathfrak{p} + 1/\mathfrak{q} = 1$ [16]. Furthermore, the input-output map $y = F_c[u]$ can often be realized by an input affine analytic state space model

$$\dot{z} = f(z) + g(z)u, \quad z(0) = z_0$$
$$y = h(z)$$

in local coordinates about z_0 such that for any word $\eta = x_{i_k} \cdots x_{i_1} \in X^*$

$$(c, \eta) = L_{g_{i_1}} \cdots L_{g_{i_k}} h(z_0),$$

where $L_{g_i} h$ denotes the Lie derivative of h with respect to vector field g_i [11], [12], [19], [28]. The concept of a Fliess operator is also well defined when the applied input is an L_2-Itô stochastic process [9] or a Poisson process [8].

B. Chen–Fliess series: discrete-time case

A discrete-time version of a Chen–Fliess series is also available. Here inputs are assumed to be sequences of vectors from the normed linear space

$$l^{m+1}_\infty(N_0) := \{\hat{u} = (\hat{u}(N_0), \hat{u}(N_0 + 1), \ldots) : \|\hat{u}\|_\infty < \infty\},$$

where $\hat{u}(N) := [\hat{u}_0(N), \hat{u}_1(N), \ldots, \hat{u}_m(N)]^T$, $N \geq N_0$ with $\hat{u}_i(N) \in \mathbb{R}$, $|\hat{u}(N)| := \max_{i \in \{0,1,\ldots,m\}} |\hat{u}_i(N)|$, and $\|\hat{u}\|_\infty := \sup_{N \geq N_0} |\hat{u}(N)|$. The subspace of finite sequences over $[N_0, N_f]$ is denoted by $l^{m+1}_\infty[N_0, N_f]$. Given a generating series $c \in \mathbb{R}^\ell\langle\langle X\rangle\rangle$, the corresponding *discrete-time Chen–Fliess series* is defined as

$$\hat{F}_c[\hat{u}](N) = \sum_{\eta \in X^*} (c, \eta) S_\eta[\hat{u}](N)$$

for any $N \geq N_0$, where

$$S_{x_i\eta}[\hat{u}](N) = \sum_{k=N_0}^{N} \hat{u}_i(k) S_\eta[\hat{u}](k)$$

with $x_i \in X$, $\eta \in X^*$, and $\hat{u} \in l^{m+1}_\infty[N_0]$. By assumption, $S_\emptyset[\hat{u}](N) := 1$. It is known that the class of truncated, discrete-time Fliess operators acts as a set of universal approximators with computable error bounds for their continuous-time counterparts [13]. Specifically, select some fixed $u \in L^m_1[0, T]$ with $T > 0$ finite. Choose an integer $L \geq 1$, let $\Delta := T/L$, and define the sequence of real numbers

$$\hat{u}_i(N) = \int_{(N-1)\Delta}^{N\Delta} u_i(t)\, dt, \quad i = 0, 1, \ldots, m, \tag{3}$$

where $N \in \{1, 2, \ldots, L\}$. Assume $u_0 = 1$ so that $\hat{u}_0(N) = \Delta$. The truncated, discrete-time Fliess operator with generating series $c \in \mathbb{R}\langle\langle X\rangle\rangle$ is

$$\hat{y}(N) = \hat{F}^J_c[\hat{u}](N) := \sum_{\eta \in X^{\leq J}} (c, \eta) S_\eta[\hat{u}](N), \tag{4}$$

where $X^{\leq J} := \cup_{k=0}^J X^k$ and $J, N \geq 1$. In this setting, an explicit and generally tight upper bound on $\left|F_c[u](T) - \hat{F}^J_c[\hat{u}](L)\right|$ is given in [13, Theorems 6 and 7] in terms of J, L, and the series growth rate parameters K and M appearing in (2). Finally, there are at least two methods for realizing the input-output map $\hat{y} = \hat{F}_c[\hat{u}]$. The first applies only for rational generating series and employs state space models [13]. The second approach is based on Chen series and is completely general [14], [15], [35]. In this paper, the latter approach is taken in all the software used for the simulations. Specifically, (4) is first rewritten as an inner product $\hat{y}(N) = \mathbf{c}^T \mathbb{S}[\hat{u}](N)$, where $\mathbf{c} \in \mathbb{R}^l$ is a column vector containing all the coefficients of c in some fixed order $\{\eta_1, \eta_2, \ldots, \eta_l\}$, $l := \mathrm{card}(X^{\leq J})$, and $\mathbb{S}[\hat{u}](N) \in \mathbb{R}^l$ is the corresponding set of iterated sums. It is then shown in [15], [35] that the update equation for $\mathbb{S}[\hat{u}](N)$ has the form

$$\mathbb{S}[\hat{u}](N + 1) = \mathcal{S}^J(N + 1)\mathbb{S}[\hat{u}](N), \quad N \geq 0,$$

where $\mathbb{S}[\hat{u}](0) := e_1 = [1\, 0 \cdots 0]^T \in \mathbb{R}^l$, and $\mathcal{S}^J(N + 1)$ has a certain recursive structure given by $\mathcal{S}^0(N + 1) = 1$, and for $J \geq 0$

$$\mathcal{S}^{J+1}(N + 1) =$$
$$\left[\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline \hat{u}(N + 1) \otimes (\mathcal{S}^J(N + 1)e_1) & \begin{array}{c}\text{block diag}(\mathcal{S}^J(N + 1), \\ \ldots, \mathcal{S}^J(N + 1))\end{array} \end{array}\right],$$

where '$\otimes$' denotes the Kronecker matrix product, and the block diagonal matrix is comprised on $m + 1$ blocks. Once J is fixed, the structure of $\mathcal{S}^J(N+1)$, which depends only on $\hat{u}(N+1)$, is also fixed and can be pre-computed to efficiently implement the update equation.

III. Nonlinear Matched Filter

Set $X = \{x_0, x_1\}$ and pick a total ordering on X^* so that

$$X^* = \left[\begin{array}{ccc} \eta_1 & \eta_2 & \cdots \end{array}\right]^T.$$

TABLE I

FILTER TYPE BASED ON SUPPORT OF GENERATING SERIES

W	filter type
$X^{\leq J}=\cup_{k=0}^{J}X^k$	finite approximation
$X_H^*=\{\eta\in X^*: \lvert\eta\rvert_{x_1}>0\}$	homogeneous
$X_H^{\leq J}=X^{\leq J}\cap X_H^*$	finite approximation, homogenous
$X_L=X_0^*x_1X_0^*$	linear
$X_{LTI}=X_0^*x_1$	linear time-invariant
$\{x_1\}$	single integrator/summer

For any $c\in\mathbb{R}\langle\langle X\rangle\rangle$ and $u\in L_1[0,T]$ define

$$\mathbf{c}=\left[\begin{array}{ccc}(c,\eta_1) & (c,\eta_2) & \cdots\end{array}\right]^T.$$
$$\mathbf{E}[u](t)=\left[\begin{array}{ccc}E_{\eta_1}[u](t,t_0) & E_{\eta_2}[u](t,t_0) & \cdots\end{array}\right]^T.$$

In which case, (1) can be written as

$$F_c[u](t)=\mathbf{c}^T\mathbf{E}[u](t).$$

It will be useful to consider special classes of Fliess operators corresponding to certain filter types. For example, if it is desirable that F_c be homogeneous, i.e., $F_c[0]=0$, then the support of c should contain no words of the form x_0^j since $E_{x_0^j}[0](t)=t^j/j!$ for $t\in[0,T]$. In which case, X^* everywhere above can be replaced with the set of words $X_H^*:=\{\eta\in X^*:|\eta|_{x_1}>0\}$, and the homogeneity property is assured. Other important classes of operators and their corresponding word sets are given in Table I.

Let $T>0$ and assume that $s\in L_1[0,T]$ is a known signal. Let $W\subseteq X^*$ be an arbitrary fixed word set. Suppose $\boldsymbol{n}(t)$ is a zero mean random process with $E_{\eta_i}[s+\boldsymbol{n}](T)$ being a well defined random variable for each $\eta_i\in W$. Let $y=F_c[u]$ be any operator with $\mathrm{supp}(c)\subseteq W$. Define $y_s=F_c[s]$ and $\boldsymbol{y}_{s+n}=F_c[s+\boldsymbol{n}]$. The input and output (power) signal-to-noise ratios at $t=T$ are taken, respectively, to be

$$\mathrm{SNR}_i^2=\frac{s^2(T)}{\mathcal{E}\{\boldsymbol{n}^2(T)\}}$$
$$\mathrm{SNR}_o^2=\frac{y_s^2(T)}{\mathcal{E}\{(\boldsymbol{y}_{s+n}(T)-y_s(T))^2\}},$$

where $\mathcal{E}\{\cdot\}$ denotes expected value. Any filter F_c that maximizes SNR_o^2 is said to be *matched* to s. The following theorem describes a matched filter in this continuous-time setting.

Theorem 1: The maximum value of SNR_o^2 is achieved when $\mathbf{c}=R^{-1}\mathbf{E}[s](T)$ and is equivalent to

$$\mathrm{SNR}_o^2=\mathbf{E}[s](T)^TR^{-1}\mathbf{E}[s](T)=\mathbf{c}^TR\mathbf{c},$$

where R is assumed to be a positive definite autocorrelation matrix with components

$$R_{ij}=\mathcal{E}\{(E_{\eta_i}[s+\boldsymbol{n}](T)-E_{\eta_i}[s](T))\\ (E_{\eta_j}[s+\boldsymbol{n}](T)-E_{\eta_j}[s](T))\}.$$

Proof: Define the mean-square performance index

$$\mathrm{J}=\mathcal{E}\{(\boldsymbol{y}_{s+n}(T)-y_s(T))^2\}=\mathbf{c}^TR\mathbf{c}.$$

Then SNR_o^2 is maximum when J is minimized subject to a fixed value of $y_s(T)$. That is, the problem reduces to a constrained quadratic optimization problem with Lagrangian

$$\mathrm{L}=\mathrm{J}-\mu y_s(T)=\mathbf{c}^TR\mathbf{c}-\mu\mathbf{c}^T\mathbf{E}[s](T),$$

where μ is the Lagrange multiplier. A necessary and sufficient condition for a minimum is:

$$\frac{d\mathrm{L}}{d\mathbf{c}}=2R\mathbf{c}-\mu\mathbf{E}[s](T)=0$$
$$\frac{d^2\mathrm{L}}{d\mathbf{c}^2}=2R>0.$$

The optimal filter follows from the first equation,

$$\mathbf{c}=(\mu/2)R^{-1}\mathbf{E}(s)(T),$$

so that

$$y_s(T)=\mathbf{c}^T\mathbf{E}(s)(T)=(\mu/2)\mathbf{E}[s](T)^TR^{-1}\mathbf{E}[s](T)$$

and

$$\mathrm{J}=\mathbf{c}^TR\mathbf{c}=(\mu/2)^2\mathbf{E}[s](T)^TR^{-1}\mathbf{E}[s](T).$$

Therefore,

$$\mathrm{SNR}_o^2=\frac{y_s^2(T)}{\mathrm{J}}=\mathbf{E}[s](T)^TR^{-1}\mathbf{E}[s](T)$$

as claimed. Note that the factor $\mu/2$ above can be dropped without affecting SNR_o^2. ■

It should be stated first that the performance of the matched filter will be highly dependent on the specific class of filters considered, i.e., on W. In fact, it is not immediately evident that for every possible filter class one has $\mathrm{SNR}_o^2>\mathrm{SNR}_i^2$ as desired. On the other hand, as in the classical case [36], the theorem holds for all noise distributions. Of course the maximum value of SNR_o^2 will be distribution dependent. In the special case of linear filters ($W=X_L$), the performance index simplifies by superposition to $\mathrm{J}=\mathcal{E}\{\boldsymbol{y}_n^2(T)\}$. If one is free to *design* s then the following corollary is useful.

Corollary 1: Consider the class of linear filters $\{F_c:\mathrm{supp}(c)\subset X_L\}$, and assume R is fixed. The maximum value of SNR_o^2 is achieved for the matched filter where $s\in L_1[0,T]$ satisfies $\mathbf{E}[s](T)=p$ with p the eigenvector corresponding to the smallest eigenvalue of R^{-1}. Specifically, if $\mathbf{c}=\lambda_{min}(R)p$ then

$$\mathrm{SNR}_{o,max}^2=\lambda_{min}^{-1}(R)\|\mathbf{E}[s](T)\|^2$$

and

$$\sup_{s'\in L_1[0,T]}\mathbf{E}[s'](T)^TR^{-1}\mathbf{E}[s'](T)\leq\mathrm{SNR}_{o,max}^2.$$

Proof: The proof follows directly from Theorem 1 and standard results regarding quadratic forms. ■

No claim is made above that for a given R such an s actually exists. In addition, as the dimension of R grows, it is likely to become ill-conditioned so that this upper bound may not be achievable in practice.

Example 1: Assume $W=X^*$. If $R=(N_0/2)I$, then

$$\mathbf{c}=(2/N_0)\mathbf{E}[s](T)$$

and

$$\mathrm{SNR}_o^2 = (2/N_0)\mathbf{E}[s](T)^T\mathbf{E}[s](T).$$

In this case, the coefficients of the filter match the iterated integrals of s modulo a constant. □

Example 2: Consider a causal linear time-invariant (LTI) filter with impulse response h. If h is assumed to be real analytic at $t = 0$ then

$$\begin{aligned} y_s(T) &= \int_0^T h(T-t)s(t)\,dt \\ &= \sum_{i=0}^{\infty} h_i \int_0^T \frac{(T-t)^i}{i!} s(t)\,dt \\ &= \sum_{i=0}^{\infty} h_i E_{x_0^i x_1}[s](t). \end{aligned}$$

(The final step requires that the integration by parts formula be applied i times.) Therefore, the filter can be written in terms of a Fliess operator, F_c, where the generating series is $c = \sum_{i\geq 0} h_i x_0^i x_1$. That is, W is equal to the set of *linear time-invariant words*

$$X_{LTI} = \left[\begin{array}{cccc} x_1 & x_0x_1 & x_0^2x_1 & \cdots \end{array}\right]^T.$$

For this case, the coefficients $h_i = (c, x_0^i x_1)$ for the optimal filter are given by Theorem 1 to be $\mathbf{c} = R^{-1}\mathbf{E}[s](T)$, where now

$$\begin{aligned} R_{ij} &= \mathcal{E}\{E_{x_0^i x_1}[\boldsymbol{n}](T)E_{x_0^j x_1}[\boldsymbol{n}](T)\} \\ &= \int_0^T \int_0^T \frac{(T-t)^i}{i!}\frac{(T-s)^j}{j!} R_n(t,s)\,dt\,ds \end{aligned}$$

with $R_n(t,s)$ being the autocorrelation function for $\boldsymbol{n}$. In the white noise case where $R_n(t,s) = (N_0/2)\delta(t-s)$ (and thus, $\mathrm{SNR}_i^2 = 0$), it follows directly that

$$R_{ij} = \frac{N_0}{2}\binom{i}{j}\frac{T^{i+j+1}}{(i+j+1)!}.$$

This implies that the filter coefficients have to satisfy

$$E_{x_0^j x_1}[s](T) = [R\mathbf{c}]_{j+1} = \sum_{i=0}^{\infty} h_i \frac{N_0}{2}\frac{T^{i+j+1}}{i!j!(i+j+1)}, \quad j \geq 0.$$

Recall that the classical theory of matched filters for LTI systems requires that $s(t) = (N_0/2)h(T-t)$ [34], [36]. Therefore,

$$\begin{aligned} E_{x_0^j x_1}[s](T) &= \int_0^T \frac{(T-\tau)^j}{j!}\frac{N_0}{2}h(T-\tau)\,d\tau \\ &= \sum_{i=0}^{\infty} h_i \frac{N_0}{2}\int_0^T \frac{\tau^{i+j}}{i!\,j!}\,d\tau \\ &= \sum_{i=0}^{\infty} h_i \frac{N_0}{2}\frac{T^{i+j+1}}{i!j!(i+j+1)} \end{aligned}$$

as expected. In addition,

$$\begin{aligned} \mathrm{SNR}_o^2 = \mathbf{c}^T R\mathbf{c} &= \sum_{i,j=0}^{\infty} \frac{h_i}{i!}\frac{h_j}{j!}\frac{N_0}{2}\frac{T^{i+i+1}}{i+j+1} \\ &= \frac{N_0}{2}\int_0^T h^2(t)\,dt = \frac{2}{N_0}\int_0^T s^2(t)\,dt. \end{aligned}$$

□

Example 3: For wide-sense stationary possibly nonwhite noise, it is known the LTI filter maximizing SNR_o^2 is the solution of the integral equation

$$s(T-t) = \int_0^T h(\tau)R_n(t-\tau)\,d\tau, \quad 0 \leq t \leq T$$

[36]. If $s \in L_2[0,T]$ and $\{f_j\}_{j\geq 0}$ is a complete orthonormal spanning set for $L_2[0,T]$ with

$$\int_0^T R_n(t-\tau)f_j(\tau)\,d\tau = \lambda_j f_j(t), \quad j \geq 0,$$

then the optimal filter has the form

$$h(t) = \sum_{j=0}^{\infty} \frac{s_j}{\lambda_j} f_j(t),$$

where $s_j := \int_0^T s(T-t)f_j^*(t)\,dt$ and provided that $\sum_{j\geq 0}(|s_j|/\lambda_j)^2 < \infty$. Take as an example the case of band-limited white noise with spectral density $S(\omega) = N_0/2$ on $[-\Omega,\Omega]$ and zero otherwise. The corresponding autocorrelation function

$$R_n(\tau) = \frac{N_0}{2}\frac{\sin(\Omega\tau)}{\pi\tau}, \tag{5}$$

has real-valued eigenfunctions $f_j(t) = \psi_j(t - T/2)$, where the ψ_j are scaled versions of angular prolate spheroidal wave functions, and positive real eigenvalues [21], [22], [33]. Computing h even for simple s is a nontrivial problem [1], [29]. If the filter class is simplified, however, the problem is more tractable as illustrated next.

Consider the optimal filter corresponding to $W = \{x_1\}$ in Theorem 1. Assume $s(t) = A > 0$, and $\boldsymbol{n}$ has the autocorrelation function (5). The optimal filter coefficient is given by $(c, x_1) = R^{-1}E_{x_1}[s](T) = R^{-1}AT$, where

$$\begin{aligned} R &= \mathcal{E}\{E_{x_1}^2[\boldsymbol{n}](T)\} = \int_0^T \int_0^T R_n(t-s)\,dt\,ds \\ &= 2T\int_0^T R_n(\tau)\,d\tau - 2\int_0^T \tau R_n(\tau)\,d\tau \\ &= N_0T\int_0^T \frac{\sin(\Omega\tau)}{\pi\tau}\,d\tau - \frac{N_0}{\pi}\int_0^T \sin(\Omega\tau)\,d\tau \\ &= N_0T\int_0^{\frac{\Omega T}{\pi}} \frac{\sin(\pi\tau)}{\pi\tau}\,d\tau + \frac{N_0}{\pi\Omega}(\cos(\Omega T) - 1) \\ &\approx \frac{N_0T}{2} \end{aligned}$$

when $\Omega T = 2\pi k$ and $k \gg 0$. Hence, the optimal filter is

$$\begin{aligned} \boldsymbol{y}(t) &= (c, x_1)E_{x_1}[s+\boldsymbol{n}](t) \\ &= \frac{2A}{N_0}\int_0^t (s+\boldsymbol{n})(\tau)\,d\tau \end{aligned}$$

for which

$$\mathrm{SNR}_o^2 = \frac{E_{x_1}^2[s](T)}{R} \approx \frac{2}{N_0}A^2T$$

$$\approx \frac{\Omega T}{\pi}\,\mathrm{SNR}_i^2. \tag{6}$$

Therefore, $\mathrm{SNR}_o^2 \gg \mathrm{SNR}_i^2$. $\square$

Nonlinear filter examples are difficult to produce analytically as the statistic R is hard to compute directly. But the filter design problem is tractable numerically using the discrete-time version of the theory as described in the next section.

IV. Discrete-Time Implementation

Let $L > 0$ be a fixed integer and assume that $\hat{s} \in l_\infty^2[0, L]$ is a known signal. Let $W \subseteq X^*$ be an arbitrary fixed word set. Suppose $\hat{\boldsymbol{n}}(N)$ is a zero mean discrete-time random process with $S_{\eta_i}[\hat{s} + \hat{\boldsymbol{n}}](L)$ a well defined random variable for each $\eta_i \in W$. Let

$$\hat{y}(N) = \hat{F}_c[\hat{u}](N) = \mathbf{c}^T \mathbb{S}[\hat{u}](N)$$

with $\mathrm{supp}(c) \subseteq W$ and

$$\mathbb{S}[\hat{u}](N) = \begin{bmatrix} S_{\eta_1}[\hat{u}](N) & S_{\eta_2}[\hat{u}](N) & \cdots \end{bmatrix}^T.$$

Define $\hat{y}_{\hat{s}} = \hat{F}_c[\hat{s}]$ and $\hat{\boldsymbol{y}}_{\hat{s}+\hat{n}} = \hat{F}_c[\hat{s} + \hat{\boldsymbol{n}}]$. The input and output (power) signal-to-noise ratios at $N = L$ are taken, respectively, to be

$$\mathrm{SNR}_i^2 = \frac{\hat{s}^2(L)}{\mathcal{E}\{\hat{\boldsymbol{n}}^2(L)\}}$$
$$\mathrm{SNR}_o^2 = \frac{\hat{y}_{\hat{s}}^2(L)}{\mathcal{E}\{(\hat{\boldsymbol{y}}_{\hat{s}+\hat{n}}(L) - \hat{y}_{\hat{s}}(L))^2\}}.$$

The following theorem is the discrete-time version of Theorem 1. Its proof is exactly analogous.

Theorem 2: The maximum value of SNR_o^2 is achieved when $\mathbf{c} = R^{-1}\mathbb{S}[\hat{s}](L)$ and is equivalent to

$$\mathrm{SNR}_o^2 = \mathbb{S}[\hat{s}](L)^T R^{-1} \mathbb{S}[\hat{s}](L) = \mathbf{c}^T R \mathbf{c}, \tag{7}$$

where R is assumed to be a positive definite autocorrelation matrix with components

$$R_{ij} = \mathcal{E}\{(S_{\eta_i}[\hat{s} + \hat{\boldsymbol{n}}](L) - S_{\eta_i}[\hat{s}](L)) \\ (S_{\eta_j}[\hat{s} + \hat{\boldsymbol{n}}](L) - S_{\eta_j}[\hat{s}](L))\}.$$

Example 4: Consider the case where $\hat{s}(N) = A \in \mathbb{R}^+$ for all $N \in \{1, \ldots, L\}$, and $\hat{\boldsymbol{n}}$ is a sequence of zero mean independent Gaussian random variables. The latter can be viewed as samples in the sense of (3) of a band-limited (not necessarily Gaussian) white noise process $\boldsymbol{n}$ on some interval $[0, T]$ as described in Example 3 with $\Delta = T/L$ and $\Omega = \pi/\Delta$. Note the random variable $S_\eta[\hat{s} + \hat{\boldsymbol{n}}](L)$ will not be Gaussian when $|\eta|_{x_1} > 1$. In this example, the filter type will be homogeneous truncated discrete-time Fliess operators corresponding to $W = X_H^{\leq J} = X^{\leq J} \cap X_F^*$, $J = 1, 2, 3$. For the particular case where $J = 1$, the corresponding discrete-time match filter is equivalent to the classical matched filter in white noise since $W = \{x_1\}$, and thus (6) provides a reference point (LTI MF (theory) in Figure 1) against which the nonlinear filters ($J = 2, 3$) can be compared. The statistic R is also estimated by Monte Carlo simulation, and the SNR_o^2 is computed from (7) as shown in Figure 1 when $A = 1$, $T = 1$ and $L = 20$ (LTI MF (MC) in Figure 1).

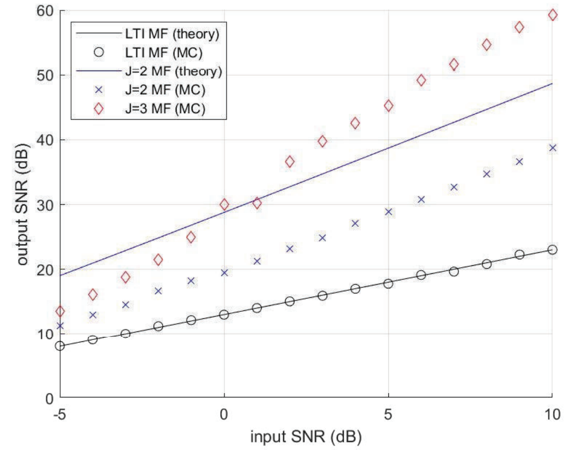

Fig. 1. Input SNR versus output SNR in Example 4

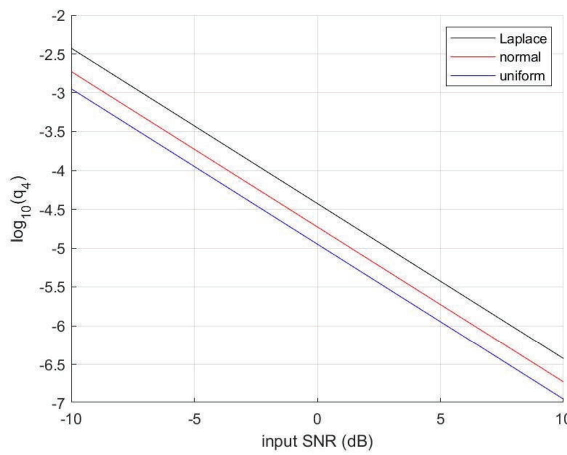

Fig. 2. Statistic $q_4 = \mathcal{E}\{\boldsymbol{n}^4\}$ for three different distributions as a function of input SNR

In the case where $J = 2$, R can be computed directly as shown in (8), where $q_i := \mathcal{E}\{\boldsymbol{n}^i\}$. The rows and columns are indexed by $X_H^{\leq 2} = \{x_1x_0, x_1, x_0x_1, x_1^2\}$. For the case under consideration, $q_3 = 0$ and $q_4 = 3\sigma_n^4$. It is important to note that while R has full rank, $\mathrm{rank}(R) \to 3$ as $\sigma_n \to 0$ since in the limit column 4 = column $2 \times (AT(L+1)/L)$. Therefore, Theorem 2 can only be applied practically by replacing the matrix inverse of R with the pseudo-inverse to solve the problem in a least-squares sense. (These numerical results were also confirmed independently for each R by solving the quadratic optimization problem $\min_{\mathbf{c} \in \mathbb{R}^4} \mathbf{c}^T R \mathbf{c} + \mathbf{c}^T \mathbb{S}[A](L)$ via the MatLab command `quadprog`.) This theoretical value of SNR_o^2 is also shown in Figure 1 ($J = 2$ MF (theory)). Note that this only upper bounds the value found by estimating R via Monte Carlo simulation ($J = 2$ MF (MC)). Computing theoretical results for the $J = 3$ is not feasible, but the outcome is likely the same, that is, the higher-order filter increases SNR_o^2 as shown in Figure 1 ($J = 3$ MF (MC)), but not to the degree predicted by computing R explicitly and applying (7). Finally, it is noted that if the distribution of the noise is changed, the results shown in Figure 1 are altered very little. Observe that the only change

$$R = \begin{bmatrix} T^2\sigma_n^2\frac{(L+1)(2L+1)}{6L} & T\sigma_n^2\frac{(L+1)}{2} & T^2\sigma_n^2\frac{(L+1)(L+2)}{6L} & AT^2\sigma_n^2\frac{(L+1)^2}{2L}+Tq_3\frac{L+1}{2} \\ T\sigma_n^2\frac{L+1}{2} & \sigma_n^2 L & T\sigma_n^2\frac{(L+1)}{2} & AT\sigma_n^2(L+1)+q_3L \\ T^2\sigma_n^2\frac{(L+1)(L+2)}{6L} & T\sigma_n^2\frac{(L+1)}{2} & T^2\sigma_n^2\frac{(L+1)(2L+1)}{6L} & AT^2\sigma_n^2\frac{(L+1)^2}{2L}+Tq_3\frac{L+1}{2} \\ AT^2\sigma_n^2\frac{(L+1)^2}{2L}+Tq_3\frac{L+1}{2} & AT\sigma_n^2(L+1)+q_3L & AT^2\sigma_n^2\frac{(L+1)^2}{2L}+Tq_3\frac{L+1}{2} & A^2T^2\sigma_n^2\frac{(L+1)^2}{L}+2ATq_3(L+1)+q_4L \end{bmatrix} \quad (8)$$

in R as shown in (8) will be q_4. This statistic is shown in Figure 2 for three distributions as a function of SNR_i. These differences are not significant enough to affect SNR_o. □

ACKNOWLEDGEMENT

This research was supported by the National Science Foundation under grant CMMI-1839378.

V. CONCLUSIONS AND FUTURE WORK

A class of nonlinear matched filters was introduced using Chen–Fliess functional series which maximizes the SNR at a given time instant in order to provide a detection test statistic for a known signal. The theory generally predicts higher SNR as the filter order is increased, but these predictions only provide upper bounds for the cases tested due to the numerical rank deficiency of the autocorrelation matrix. Noise with nonsymmetric distributions should be investigated ($q_3 \neq 0$), and the new test statistics should next be evaluated for detection performance.

REFERENCES

[1] R. Adelman, N. A. Gumerov, and R. Duraiswami, Software for computing the spheroidal wave functions using arbitrary precision arithmetic, arXiv:1408.0074v1, 2014.

[2] S. Babak1, R. Balasubramanian, D. Churches, T. Cokelaer, and B. S. Sathyaprakash, A template bank to search for gravitational waves from inspiralling compact binaries: I. Physical models, *Classical Quant. Grav.*, 23 (2006) 5477–5504.

[3] Y. Boers and H. Driessen, A particle-filter-based detection scheme, *IEEE Signal Process. Lett.*, 10 (2003) 300–302.

[4] Y. Boers and P. K. Mandal, Optimal particle-filter-based detector, *IEEE Signal Process. Lett.*, 26 (2019) 435–439.

[5] P. Bondon, P. L. Combettes, and B. Picinbono, Volterra filtering and higher order whiteness, *IEEE Trans. Signal Process.*, 43 (1995) 2209–2212.

[6] M. Bouvet, Expansions of the likelihood ratio and applications *IEEE Trans. Acoust., Speech, Signal Process*, ASSP-34 (1986) 653–660.

[7] P. Chevalier, P. Duvaut, and B. Picinbono, Complex transversal Volterra filters optimal for detection and estimation, *Proc. 1991 Inter. Conf. Acoustics, Speech, Signal Process.*, Toronto, Canada, 1991, pp. 3537–3540.

[8] L. A. Duffaut Espinosa and W. S. Gray, Functional series expansions for continuous-time switched systems, *J. Dyn. Control Syst.*, 21 (2015) 211–237.

[9] L. A. Duffaut Espinosa, W. S. Gray, and O. R. González, On Fliess operators driven by L2-Itô processes, *Stochastics*, 84 (2012) 507–532.

[10] D. Erdogmus, R. Agrawal, and J, C. Principe, A mutual information extension to the matched filter, *Signal Process.*, 85 (2005) 927–935.

[11] M. Fliess, Fonctionnelles causales non linéaires et indéterminées non commutatives, *Bull. Soc. Math. France*, 109 (1981) 3–40.

[12] M. Fliess, Réalisation locale des systèmes non linéaires, algèbres de Lie filtrées transitives et séries génératrices non commutatives, *Invent. Math.*, 71 (1983) 521–537.

[13] W. S. Gray, L. A. Duffaut Espinosa, and K. Ebrahimi-Fard, Discrete-time approximations of Fliess operators, *Numer. Math.*, 137 (2017) 35–62, arXiv:1510.07901.

[14] W. S. Gray, G. S. Venkatesh, and L. A. Duffaut Espinosa, Discrete-time Chen series for time discretization and machine learning, *Proc. 53rd Conf. on Information Sciences and Systems*, Baltimore, Maryland, 2019.

[15] W. S. Gray, G. S. Venkatesh, and L. A. Duffaut Espinosa, Nonlinear system identification for multivariable control via discrete-time Chen-Fliess series, *Automatica*, 119 (2020), article 109085.

[16] W. S. Gray and Y. Wang, Fliess operators on L_p spaces: Convergence and continuity, *Systems Control Lett.*, 46 (2002) 67–74.

[17] B. Hambly and T. Lyons, Uniqueness for the signature of a path of bounded variation and the reduced path group, *Ann. Math.*, 171 (2010) 109–167.

[18] C. W. Helstrom, Solution of the detection integral equation for stationary filtered white noise, *IEEE Trans. Inf. Theory*, 11 (1965) 335–339.

[19] A. Isidori, *Nonlinear Control Systems*, 3rd Ed., Springer, London, 1995.

[20] R. J. Kenefic and D. D. Weiner, Application of the Volterra funtional expansion in the detection of nonlinear functions of Gaussian processes, *IEEE Trans. Commun.*, 33 (1985) 276–279.

[21] H. J. Landau and H. O. Pollak, Prolate spheriodal wave functions, Fourier analysis and uncertainty - II, *Bell Sys. Tech. J.*, 40 (1961) 65–84.

[22] H. J. Landau and H. O. Pollak, Prolate spheriodal wave functions, Fourier analysis and uncertainty - III: The dimension of the space of essentially time- and band-limited signals, *Bell Sys. Tech. J.*, 41 (1962) 1295–1336.

[23] T. Lyons, Rough paths, signatures and the modelling of functions on streams, arXiv:1405.4537, 2014.

[24] T. Lyons and Z. Qian, *System Control and Rough Paths*, Clarendon Press, Oxford, U.K., 2002.

[25] J. H. Miller and J. B. Thomas, Detectors for discrete-time signals in non-Gaussian noise, *IEEE Trans. Inf. Theory*, 18 (1972) 241–250.

[26] J. H. Miller and J. B. Thomas, Robust detectors for signals in non-Gaussian noise, *IEEE Trans. Commun.*, 25 (1977) 686–690.

[27] I. J. Morrison, The Application of Volterra Series to Signal Detection and Estimation, doctoral dissertation, University of Cambridge, 1990.

[28] H. Nijmeijer and A. J. van der Schaft, *Nonlinear Dynamical Control Systems*, Springer, New York, 1990.

[29] A. Osipov, V. Rokhlin, and H. Xiao *Prolate Spheroidal Wave Functions of Order Zero*, Springer, New York, 2013.

[30] B. Picinbono, Higher-order statistical signal processing with Volterra filters, *Proc. Workshop on Higher-Order Spectral Analysis*, Vail, CO, 1989, pp. 62–67.

[31] B. Picinbono, Generalized matched filters, *Proc. Fifth ASSP Workshop on Spectrum Estimation and Modeling*, Rochester, NY, 1990, pp. 182–186.

[32] B. Picinbono and P. Duvaut, Geometrical properties of optimal Volterra filters for signal detection, *IEEE Trans. Inf. Theory*, 36 (1990) 1061–1068.

[33] D. Slepian and H. O. Pollak, Prolate spheriodal wave functions, Fourier analysis and uncertainty - I, *Bell Sys. Tech. J.*, 40 (1961) 43–64.

[34] G. L. Turin, An introduction fo matched filters, *IRE Trans. Inf. Theory*, 6 (1960) 311–329.

[35] G. S. Venkatesh, W. S. Gray, and L. A. Duffaut Espinosa, Combining learning and model based multivariable Control, *Proc. 58th IEEE Conf. on Decision and Control*, Nice, France, 2019, pp. 1013–1018.

[36] A. D. Whalen, *Detection of Signals in Noise*, Academic Press, San Diego, 1971.

[37] Z. Xie, Z. Sun, L. Jin, H. Ni, and T. Lyons, Learning spatial-semantic context with fully convolutional recurrent network for online handwritten Chinese text recognition, *IEEE Trans. Pattern Anal. Mach. Intell.*, 40 (2018) 1903–1917.