Overview:

- General setup as before: $s(t) \rightarrow F \rightarrow r(t)$
- We want a principled way to describe $F$, but without assuming linearity.
- A more concise description then supplies a list of $(s, r)$-pairs.
- A description and suggests (or tests) ideals for the internals of $F$.

In Linear Systems Theory, we assume that if $F(s_1) = r_1, F(s_2) = r_2$, then

$$F(s_1 + s_2) = r_1 + r_2$$

and

$$F(c s) = c r.$$ 

This allows us to use the vector space structure on $V$, the space of all signals $s$, namely, $F$ is in $\text{Hom}(V, V)$.

We then make use of time-translation invariance:

- If $s' = D\tau s$, i.e., $s'(t) = s(t + \tau)$, then

$$D\tau F = F D\tau.$$ 

This implies that $F$ is diagonal in the Fourier basis of $S$,

$$F(e^{i\omega t}) = F(\omega) e^{i\omega t}.$$ 

We're dispensing with linearity of $F$—but we still have time translation invariance.

(And we still have smoothness, boundedness, finite memory, ...).
A "lower bound" for how hard this can be:

Say \( F_0(s) \) depends only on the current value of \( s \).

\[ F_0(s) \] depends only on the current value of \( s \).

\[ F_0(s) ] \] for some s.t.

\[ F_0(s) \] does not depend on \( s \).

These are the "static nonlinearities." At the very least, describing all

\( F \)'s is at least as hard as describing all \( F_0 \)'s.

\[ \begin{aligned}
 r(t) \\
 s(t)
\end{aligned} \]

\[ \begin{aligned}
 f
\end{aligned} \]

We can, for example, describe \( f \) by its Taylor series

\[ f(s) = f_0 + s f_1 + \frac{s^2}{2!} f_2 + \frac{s^3}{3!} f_3 + \cdots \]

where \( f_k = \frac{d^k}{ds^k} f \bigg|_{s=0} \).

Pros: it is universal* and principled (*: for \( f \)'s that have a

Taylor series)

Cons: too hard to measure \( \frac{d^k}{ds^k} f \), because of noise.

* What if \( f \) looks like

\[ f \]

i.e., Taylor expansion might be useful only in a

narrow range

(Taylor series requires that \( f(s) \) be analytic, \( f(0) = 1 \) is not.)
An alternative is to express $f$ as an orthogonal expansion

$$f(s) = \sum_{k=0}^{\infty} a_k \psi_k(s), \text{ where } \psi_k(s) \text{ are orthogonal in the sense that } \int_{-\infty}^{\infty} \psi_k(s) \psi_l(s) w(s) ds = \delta_{kl}$$

for some $w(s) > 0$.

Then $$a_k = \frac{1}{\delta_{kk}} \int_{-\infty}^{\infty} f(s) \psi_k(s) w(s) ds.$$

Typical example: $w(s) = \frac{1}{\sqrt{2\pi}} e^{-s^2/2}$ (6.5557).

The $\psi$'s are the Hermite polynomials (s=4):

$$\begin{align*}
\psi_0(s) &= 1, \\
\psi_1(s) &= s, \\
\psi_2(s) &= s^2 - 1, \\
\psi_3(s) &= s^3 - 3s, \\
\psi_4(s) &= s^4 - 6s^2 + 3, \\
&\vdots
\end{align*}$$

A truncated approximation $\sum_{k=0}^{M} a_k \psi_k(s)$ is the best approximation to $f(s)$ among all $M$-th order polynomials, in the sense that

$$\int_{-\infty}^{\infty} \left( f(s) - \sum_{k=0}^{M} a_k \psi_k(s) \right) w(s) ds$$

is minimized.

This is a reasonable definition of "best" if your inputs are drawn from $w(s)$.
The orthogonal expansion coefficients are more gradual to precise (don't require a limit of $s \to 0$) but they will
either work (i.e., not universal).
The orthogonal expansion does not require the existence of
derivatives of $f$.

How will these ideas (Taylor - orthonal) generalize to
$F$'s that care about history?

$$F(s)(t) = F(s(t), s(t-\alpha t), s(t-2\alpha t), \ldots)$$

so we'd need to consider a multivariate Taylor series

$$F(s)(t) = f_0 + \sum_{\ell} \frac{\partial F}{\partial (s(t-\ell \alpha t))} \bigg|_{s=0} \left( s(t-\ell \alpha t) - s_0 \right)$$

$$+ \frac{1}{2} \sum_{l_1, l_2} \frac{\partial^2 F}{\partial (s(t-l_1 \alpha t)) \partial (s(t-l_2 \alpha t))} \bigg|_{s=0} \left( s(t-l_1 \alpha t) - s_0 \right) \left( s(t-l_2 \alpha t) - s_0 \right)$$

$$+ \frac{1}{3} \sum_{l_1, l_2, l_3} \frac{\partial^3 F}{\partial (s(t-l_1 \alpha t)) \partial (s(t-l_2 \alpha t)) \partial (s(t-l_3 \alpha t))} \bigg|_{s=0} \left( s(t-l_1 \alpha t) - s_0 \right) \left( s(t-l_2 \alpha t) - s_0 \right) \left( s(t-l_3 \alpha t) - s_0 \right)$$

or, a multivariate orthogonal series

$$F(s)(t) = a_0 + \sum_{l} a_{l, t} \Phi_l(s(t-\ell \alpha t))$$

$$+ \sum_{l_1, l_2} a_{l_1, l_2} \Phi_{l_1, l_2}(s(t-l_1 \alpha t), s(t-l_2 \alpha t)) + \ldots$$
Each $\phi_{l_{1}} \ldots \phi_{l_{r}} (x_{1}, \ldots, x_{n})$ is a polynomial with leading term

$$x_{i_{1}} \cdot x_{i_{2}} \ldots \cdot x_{i_{r}};$$

they are orthogonal in the sense

$$\int \phi_{l_{1}} \ldots \phi_{l_{r}} (s_{1}, \ldots, s_{r}) \cdot \phi_{j_{1}} (s_{1}) \ldots \phi_{j_{q}} (s_{r}) W(s) ds = 0$$

unless $r = q$ and $l_{i} = m_{i}, \ldots, l_{r} = m_{r}$.

$W(s)$ is probability of a stimulus $s$.

Terminals

\[\begin{array}{ccc}
\text{Universality} & \text{Multi-volt (covariances limit)} \\
\text{Taylor} & \rightarrow & \text{Volterra series} \\
\text{orthogonal} & \rightarrow & \text{Wiener series}
\end{array}\]

Relationship between Wiener + Volterra strategy is a red of Taylor orthogonal strategies.

\[\{\begin{array}{ll}
\text{Volterra requires analyticity + limits} & S \rightarrow 0 \\
\text{Wiener does not require analyticity but depends on W(s)}.
\end{array}\]

A truncated Wiener series is best polynomial approx of any order given weighting $W(s)$, the Volterra series is the best local approximation near $0$.

A truncated Wiener series is a polynomial -- but it is not the same polynomial as the Volterra series of the same order.

Adding on additional term in the Wiener series results in the Volterra term missing.
Given a function of any order, Volterra & Wiener expansions constitute different bases:

<table>
<thead>
<tr>
<th>Volterra</th>
<th>Wiener ( (P = \sigma^2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_0 = 1 )</td>
<td>( \phi_0 = 1 )</td>
</tr>
<tr>
<td>( v_1 = x )</td>
<td>( \phi_1 = x )</td>
</tr>
<tr>
<td>( v_2 = x^2 )</td>
<td>( \phi_2 = x^2 - \sigma^2 )</td>
</tr>
<tr>
<td>( v_3 = x^3 )</td>
<td>( \phi_3 = x^3 - 3 \sigma^2 x )</td>
</tr>
<tr>
<td>( v_4 = x^4 )</td>
<td>( \phi_4 = x^4 - 6 \sigma^2 x^2 + 3 \sigma^4 )</td>
</tr>
</tbody>
</table>

Bessel,

| \( v_0 = \phi_0 \) |
| \( v_1 = \phi_1 \) |
| \( v_2 = \phi_2 + \sigma^2 \phi_0 \) |
| \( v_3 = \phi_3 + 3 \sigma^2 \phi_1 \) |
| \( v_4 = \phi_4 + 6 \sigma^2 \phi_2 + 3 \sigma^4 \phi_0 \) |

\( \sigma \) not generic (not costs match except for \( \phi_0 \))
Generic problems.

- Lots of parameters to measure.
  Need to choose a reasonable set: history (length), amplitude (A).

- Polynomials are not likely to be good global approximations.
  Alternative strategy: put together multiple local approximations.

\[ f \]

i.e., quasilinear near an operating point.

- Wiener expansion is polynomial in A.

- Try to build a model based on a limited V-W expansion.
  e.g., $e(t) \rightarrow \overline{L_1} \rightarrow \overline{L_2} \rightarrow \text{not polynomial}$

- Composition of subsystems: helpful if the were simple.

We can make headway on the "composition" problem by using
graph-theoretic tools, & also on the # of parameters problem.
We haven't used those tools in variable.
Using time-traveling machines

Rather than focus on the vector space of signals, focus on the vector space of systems, $M$.

Need to check if vector-space operations in $M$ make sense:

$$(F + G)(s) = F(s) + G(s)$$

$$(\alpha F)(s) = \alpha \cdot F(s)$$

Time-traveling ode on $M$ too:

$$(D_{t} F)(s) (t) = F(s) (t + \tau)$$

So $D_{t}$ is in $\text{Hom}(M, M)$, and commutes with all $F$.

Can we complexify $M$?

$$(F + i G)(s) = F(s) + i G(s)$$

Inner product on $M$?

Generalize $S_{1} \cdot S_{2} = \int_{-\infty}^{\infty} S_{1}(t) S_{2}(t) \, dt$.

$$(F, G) = \left< F(s), \overline{G(s)} \right>$$

So here, the choice of $s$ (and thereby the ensemble $\mathcal{E}$) needed to postulate $\mathcal{M}_{\mathcal{E}}$ is throughly matter.

Need to postulate $\mathcal{M}_{\mathcal{E}}$ is tracial—meaning, collect

$$\left< F, G \right> = \left< F(s)(0), \overline{G(s)(0)} \right> = \left< F(\mathcal{E})(t), \overline{G(\mathcal{E})(t)} \right> = \left< \mathcal{E} \right>$$
We now have all the familiar machinery available:

\[ D \gamma, \text{ a group of } \mathbb{R} \text{ acts on } \text{Hom}(M, \mathbb{R}) \text{ via matrix product} \]

so we can expect that the action of \( D \gamma \) decomposes \( M \) into eigenspaces, one for each irreducible representation of the time-translation group.

I.e., for each \( \omega \), there is a subspace \( M_\omega \) of \( M \):

\[ D \gamma, F = e^{i \omega \tau} F, \text{ for } F \in M_\omega. \]

\( M_\omega \) is the space of systems for which translation by \( \tau \) results in multiplication of the output by \( e^{i \omega \tau} \).

\( M_\omega \) contains any system whose output is \( e^{i \omega \tau} \).

For linear systems, we had characterized a system by its impulse response:

\[ r(t) = \int L(p) s(t - p) \, dp \]

or equivalently, its transfer function:

\[ \tilde{L}(\omega) = \int e^{-i \omega t} L(t) \, dt. \]

Now we want to think of \( L \) as a superposition of systems:

\[ L = \sum \tilde{L}_\omega, \text{ where } \tilde{L}_\omega \text{ is in } M_\omega \text{ and its response to } \]

\[ s(t) = \tilde{L}_\omega e^{i \omega t}. \]

I.e., \( \tilde{L}_\omega \) is \( L \), followed by a narrow-band filter at the frequency \( \omega \).
But there are other members of \( M \) besides \( L_w \); for example, \( s(t)^2 \) followed by hardpass at \( \omega \)

\[
\frac{s(t) \cdot s(t - \tau)}{[1 + s(t - 3\tau)]^{2/3}}
\]

Assume (see 2003-4 note) that we can construct a basis for \( M \):

"1st order" systems: \( L_w \), coarure.

"2nd order" systems: \( L_w \oplus L_{w_2} \)

\[
\begin{align*}
&L_{w_2} \\
&\times \\
&L_w
\end{align*}
\]

provided \( \omega + \omega_2 = \omega \).

"3rd order" systems \( L_w \oplus L_{w_2} \oplus L_{w_3} \):

\[
\begin{align*}
&L_{w_3} \\
&\times \\
&L_{w_2} \\
&\times \\
&L_w
\end{align*}
\]

provided \( \omega + \omega_2 + \omega_3 = \omega \).
Why does this work:

We need a larger group, since \( MC \) is too large.

Back to vector spaces of singular \( V \).

Each irreducible group acts on \( V \), \( k \) times on \( V \otimes \cdots \otimes V \).

\[
\begin{align*}
S_k(t) & \rightarrow L_k \\
S_3(t) & \rightarrow L_3 \\
\vdots & \\
S_1(t) & \rightarrow L_1
\end{align*}
\]

The action of \( G \otimes \cdots \otimes G \) on \( V \otimes \cdots \otimes V \) decomposes it into 1-d subspaces, namely, the subspaces of states for which the action by \( T_1, T_2, \ldots, T_k \) is equivalent to multiplication of \( s_1 \otimes \cdots \otimes s_k \) by \( \lambda \), \( \lambda \) is a \( \lambda \)th root of unity.

\[
(L_1 \otimes \cdots \otimes L_k)(s_1 \otimes \cdots \otimes s_k) = L_1(s_1(t)) \cdot L_2(s_2(t)) \cdots L_k(s_k(t))
\]

This is not linear

\[
(s_1 \otimes s_2 \otimes \cdots \otimes s_k) + (s_1' \otimes \cdots \otimes s_k') \neq (s_1 + s_1') \otimes \cdots \otimes (s_k + s_k')
\]

so only add tensor product if all \( k \) one-term match
We've constructed a standard non-linear system \( L_1 \otimes \ldots \otimes L_k \) that acts on \( s(t) \) by

\[
\begin{align*}
    s_1(t) & \rightarrow L_1 \\
    s_2(t) & \rightarrow L_2 \\
    \vdots & \quad \vdots \\
    s_k(t) & \rightarrow L_k \\
\end{align*}
\]

\( s(t) \)

For which transfer of \( s(t) \) by \( T \) results in multiplication of the response by

\[
e^{j\omega_1 t + j\omega_2 t + \ldots + j\omega_k t}
\]

\[
e^{j(\sum \omega_j) t}
\]

provided that each \( L_j \) is narrowband at \( \omega_j \).

How does the decomposition of \( M \) into \( L_1, L_2, \ldots, L_k \) work via the practical case?

\( s(t) \rightarrow G \rightarrow X \rightarrow H \rightarrow \) where \( G \neq H \) one linear.

\[
\begin{align*}
    s(t) & \rightarrow G \\
    \vdots & \quad \vdots \\
    s(t) & \rightarrow G \\
    \vdots & \quad \vdots \\
\end{align*}
\]

\( G \) has t.f. \( G(w) \). \( G \) is a sum of narrow-bd systems \( L_w \), each weighted by \( G(w) \).
\[ g = \sum \tilde{G}(\omega) L_\omega \] so there is a contribution \( \otimes \)

For each \( \omega_1, \omega_2 \) of

\[ s(t) = \begin{cases} L_{\omega_1} & \text{if } \omega_1 < \omega_2 \\ L_{\omega_2} & \text{if } \omega_2 < \omega_1 \end{cases} \]

weighted by \( \tilde{G}(\omega_1) \tilde{G}(\omega_2) \)

Output above module is \( \tilde{G}(\omega_1) \tilde{G}(\omega_2) e^{i(\omega_1 t + \omega_2 t)} \) for

\[ s(t) = e^{i\omega_1 t} + e^{i\omega_2 t} \]

\[ \text{So, at least formally, } s(t) \rightarrow \tilde{G}(\omega_1) \tilde{G}(\omega_2) e^{i(\omega_1 t + \omega_2 t)} \]

\[ \tilde{H}(\omega_1, \omega_2) \] for

\[ \tilde{H}(\omega_1, \omega_2) \]

\[ \tilde{G}(\omega_1, \omega_2) \tilde{H}(\omega_1 + \omega_2). \]

\[ \text{\textit{at least formally}.} \]

\[ \text{\textit{Above } } s \text{ \textit{is complex.}} \]

\[ \text{\textit{What about other frequencies?}} \]

\[ \text{\textit{What if } } \Delta \text{ \textit{is not } } x^2? \]