L41 Nonlinear Systems Theory - Part II (Restart)

PLAN

Orthogonalization (Gram Schmidt Procedure)
↓
Orthogonal Polynomials : approximation of nonlinear functions $f(x)$
↓
Orthogonal Multinomials : approximation of nonlinear functions $f(x_1, x_2, \cdots, x_k)$
↓
Orthogonal Functionals : approximation of $f[x(t)]$

→ Discrete time approach → "White Noise" ↗ m-sequences
→ Frequency domain approach ↗ "White Noise" ↘ sinusoids

We will see that

* describing a nonlinear system can be viewed as a regression problem

* in contrast to the analysis of linear systems, there is no universal choice of "regressors" (in linear systems, the sinusoids)

* the description of a nonlinear system depends on the choice of regressors ("context")

* some choices are better than others
  - efficiency of analysis
  - usefulness of description

Orthogonalization, approximation, Gram-Schmidt procedure.

Working in a vector space V (over $\mathbb{R}$) with an inner product $\langle , \rangle$.

Here, $v, \varphi, f \in V$, abstract - but we have in mind vectors that represent functions of single variable $f(x)$
functions of multiple discrete variables $f(x_1, \dots, x_k)$
"functionals": functions of a continuum of variables $f[x(t)]$

Say we want to approximate

$f \approx \sum_{j=1}^{r} \alpha_j v_j$, for some given library $v_1, \dots, v_r$

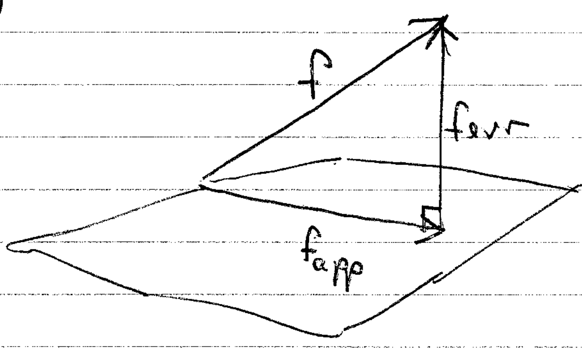
That is, we want to find the $\{\alpha_j\}$ that minimize $R = \left| f - \sum_{j=1}^{r} \alpha_j v_j \right|^2$,
where $|u|^2 = \langle u, u \rangle$.

We can write $f = f_{app} + f_{err}$, where $f_{app} = \sum_{j=1}^{r} \alpha_j v_j$:
$R = |f_{err}|^2$.

We'd like to show that when $|f_{err}|^2$ is minimized, f_{err} is orthogonal to f_{app}, i.e.,

$$\langle f_{app}, f_{err} \rangle = 0$$

i.e., we are projecting f into the subspace spanned by $v_1, \dots, v_r$.



Say, at the minimum, $f_{err} = \sum_{j=1}^{r} \beta_j v_j + \varepsilon$,

where

$$\varepsilon \perp v_j \quad \text{, i.e.,} \quad \langle \varepsilon, v_j \rangle = \langle v_j, \varepsilon \rangle = 0.$$

We have $f = \underbrace{\sum_{j=1}^{r} \alpha_j v_j}_{f_{app}} + \underbrace{\sum_{j=1}^{r} \beta_j v_j + \varepsilon}_{f_{err}}$

But this can be reorganized into

$$f = \underbrace{\sum_{j=1}^{r} (\alpha_j + \beta_j) v_j}_{f_{new\text{-}app}} \quad \underbrace{+ \varepsilon}_{f_{new\text{-}err}}$$

We know that $|f_{new-err}|^2 \geq |f_{err}|^2$,

because we hypothesized that f_{err} was minimum.

$$|f_{err}|^2 = \left|\sum_{j=1}^{r} \beta_j v_j + \varepsilon\right|^2 = \left\langle \left(\sum_{j=1}^{r} \beta_j v_j + \varepsilon\right), \left(\sum_{k=1}^{r} \beta_k v_k + \varepsilon\right) \right\rangle$$

$$= \sum_{j=1}^{r} \sum_{k=1}^{r} \beta_j \bar{\beta}_k \langle v_j, v_k \rangle + \sum_{j=1}^{r} \beta_j \langle v_j, \varepsilon \rangle + \sum_{k=1}^{r} \bar{\beta}_k \langle \varepsilon, v_k \rangle + \langle \varepsilon, \varepsilon \rangle$$

$$= \left|\sum_{j=1}^{r} \beta_j v_j\right|^2 + |\varepsilon|^2 = \left|\sum_{j=1}^{r} \beta_j v_j\right|^2 + |f_{new-err}|^2$$

So $|\sum \beta_j v_j|^2 = 0$, $\Rightarrow f_{err} = f_{new-err} = \varepsilon$.

$\boxed{L7}$

In words, f_{err} has no component in the subspace spanned by $v_1, \dots, v_r$, since if it did, we could improve the approximation by adding this component into f_{app}

Key ingredient was $\langle \, , \, \rangle$.

Geometrically, f_{app} is the projection of f onto the subspace spanned by the v_j. Projection is linear.

Formal solution: Let $A =$ matrix of columns $v_1, \dots, v_r$.

The projection is $P = A(A^TA)^{-1}A^T$.

$$\text{Verify } P^2 = A(A^TA)^{-1}A^T A(A^TA)^{-1}A^T$$
$$= A(A^TA)^{-1}\cancel{(A^TA)(A^TA)^{-1}}A^T$$
$$= A(A^TA)^{-1}A^T = P$$

and that span of P include each column ($P = A \cdot X$)

Not a "useful" solution, in the sense that we'd need to calculate $(A^TA)^{-1}$.

[When does this not exist?]

We could also try to minimize $R(\alpha_1, \dots, \alpha_r) = |f - \sum_{j=1}^{r} \alpha_j v_j|^2$

by $\frac{\partial R}{\partial \alpha_k} = 0$, leads to a linear system of equations for the r α_j's

Better solution is to replace $\{v_1, \dots, v_r\}$ by an orthogonal set

$\{\varphi_1, \dots, \varphi_r\}$ with the same span.

Then the approximation $f_{app} = \sum_{j=1}^{r} \alpha_j v_j = \sum_{j=1}^{r} a_j \varphi_j$,

and $\langle f_{app}, \varphi_h \rangle = \sum_{j=1}^{r} a_j \langle \varphi_j, \varphi_h \rangle = a_h \langle \varphi_h, \varphi_h \rangle$

so $$a_h = \frac{\langle f_{app}, \varphi_h \rangle}{\langle \varphi_h, \varphi_h \rangle} = \frac{\langle f, \varphi_h \rangle}{\langle \varphi_h, \varphi_h \rangle}$$, since $(f - f_{app}) \perp \varphi_h$.

Advantage 1: No system of equations to solve

Advantage 2: Straightforward to improve the approx by adding new terms.

Suppose we have $f_{app}^{[r]} = \sum_{j=1}^{r} \alpha_j v_j$ & want to add v_{j+1}.

$$f_{app}^{[r+1]} = \sum_{j=1}^{r+1} \alpha_j' v_j',$$ no guarantee that $\alpha_j = \alpha_j'$.

In fact if $\langle v_{r+1}, v_j \rangle \neq 0$, then typically $\alpha_j \neq \alpha_j'$.

But adding a new φ_{r+1} doesn't revise previous a_j's: $a_h = \frac{\langle f, \varphi_h \rangle}{\langle \varphi_h, \varphi_h \rangle}$

Think of $f \approx \sum_{j=1}^{r} \alpha_j v_j$ as a regression, and

$f \approx \sum_{j=1}^{r} a_j \varphi_j$ is a way to solve it.

How to create the φ_j's, such that the span of

$\{v_1, \cdots, v_n\}$ = span of $\{\varphi_1, \cdots, \varphi_n\}$, and φ_j's orthogonal?

"Gram Schmidt" procedure.

$$\varphi_1 = v_1$$

$\varphi_2 = v_2 -$ projection of v_2 onto space spanned by $\{v_1\}$

$\varphi_3 = v_3 -$ " " v_3 " " " " $\{v_1, v_2\}$

$\varphi_4 = v_4 -$ " " " " " " " $\{v_1, v_2, v_3\}$

At each stage, space spanned by $\{v_1, \cdots, v_n\}$ = space spanned by $\{\varphi_1, \cdots, \varphi_n\}$.

So calculation of the projection is easy, if you proceed iteratively

$$\varphi_1 = v_1$$

$$\varphi_2 = v_2 - \frac{\langle v_2, \varphi_1\rangle}{\langle \varphi_1, \varphi_1\rangle}\varphi_1$$

$$\varphi_3 = v_3 - \frac{\langle v_3, \varphi_2\rangle}{\langle \varphi_2, \varphi_2\rangle}\varphi_2 - \frac{\langle v_3, \varphi_1\rangle}{\langle \varphi_1, \varphi_1\rangle}\varphi_1$$

etc.

Will fail if some $\varphi_n = 0$, which will happen if v_n is in span $\{\varphi_1, \cdots, v_{n-1}\}$.

An example: vectors are functions of a single variable, $f(x)$.

Inner product: $\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)\, g(x)\, W(x)\, dx$

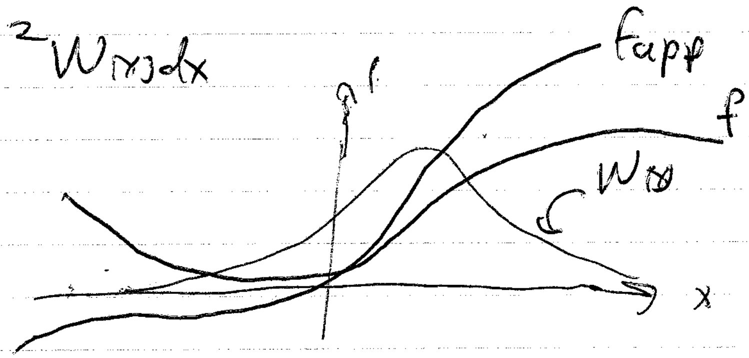
for some $W(x) \geq 0$ $\quad$ V is the v.s. of functions for which

(why?) $\qquad \int_{-\infty}^{\infty} |f(x)|^2 W(x)\, dx < \infty$.

By minimizing f_{err}, we minimize

$$\int |f - f_{app}|^2 W(x)\, dx$$

i.e., f_{app} is a good approximation where W is large.



Say $v_0 = x^0$, $v_1 = x^1$, $v_2 = x^2$, etc.

Say $\int x^k W(x)\, dx = M_k$, $\left(\begin{array}{l}\text{choose } W(x) \text{ so that} \\ M_0 \equiv 1 \quad \text{i.e., } W(x) \text{ is} \\ \text{a probability distribution}\end{array}\right)$

$\varphi_0 = x^0$.

$$\varphi_1 = x^1 - \frac{\langle x^1, \varphi_0\rangle}{\langle \varphi_0, \varphi_0\rangle}\varphi_0 = x^1 - \frac{M_1}{M_0}x^0 = x^1 - M_1$$

$$\left(\langle x^1, \varphi_0\rangle = \int x^1 \cdot x^0\, W(x)\, dx = \int x^1 W(x)\, dx = M_1\right)$$

$\boxed{34}$

$$\varphi_2 = x^2 - \frac{\langle x^2, \varphi_1 \rangle}{\langle \varphi_1, \varphi_1 \rangle} \varphi_1 - \frac{\langle x^2, \varphi_0 \rangle}{\langle \varphi_0, \varphi_0 \rangle} \varphi_0$$

$$\langle x^2, \varphi_1 \rangle = \int x^2 (x^1 - M_1) W(x) dx = \int (x^3 - M_1 x^2) W(x) dx$$
$$= M_3 - M_1 M_2$$

$$\langle x^2, \varphi_0 \rangle = \int x^2 W(x) dx = M_2 .$$

$$\langle \varphi_1, \varphi_1 \rangle = \int (x^1 - M_1)^2 W(x) dx = \int (x^2 - 2M_1 x^1 + M_1^2) W(x) dx$$
$$= M_2 - M_1^2 .$$

$$\varphi_2 = x^2 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2} (x^1 - M_1) - M_2$$

$$= x^2 - \left(\frac{M_3 - M_1 M_2}{M_2 - M_1^2} \right) x^1 + \frac{M_1 M_3 - M_2^2}{M_2 - M_1^2}$$

It can be done, but it's messy. Important special case is symmetric $W(x)$

Say $W(x) = W(-x)$. Then $M_1, M_3, M_5, \ldots$ are 0.

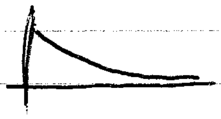
$$\varphi_0 = x^0$$

$$\varphi_1 = x^1$$

$$\varphi_2 = x^2 - M_2 x^0$$

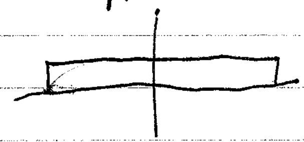
$$\varphi_3 = x^3 - \frac{M_4}{M_2} x^1$$

"Classical" special cases: $W(x) = \text{Gaussian} \rightarrow$ Hermite polynomials

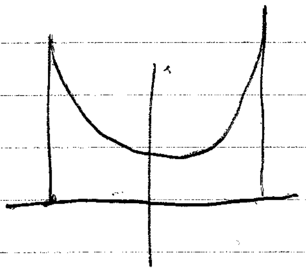


$$W(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & x < 0 \end{cases} \rightarrow \text{Laguerre polynomials}$$

uniform approx in an interval



$$W(x) = \begin{cases} \frac{1}{2}, & |x| < 1 \\ 0, & |x| > 1 \end{cases} \rightarrow \text{Legendre polynomials}$$



$$W(x) = \begin{cases} \dfrac{1/\pi}{\sqrt{1-x^2}}, & |x| < 1 \\ 0, & |x| > 1 \end{cases} \rightarrow \text{Chebyshev p-n's}$$

($x = \sin\theta$ makes connection with frequency-domain approach)

<u>Hermites</u> $\quad W(x) = \dfrac{1}{\sqrt{2\pi\rho}} e^{-x^2/2\rho}$

$$M_1, M_3, M_5, \cdots = 0$$

$$M_0 = 1, \quad M_2 = \rho, \quad M_4 = 3\rho^2 \quad M_6 = 15\rho^3 \cdots$$

$$M_{2n} = \frac{(2n)!}{2^n \cdot n!}\rho^n = \frac{(2n)(2n-1)(2n-2)(2n-3)\cdots 1)}{2^n \cdot n \cdot (n-1) \cdots (1)}\rho^n$$

$$= (2n-1)(2n-3)(2n-5)\cdots 1 \cdot \rho^n$$

(with $h_n = \varphi_n$)

$h_0 = 1$	$1 = h_0$
$h_1 = x$	$x = h_1$
$h_2 = x^2 - P$	$x^2 = h_2 + Ph_0$
$h_3 = x^3 - 3Px$	$x^3 = h_3 + 3Ph_1$
$h_4 = x^4 - 6Px^2 + 3P^2$	$x^4 = h_4 + 6Ph_2 + 3P^2h_0$
$h_5 = x^5 - 10Px^3 + 15P^2x$	$x^5 = h_5 + 10Ph_3 + 15P^2h_1$

$\vdots$ The Hermites have many other properties, usually with parallels in the other classical families!

$$\frac{dh_n}{dx} = n h_{n-1}$$

$$\frac{dh_n}{dP} = \frac{n(n-1)}{2} h_{n-2} \qquad \text{(non-generic)}$$

$$h_{n+1} = x h_n - n h_{n-1} \qquad \text{"recursion"}$$

$$|h_n|^2 = n!\,P^n \qquad \text{"normalization"}$$

h_n has n roots, all real.

$$\sum_{n=0} \frac{h_n(x)\,t^n}{n!} = e^{tx - Pt^2/2} \qquad \text{"Generating function"}$$

Generating function is the royal road to deducing all the other properties.

coef of t^n in $e^{tx - \rho t^2/2}$ has x^n, x^{n-2}, x^{n-4}, $\ldots$, since

$$e^u = \sum_{k=0}^{\infty} \frac{u^k}{k!} \quad ; \left(u = tx - \frac{\rho t^2}{2} \right)$$

For example, to see that the h_k's are orthogonal:

Let $c_{jk} = \langle h_j, h_k \rangle = \int h_j(x)\, h_k(x)\, W(x)\, dx.$

Let $$C = \sum_{\substack{j=0\\k=0}}^{\infty} \frac{c_{jk}\, s^j t^k}{j!\, k!} = \sum_{j,k} \int \frac{s^j t^k}{j!\, k!} h_j(x)\, h_k(x)\, W(x)\, dx$$

$$C = \int \sum_j \frac{s^j h_j(x)}{j!} \sum_k \frac{t^k h_k(x)}{k!} W(x)\, dx$$

$$= \int e^{sx - \rho s^2/2}\, e^{tx - \rho t^2/2}\, e^{-x^2/2\rho} \frac{1}{\sqrt{2\pi\rho}}\, dx$$

$$= \frac{1}{\sqrt{2\pi\rho}} e^{-\rho s^2/2 - \rho t^2/2} \int e^{-x^2/2\rho + sx + tx}\, dx$$

$\boxed{251}$

Complete the square: $-\frac{x^2}{2\rho} + sx + tx = -\frac{x^2}{2\rho} + sx + tx - \frac{\rho}{2}(s+t)^2$

$$+\frac{\rho}{2}(s+t)^2$$

So $-\frac{x^2}{2\rho} + sx + tx = -\frac{1}{2\rho}\left(x - \rho(s+t)\right)^2 + \frac{\rho}{2}(s+t)^2$

$$e = \frac{1}{\sqrt{2\pi\rho}}\, e^{-\rho s^2/2 - \rho t^2/2}\, e^{\frac{\rho}{2}(s+t)^2} \int e^{-\frac{1}{2\rho}(x-\rho(s+t))^2}\, dx$$

A de-centred Gaussian of variance ρ

$$e = e^{-\rho s^2/2 - \rho t^2/2 + \frac{\rho}{2}(s+t)^2} = e^{\rho s t}$$

But $e_{\text{calc}} = \sum \frac{c_{jk} s^j t^k}{j!k!}$

So $\sum \frac{c_{jk} s^j t^k}{j!k!} = e^{\rho s t} = \sum \frac{(st)^n}{n!} \rho^n$

So $c_{jk} = 0$ unless $j = k = n$

$\frac{c_{nn}}{n!^2} = \frac{\rho^n}{n!}$, so $c_n = n!$

From orthogonal polynomials to orthogonal multinomials

Above, $\{v_k\} = \{1, x, x^2, x^3, \cdots\}$ (x represents an "input" value we're approximating $f(x)$)

What if we wanted to approximate $f(x,y)$?

(x, y represent two inputs, or inputs at two times)

$\{v_k\} = \{1, x, x^2, x^3, \cdots, y, y^2, y^3, \cdots, xy, x^2y, \cdots, xy^2, \cdots\}$

The orthogonalization procedure will depend on the order chosen, as will the approximations

But we'd like $f(x,y) = A(x) + B(y)$ to be simple to represent if A + B are both linear

So,

$\{1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3, \cdots\}$

With multiple variables, we'd like to do the analogous:

$\{1; x_1, x_2, x_3, \cdots, x_Q; x_1^2, x_1x_2, \cdots, x_2^2, x_2x_3, \cdots, \cdots\}$

a notational catastrophe!

Use vectorized subscripts + exponents: $v_{\vec{k}} = x_1^{k_1} x_2^{k_2} \cdots x_Q^{k_Q} = \vec{x}^{\vec{k}}$

"order" of $\vec{k} = \sum k_i$

1 is zeroth-order

$x_1, \cdots, x_Q$ are 1st-order

$x_1^2, \cdots, x_Q^2, x_1x_2, \cdots, x_{Q-1}x_Q$ are 2nd-order (2 kinds)

Orthogonalization can proceed as before for any $W(\vec{x}) \geq 0$, for functions $f(\vec{x})$ with $\int_{R^Q} |f(\vec{x})|^2 W(\vec{x})\, d\vec{x} < \infty$.

Proceed order-by-order.

$W(\vec{x})$ is the distribution of inputs, i.e., a multivariate dist. over $x_1, \cdots, x_Q$.

If $W(\vec{x}) = W_0(x_1) W_0(x_2) \cdots W_0(x_Q)$, then

$$\langle \vec{x}^{\vec{k}}, \vec{x}^{\vec{l}} \rangle = \int \vec{x}^{\vec{k}+\vec{l}}\, W(\vec{x}) d\vec{x}$$

$$= \left(\int x_1^{k_1+l_1} W(x_1) dx_1\right) \cdots \int x_Q^{k_Q+l_Q} W(x_Q) dx_Q$$

$$= M_{k_1+l_1} \cdots M_{k_Q+l_Q}$$ and we immediately have

$$\varphi_{\vec{k}} = \varphi_{k_1}(x_1)\, \varphi_{k_2}(x_2) \cdots \varphi_{k_Q}(x_Q).$$

For Gaussian case:

$$\varphi_0 = 1$$

$$\varphi_{0\cdots 1 \cdots 0}(\vec{x}) = \varphi_1(x_k) = x_k$$

$$\varphi_{0\cdots 2 \cdots 0}(\vec{x}) = \varphi_2(x_k) = x_k^2 - P, \quad |\varphi|^2 = 2P^2$$

$$\varphi_{0\cdots 1 \cdots 1 \cdots 0}(\vec{x}) = \varphi_1(x_k)\varphi_1(x_l) = x_1 x_2, \quad |\varphi|^2 = P^2$$

28

$$\varphi_{0\cdots 3\cdots 0}(\vec{x}) = \varphi_3(x_k) = x_k^3 - 3P x_k \quad , |\varphi|^2 = 6P^3$$

$$\varphi_{0\cdots \underset{k}{2}\cdots \underset{l}{1}\cdots 0}(\vec{x}) = \varphi_2(x_k)\varphi_1(x_l) = x_k^2 x_l - P x_l \, , |\varphi|^2 = 2P^3$$

$$\varphi_{0\cdots \underset{k}{1}\cdots \underset{l}{1}\cdots \underset{m}{1}\cdots 0}(\vec{x}) = \varphi_1(x_k)\varphi_1(x_l)\varphi_1(x_m) = x_k x_l x_m \; ; \quad |\varphi|^2 = P^3.$$

Does this make sense. $Q \Rightarrow \infty$? Yes, infinite number of basis elements at each order

$$f(x) = \sum a_{\vec{k}} \varphi_{\vec{k}} = a_0$$
$$+ \sum_k a_{0\cdots 1\cdots 0}\, \varphi_{0\cdots 1\cdots 0}(\vec{x})$$
$$+ \sum_k a_{0\cdots 2\cdots 0}\varphi_{0\cdots 2\cdots 0}(\vec{x}) + \sum_{k,l} a_{0\cdots 1\cdots 1\cdots 0}\varphi_{0\cdots 1\cdots 1\cdots 0}(\vec{x}) \cdots$$

$$= a_0 + \sum a_k x_k + \sum_k b_k (x_k^2 - P) + \sum_{k<l} c_{kl}\, x_k x_l + \cdots$$

$$b_k = a_{0\cdots 2\cdots 0}, \quad c_{k,l} = a_{0\cdots \underset{k}{1}\cdots \underset{l}{1}\cdots 0} \text{ etc.}$$

Each $a, b, \cdots$ obtained,

$$a_{\vec{k}} = \frac{\langle f(x), \varphi_{\vec{k}}(x) \rangle}{\langle \varphi_{\vec{k}}(x), \varphi_{\vec{k}}(x) \rangle}$$