Nonlinear Systems Theory - Part I

Overview:

General setup as before $s(t) \to \boxed{F} \to r(t)$

Want a principled way to describe F, but without assuming linearity

→ a more concise description than simply a list of (s, r)-pairs

→ a description that suggests (or tests) ideas for the internals of F

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

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

and

$$F(\alpha s_1) = \alpha r_1$$

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

We then made 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, i.e.,

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

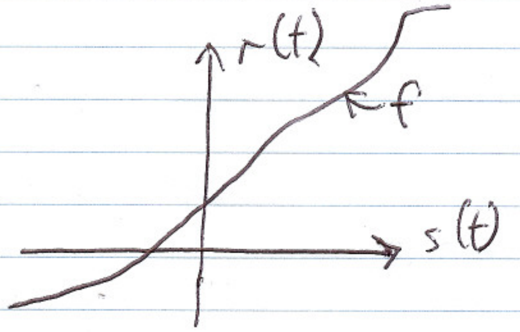
We're dispensing with linearity of F - but we still have time translation invariance.

(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.
i.e.,
$$[F_0(s)](t) = f(s(t)) \text{ for some ordinary function } f.$$

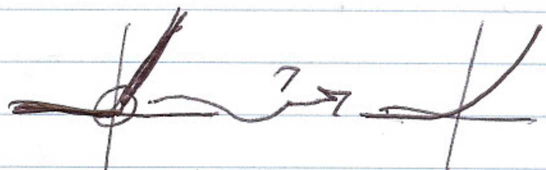
These are the "static nonlinearities". At the very least, describing all F's is at least as hard as describing all f_0's.



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 $f_k = \left.\frac{d^k}{ds^k} f\right|_{s=0}$.

Pros: • it is universal* + principled (*: for f's that have a Taylor series)

Cons: • tough to measure $\frac{d^k}{ds^k} f$, because of noise.

• What if f looks like 

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

(Taylor series requires that f is analytic, $f(s) = |s|$ is not.)

An alternative is to express f as an orthogonal expansion

$$f(s) = \sum_{k=0}^{\infty} a_k \varphi_k(s)\,,$$ where $\varphi_k(s)$ are orthogonal in the sense that

$$\int_{-\infty}^{\infty} \varphi_k(s)\, \varphi_l(s)\, w(s)\, ds = \delta_{kl}\, c_{kl}$$

for some $w(s) \geq 0$

Then $$a_k = \frac{1}{c_k} \int_{-\infty}^{\infty} f(s)\, \varphi(s)\, w(s)\, ds\,.$$

Typical example: $$W(s) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-s^2/2\sigma^2}$$ (Gaussian)

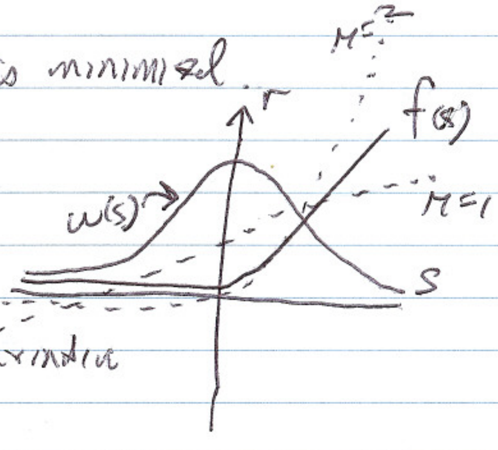
The φ's are the Hermite polynomials

$(\sigma = 1)$ $\quad \varphi_0(s) = 1,\quad \varphi_1(s) = s,\quad \varphi_2(s) = s^2 - 1,$

$\varphi_3(s) = s^3 - 3s,\quad \varphi_4(s) = s^4 - 6s^2 + 3,\ \ldots$

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

$$\int_{-\infty}^{\infty} \Big(f(s) - \sum_{k=0}^{M} a_k \varphi_k(s) \Big)^2 w(s)\, ds$$ is minimized.



This is a reasonable definition of "best" if your inputs are drawn from $W(s)$.

The orthog. expansion coefficients are more practical to measure (don't require a limit of $s \to 0$) but they will depend on W (i.e., NOT universal).
The orthogonal expansion does not require the existence of derivatives of f.

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

$$F(s)(t) = F(s(t), s(t-\Delta t), s(t-2\Delta t), \cdots)$$

so we'd need to consider a multivariate Taylor series

$$
\begin{aligned}
F(s)(t) &= f_0 && \text{"offset"} \\
&+ \sum_{l} f_{1,l} \left.\frac{\partial F}{\partial(s(t-l\Delta t))}\right|_{s=0} && \text{"linear"} \\
&+ \frac{1}{2}\sum_{l_1,l_2} f_{2,l_1,l_2} \left.\frac{\partial^2 F}{\partial(s(t-l_1\Delta t))\,\partial(s(t-l_2\Delta t))}\right|_{s=0} && \text{"quadratic"} \\
&+ \frac{1}{3}\sum_{l_1,l_2,l_3} f_{3,l_1,l_2,l_3} \left.\frac{\partial^3 F}{\partial(s(t-l_1\Delta t))\,\partial(s(t-l_2\Delta t))\,\partial(s(t-l_3\Delta t))}\right|_{s=0} \cdots \\
&\cdots
\end{aligned}
$$

or, a multivariate orthogonal series

$$F(s)(t) = a_0 + \sum_{l} a_{1,l}\, \varphi_{1,l}(s(t-l\Delta t))$$

$$+ \sum_{l_1,l_2} a_{2,l_1,l_2}\, \varphi_{2,l_1,l_2}(s(t-l_1\Delta t), s(t-l_2\Delta t)) + \cdots$$

5

Each $\varphi_{r;l_1,\ldots,l_r}(x_{l_1},\ldots,x_{l_r})$ is a polynomial with leading term $x_{l_1}\cdot x_{l_2}\cdot\cdots\cdot x_{l_r}$; they are orthogonal in the sense

$$\int \varphi_{r;l_1,\ldots,l_r}(s_{l_1},\ldots,s_{l_r})\cdot\varphi_{q;m_1,\ldots,m_q}(s_{m_1},\ldots,s_{m_q})\,W(s)\,ds$$

$$= 0 \text{ unless } r=q \text{ and } l_1=m_1,\ldots,l_r=m_q.$$

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

Terminology: Univariate → Multivariate (continuous limit)

Univariate		Multivariate (continuous limit)
Taylor	→	Volterra series
orthogonal	→	Wiener series

Relationship betw Wiener + Volterra strategy is ~ rel of Taylor orthog'l strategies.

{ Volterra requires analyticity + limit of $s \to 0$,
Wiener does not require analyticity but depends on $W(s)$.

A truncated Wiener series is best polynomial approx of a given 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 an additl term in the Wiener series revises the monomials of lower order; adding an additional Volterra term does not.

6

Given a truncation at any order, Volterra & Wiener expansions constitute different bases:

Volterra		Wiener ($P = \sigma^2$)	
$v_0 =$	1	$\varphi_0 =$	1
$v_1 =$	x	$\varphi_1 =$	x
$v_2 =$	x^2	$\varphi_2 =$	$x^2 - P$
$v_3 =$	x^3	$\varphi_3 =$	$x^3 - 3Px$
$v_4 =$	x^4	$\varphi_4 =$	$x^4 - 6Px^2 + 3P^2$

Besides,

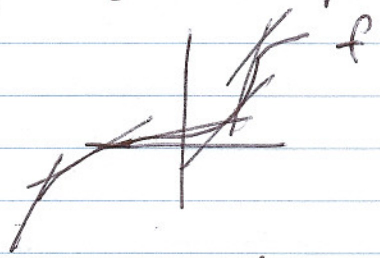
$$v_0 = \varphi_0$$
$$v_1 = \varphi_1$$
$$v_2 = \varphi_2 + P\varphi_0$$
$$v_3 = \varphi_3 + 3P\varphi_1$$
$$v_4 = \varphi_4 + 6P\varphi_2 + 3P^2\varphi_0$$

[Not generic that coef's match except for sign]

Generic problems.

- Lots of parameters to measure
  [Need to choose a reasonable Δt, history length, amplitude (σ), ? other aspects]

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



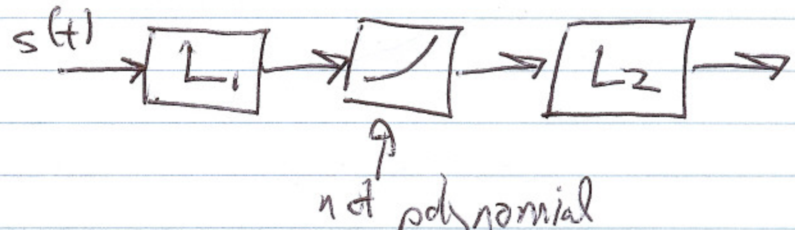
i.e., quasilinearity based on operating point

or

Wiener expansion but parametric in σ.

or, build a model based on a limited V-W expansion

e.g.



- Composition of subsystems – helpful if they were simple

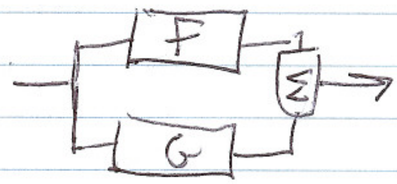
We can make headway on the "composition" problem by using the group-theoretic tools, & also on the #-of-parameter problem – we haven't yet used time-translation invariance.

Using time-translation invariance

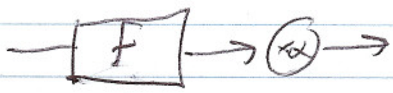
Rather than focus on the vector space of signals, focus on the vectorspace of systems, $\mathcal{M}$.

Need to check that vector-space operations in $\mathcal{M}$ make sense.

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



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



Time-translation acts on $\mathcal{M}$ too: $(D_T F)(s)(t) = F(s)(t+T)$

so D_T is in $\mathrm{Hom}(\mathcal{M}, \mathcal{M})$ & commutes with all of F.

Can we complexify $\mathcal{M}$? $(F+iG)(s) = F(s) + i\,G(s)$

Inner product on $\mathcal{M}$? (generalize $s_1 \cdot s_2 = \int s_1(t)\,\overline{s_2(t)}\,dt$):

$$(F, G) = \left\langle F(s)(t) \cdot \overline{G(s)(t)} \right\rangle_{t,\,\Omega}$$

So here, the choice of signals matters - the ensemble Ω.
Need to postulate that Ω is translation-invariant, so that

$$(F,G) = \left\langle F(s)(0)\,\overline{G(s)(0)} \right\rangle_\Omega = \left\langle F(s)(T)\,\overline{G(s)(T)} \right\rangle_\Omega$$

for any T

We now have all the familiar machinery in place:

D_τ, a group that acts in $Hom(\mathcal{M}, \mathcal{M})$ & preserves the inner product

So we can expect that the actions of D_τ decompose $\mathcal{M}$ into eigenspaces, one for each irreducible representation of the time-translation group

I.e., for each ω, there is a subspace $\mathcal{M}_\omega$ of $\mathcal{M}$

$$D_\tau F = e^{i\omega\tau} F, \text{ for } F \in \mathcal{M}_\omega.$$

$\mathcal{M}_\omega$ is the space of systems for which translation by τ results in multiplication of the output by $e^{i\omega\tau}$.

so

$\mathcal{M}_\omega$ contains ANY system whose output is $e^{i\omega\tau}$.

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

$$r(t) = \int L(\tau)\, s(t-\tau)\, d\tau$$

or equivalently its transfer function $\tilde{L}(\omega) = \int_0^\infty e^{-i\omega t} L(t)\, dt$.

Now we want to think of L as a superposition of systems

$L = \sum_\omega L_\omega$, where L_ω is in $\mathcal{M}_\omega$ and its response to $s(t)$ is $\tilde{L}(\omega)\, e^{i\omega t}$.

i.e., L_ω is L, followed by a narrow-band filter at the frequency ω.

But there are other members of M_ω besides L_ω; for example

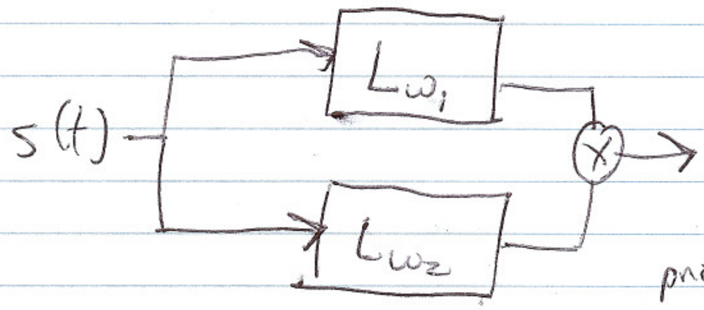
$s(t)^2$ followed by bandpass at ω

or $\dfrac{s(t)\,s(t-\tau)}{[1+s(t-3\tau)]^{2/3}}$ followed by bandpass at ω, etc.

Assert (see 2003-4 notes) that we can construct a basis for M_ω:

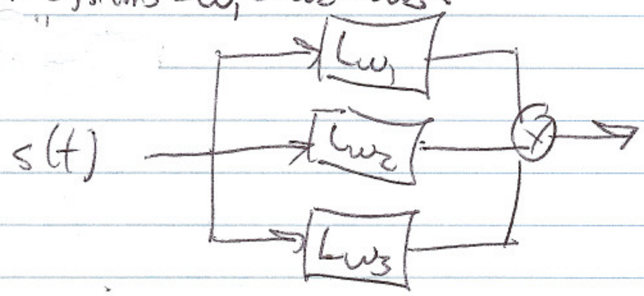
"1st-order" systems: L_ω, as above.

"2nd-order" systems: $L_{\omega_1} \otimes L_{\omega_2}$



provided that $\omega_1 + \omega_2 = \omega$.

"3rd order" systems $L_{\omega_1} \otimes L_{\omega_2} \otimes L_{\omega_3}$:



provided that $\omega_1 + \omega_2 + \omega_3 = \omega$.

$\vdots$

11

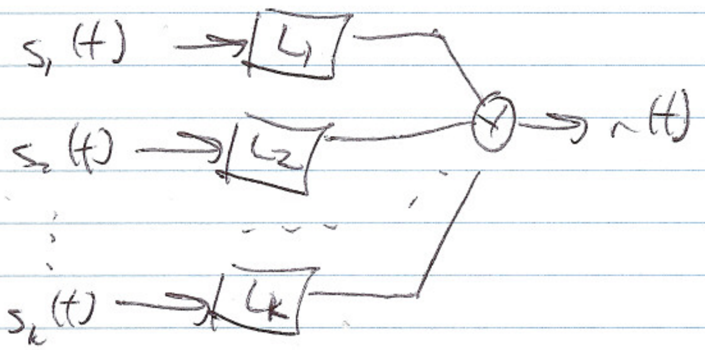
Why does this work?

We need a larger group, since $\mathfrak{M}_\omega$ is too large.

Back to vector spaces of signals, V.

G, time-translation group acts on V, ∴ also on $\underbrace{V\otimes\cdots\otimes V}_{k \text{ times}}$.

So $G\otimes\cdots\otimes G$ acts on $V\otimes\cdots\otimes V$.



The action of $G\otimes\cdots\otimes G$ on $V\otimes\cdots\otimes V$ decomposes it into 1-d subspaces, namely, the subspace of signals for which translation by $\tau_1, \tau_2, \cdots, \tau_k$ ~~leads~~ is equivalent to multiplication of $s_1\otimes\cdots\otimes s_k$ by $e^{i\omega_1\tau_1+\cdots+i\omega_k\tau_k}$.

$L_1\otimes\cdots\otimes L_k$ is linear on $s_1\otimes\cdots\otimes s_k$, i.e.,

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

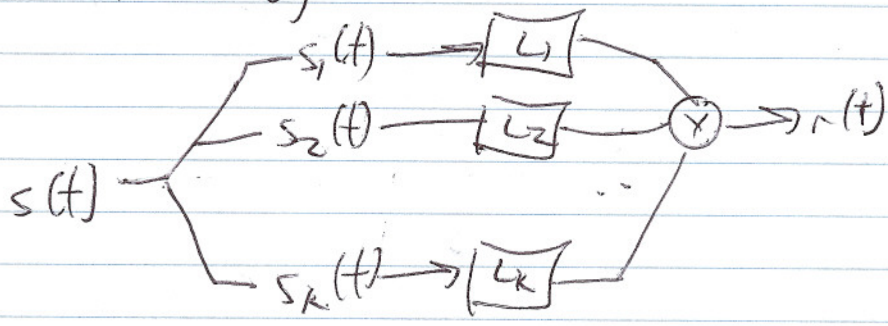
but now we can also let $L_1\otimes\cdots\otimes L_k$ act on $s(t)\otimes\cdots\otimes s(t)$ –

This is <u>NOT</u> linear

$$(s_1\otimes s_2\otimes\cdots\otimes s_k)+(s_1'\otimes\cdots\otimes s_k') \neq (s_1+s_1')\otimes\cdots\otimes(s_k+s_k')$$

can only add tensor product if all but one term match

We've constructed a standard nonlinear system $L_1 \otimes \cdots \otimes L_k$ that acts on $s(t)$ by

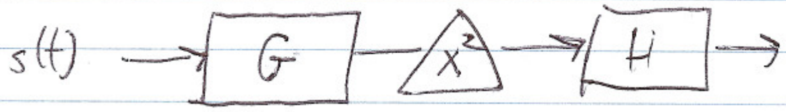


For which translation of $s(t)$ by τ results in multiplication of the response by

$$e^{i\omega_1\tau + i\omega_2\tau + \cdots + i\omega_k\tau}$$

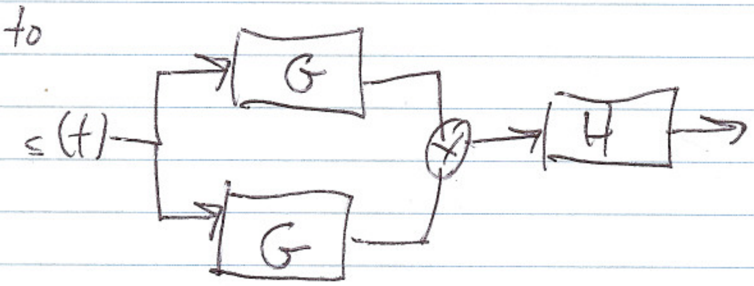
$$= e^{i(\sum \omega_j)\tau},$$ provided that each L_p is narrowband at ω_p.

How does the decomposition of $\mathcal{M}$ into L_ω, $L_{\omega_1} \otimes L_{\omega_2}$, etc. work in a practical case?



where G & H are linear.

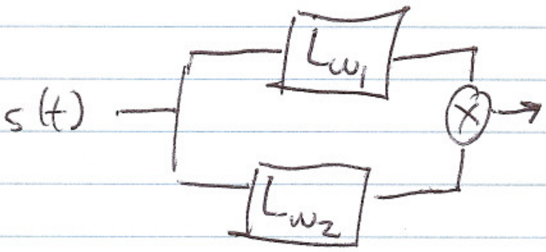
Equiv to



Say G has t.f. $\tilde{G}(\omega)$. G is a sum of narrow-bd systems L_ω, each weighted by $\tilde{G}(\omega)$.

$G = \sum' \tilde{G}(\omega) L_\omega$ so there is a contribution of $\otimes$

for each ω_1, ω_2 of



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

Output of above module is $\tilde{G}(\omega_1)\, \tilde{G}(\omega_2)\, e^{i(\omega_1+\omega_2)t}$, if

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

So, <u>at least formally</u>, $s(t) \to \boxed{G} \to \boxed{x^2} \to \boxed{H} \to$

has output $\tilde{G}(\omega_1)\, \tilde{G}(\omega_2)\, e^{i(\omega_1+\omega_2)t}\, \tilde{H}(\omega_1+\omega_2)$ for

$s(t) = e^{i\omega_1 t} + e^{i\omega_2 t}$, i.e., the component of this system in $L_{\omega_1} \otimes L_{\omega_2}$ is

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

1) "at least formally": ① Above s is complex.
② What about other frequencies?
③ What if $\triangle$ is not x^2?