

**STSCI 4780/5780**  
**Relationships between variables:**  
**Regression, 2**

Tom Lored, CCAPS & SDS, Cornell University

© 2022-03-31

# Agenda

- Recap of linear regression setup
- Connection to least squares
- Posterior normality
- Nonlinear curve fitting
  - ▶ Separable nonlinear models & Jaynes-Bretthorst algorithm
  - ▶ (Spectrum analysis)

# Simple normal linear regression

$$y_i = f(x_i; \theta) + \epsilon_i; \quad \epsilon_i \sim \text{Norm}(0; \sigma^2)$$

$$f(x; \theta) = \sum_{\alpha=1}^M A_{\alpha} g_{\alpha}(x)$$

- Parameters are  $M$  coefficients/amplitudes:  $\theta = \{A_{\alpha}\}$ ,  $\alpha = 1$  to  $M$
- Regression function is *linear wrt  $A_{\alpha}$*  (not necessarily wrt  $x$ !)
- $M$  *basis functions*  $g_{\alpha}(x)$ 
  - Polynomials:  $\{1, x, x^2, \dots\}$  (or orthogonal polynomials)
  - Sinusoids/Fourier series:  $\{\sin(\omega x), \cos(\omega x), \dots\}$   
(with  $\omega$  fixed/known)
- PDFs for errors are *normal*, with *known*  $\sigma$

## Likelihood function

Abbreviating  $f_i = f(x_i; \{A_\alpha\}) = f_i(\{A_\alpha\})$ ,

$$\begin{aligned} p(\{y_i\}|\{x_i\}, \{A_\alpha\}) &= \frac{1}{\sigma^N (2\pi)^{N/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^N (y_i - f_i)^2 \right] \\ &= \frac{1}{\sigma^N (2\pi)^{N/2}} e^{-Q/2\sigma^2} \end{aligned}$$

$$\begin{aligned} Q(\{A_\alpha\}) &= \sum_{i=1}^N (y_i - f_i)^2 \\ &= \sum_{i=1}^N \left( y_i - \sum_{\alpha=1}^M A_\alpha g_{\alpha i} \right)^2 \quad \text{with } g_{\alpha i} \equiv g_\alpha(x_i) \\ &= \sum_{i=1}^N y_i^2 + \sum_{i=1}^N \left( \sum_{\alpha=1}^M A_\alpha g_{\alpha i} \right)^2 - 2 \sum_{i=1}^N y_i \sum_{\alpha=1}^M A_\alpha g_{\alpha i} \end{aligned}$$

## Vector notation

Eliminate Roman (data) indices by denoting such quantities as  $N$ -vectors:  $\vec{f} = [f_1, \dots, f_N]^T$ , etc.

Model expresses  $\vec{f}$  as a sum of  $M$  basis vectors:

$$\vec{y} = \vec{f}(\{A_\alpha\}) + \vec{\epsilon}; \quad \vec{f}(\{A_\alpha\}) = \sum_{\alpha=1}^M A_\alpha \vec{g}_\alpha$$

Quadratic form is the squared magnitude of the misfit vector:

$$\begin{aligned} Q(\{A_\alpha\}) &= \left[ \vec{y} - \vec{f}(\{A_\alpha\}) \right]^2 \\ &= y^2 + f^2 - 2\vec{y} \cdot \vec{f} \\ &= y^2 + \sum_{\alpha\beta} A_\alpha A_\beta \vec{g}_\alpha \cdot \vec{g}_\beta - 2 \sum_{\alpha} A_\alpha \vec{y} \cdot \vec{g}_\alpha \end{aligned}$$

## Posterior mode

Adopt a flat prior; the posterior mode (the MAP estimate—“maximum a posteriori”) is then the maximum likelihood estimate (MLE), which satisfies (for  $\gamma = 1$  to  $M$ )

$$\left. \frac{\partial Q}{\partial A_\gamma} \right|_{A=\hat{A}} = 2 \sum_{\beta} \hat{A}_\beta \vec{g}_\beta \cdot \vec{g}_\gamma - 2 \vec{y} \cdot \vec{g}_\gamma = 0$$

Let  $\hat{\vec{f}} \equiv \sum_{\beta} \hat{A}_\beta \vec{g}_\beta$  (function estimate at the mode); then

$$\hat{\vec{f}} \cdot \vec{g}_\gamma = \vec{y} \cdot \vec{g}_\gamma$$

*The modal model is the one whose projection on each basis function matches the data's projection on each basis function*

In terms of the  $M \times M$  **model metric matrix**,  $\eta_{\alpha\beta} \equiv \vec{g}_\alpha \cdot \vec{g}_\beta$ ,

$$\sum_{\beta} \eta_{\gamma\beta} \hat{A}_\beta = \vec{y} \cdot \vec{g}_\gamma$$

## Regression geometry

Sums of squares in normal-based likelihood exponents makes normal linear regression look like Euclidean geometry in  $N$ -D space, with projections into the  $M$ -D subspace spanned by the model basis

The metric generalizes the Pythagorean theorem to non-orthonormal coordinates

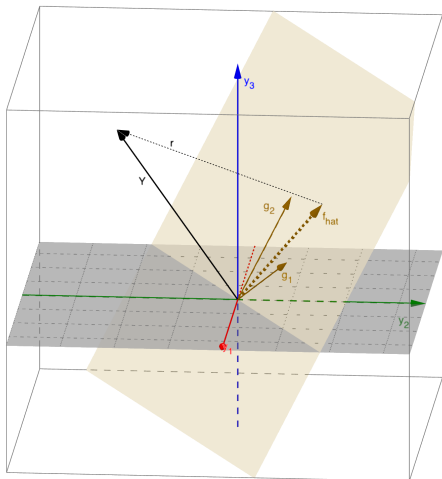
Geometry for linear regression,  
 $M = 2$  bases,  $N = 3$  samples

$$\vec{x} = [0, 1, 2]^T; \quad \vec{y} = [3, -2, 4]^T$$

$$f(x) = A_1 + A_2 x$$

$$g_1(x) = 1 \rightarrow \vec{g}_1 = [1, 1, 1]^T$$

$$g_2(x) = x \rightarrow \vec{g}_2 = [0, 1, 2]^T$$



## Metric and mode equation

To solve for the mode, define the matrix  $\eta_{\alpha\beta}$ :

$$\eta_{\alpha\beta} \equiv \vec{g}_\alpha \cdot \vec{g}_\beta = \eta_{\beta\alpha}$$

This is a symmetric matrix; it plays the role of a metric on the  $M$ -dimensional subspace spanned by the model functions

The mode condition is now (switch  $\gamma$  to  $\alpha$  above, & use symmetry)

$$\sum_{\beta} \eta_{\alpha\beta} \hat{A}_{\beta} = \vec{y} \cdot \vec{g}_{\alpha}$$

The LHS describes the product of an  $M \times M$  matrix and a column vector of  $M$  components; the RHS comprises a vector with  $M$  components—this is just a matrix equation (in the  $M$ -D model space, not the  $N$ -D sample space)

$$\Rightarrow \hat{A}_{\alpha} = \sum_{\beta} [\eta^{-1}]_{\alpha\beta} \vec{y} \cdot \vec{g}_{\beta}$$

(numerically, the most stable solvers backsolve rather than invert)



### *Aside on metrics*

A metric defines dot products in terms of coordinates in an arbitrary (e.g., non-orthonormal) basis:

$$\vec{v}_1 = \sum_{\alpha} a_{\alpha} \vec{g}_{\alpha}$$

$$\vec{v}_2 = \sum_{\beta} b_{\beta} \vec{g}_{\beta}$$

$$\begin{aligned} \Rightarrow \vec{v}_1 \cdot \vec{v}_2 &= \sum_{\alpha} \sum_{\beta} a_{\alpha} b_{\beta} \vec{g}_{\alpha} \cdot \vec{g}_{\beta} \\ &= \sum_{\alpha} \sum_{\beta} a_{\alpha} b_{\beta} \eta_{\alpha\beta} \end{aligned}$$

(If the  $\vec{g}_{\alpha}$  were orthogonal, only the  $\alpha = \beta$  terms would be nonzero)

Key use is finding distance from coordinate differences:

$$\begin{aligned}d_{12}^2 &= |\vec{b} - \vec{a}|^2 \\&= \left[ \sum_{\alpha} (b_{\alpha} - a_{\alpha}) \vec{g}_{\alpha} \right] \cdot \left[ \sum_{\beta} (b_{\beta} - a_{\beta}) \vec{g}_{\beta} \right] \\&= \sum_{\alpha} \sum_{\beta} \Delta_{\alpha} \Delta_{\beta} \eta_{\alpha\beta}\end{aligned}$$

with coordinate differences  $\Delta_{\alpha} \equiv b_{\alpha} - a_{\alpha}$

In an *orthonormal* basis,  $\eta_{\alpha\beta} = \delta_{\alpha\beta}$ , the identity matrix; in this case

$$d_{12}^2 = \sum_{\alpha} \Delta_{\alpha}^2,$$

i.e., the *Pythagorean theorem*

*Metrics generalize the Pythagorean theorem to non-orthonormal coordinate systems*

## Connections to least squares estimation

For a flat prior and fixed  $\sigma$ , the posterior mode minimizes

$$Q(\{A_\alpha\}) = \sum_{i=1}^N [y_i - f_i(\{A_\alpha\})]^2$$

→ the flat-prior mode gives the *least squares estimates of the amplitudes*

The  $N \times M$  matrix of model vector coordinates  $[g_{\alpha i}]^T$  is the *design matrix*; it is often denoted  $\mathbf{X} = X_{i\alpha}$ , even though it consists of *response* values (the model basis in the  $y$  space—functions of  $x_i$ s)

The  $M \times M$  metric

$$\eta_{\alpha\beta} \equiv \vec{g}_\alpha \cdot \vec{g}_\beta = \sum_i g_{\alpha i} g_{\beta i} = \mathbf{X}^T \mathbf{X}$$

is sometimes called the *Gram matrix* or *Gramian matrix*

The mode condition

$$\sum_{\beta} \eta_{\alpha\beta} \hat{A}_\beta = \vec{y} \cdot \vec{g}_\alpha$$

is a set of  $M$  equations called the *normal equations* when expressed in terms of the design matrix

## Posterior is multivariate normal

Write  $A_\alpha = \hat{A}_\alpha(\vec{x}, \vec{y}) + \delta A_\alpha$  (expressing  $A_\alpha$ 's in terms of  $\delta A_\alpha$ 's); then

$$\begin{aligned} Q(\{A_\alpha\}) &= (\vec{y} - \vec{f})^2 = \left( \vec{y} - \sum_{\alpha} \hat{A}_\alpha \vec{g}_\alpha - \sum_{\beta} \delta A_\beta \vec{g}_\beta \right)^2 \\ &= \left( \vec{y} - \sum_{\alpha} \hat{A}_\alpha \vec{g}_\alpha \right)^2 + \left( \sum_{\alpha} \delta A_\alpha \vec{g}_\alpha \right)^2 \\ &\quad - 2 \left( \sum_{\beta} \delta A_\beta \vec{g}_\beta \right) \cdot \left( \vec{y} - \sum_{\alpha} \hat{A}_\alpha \vec{g}_\alpha \right) \\ &= Q_{\min} + \left( \sum_{\alpha} \delta A_\alpha \vec{g}_\alpha \right) \cdot \left( \sum_{\beta} \delta A_\beta \vec{g}_\beta \right) \\ &\quad - 2 \sum_{\beta} \delta A_\beta \left( \vec{g}_\beta \cdot \vec{y} - \sum_{\alpha} \hat{A}_\alpha \eta_{\alpha\beta} \right) \quad \leftarrow \text{mode cond'n} \end{aligned}$$

$$\begin{aligned}
Q(\{A_\alpha\}) &= Q_{\min} + \sum_{\alpha} \sum_{\beta} \delta A_{\alpha} \delta A_{\beta} \eta_{\alpha\beta} \\
&= r^2 + \sum_{\alpha} \sum_{\beta} (A_{\alpha} - \hat{A}_{\alpha}) \eta_{\alpha\beta} (A_{\beta} - \hat{A}_{\beta})
\end{aligned}$$

where  $\vec{r} \equiv \vec{y} - \hat{\vec{f}}$  is the *residual vector* between the data and the best-fit model

The posterior is thus a multivariate normal (MVN) distribution for  $A = \{A_{\alpha}\}$ :

$$\begin{aligned}
p(\{A_{\alpha}\} | \vec{y}, \sigma) &\propto \frac{1}{\sigma^N} \exp \left[ -\frac{Q(\{A_{\alpha}\})}{2\sigma^2} \right] \\
&\propto \frac{1}{\sigma^N} \exp \left[ -\frac{r^2}{2\sigma^2} \right] \exp \left[ -\frac{1}{2} (A - \hat{A}) \cdot \mathbf{V}^{-1} \cdot (A - \hat{A}) \right]
\end{aligned}$$

MVN for the  $A$ 's with (marginal) means  $\hat{A}_{\alpha}$ , inverse covariance matrix  $\mathbf{V}^{-1} = \boldsymbol{\eta}/\sigma^2$ , and covariance matrix

$$\mathbf{V} = \sigma^2 \boldsymbol{\eta}^{-1}$$

# Consequences of posterior normality

## *Joint HPD regions for coefficients*

Write  $p(A|\vec{y}, \sigma) = C e^{-\chi^2/2}$  with

$$\chi^2(A) = \frac{Q(A)}{\sigma^2} = \frac{r^2}{\sigma^2} + \Delta\chi^2(A),$$

$$\text{with } \Delta\chi^2(A) \equiv (A - \hat{A}) \cdot \mathbf{V}^{-1} \cdot (A - \hat{A})$$

$$\Rightarrow p(A|\vec{y}, \sigma) \propto e^{-\Delta\chi^2(A)^2/2}$$

An HPD region with probability  $C$  is bounded by a surface of constant density, i.e., a surface of constant  $\Delta\chi^2(A) = \Delta\chi_{\text{crit}}^2$ , chosen so

$$C = \int_{\Delta\chi^2 < \Delta\chi_{\text{crit}}^2} d^M A p(A|\vec{y}, \sigma)$$

Normality  $\rightarrow$  choose  $\Delta\chi_{\text{crit}}^2$  so that  $C$  is the probability that  $\chi^2 < \Delta\chi_{\text{crit}}^2$  in the  $\chi^2$  distribution with  $M$  degrees of freedom

## Uncertain $\sigma$

Adopt a log-flat prior for  $\sigma$ , i.e.,  $p(\sigma) \propto 1/\sigma$ ; then as a function of  $\sigma$ , the posterior is

$$p(\sigma, A|\vec{y}) \propto \frac{1}{\sigma^{N+1}} e^{-Q(A)/2\sigma^2}$$

Marginalize over  $\sigma$  just as we did for normal  $(\mu, \sigma)$  inference; this gives

$$p(\sigma, A|\vec{y}) \propto \left[ 1 + \frac{\Delta Q(A)}{r^2} \right]^{-N/2},$$

where  $\Delta Q(A) \equiv (A - \hat{A}) \cdot \boldsymbol{\eta} \cdot (A - \hat{A})$

This is a *multivariate Student's t distribution*

## Marginalizing over coefficients

- Marginalizing over a *subset* of coefficients is straightforward using the fact that MVN conditional distributions are normal with fixed variance; can show the marginal is proportional to the *profile likelihood*
- Marginalizing over *all* coefficients can be done analytically by *diagonalizing the metric*  $\rightarrow$  the MVN normalization constant is  $\sqrt{\det \mathbf{V}}$

## Conjugate priors

Since the likelihood function is MVN with respect to  $A$ , a MVN prior for  $A$  is a *conjugate prior*, resulting in a posterior that remains MVN



## Heteroskedastic cases

If the errors are correlated rather than IID, then the quadratic form is a double sum using the noise covariance matrix,  $\mathbf{E}$ :

$$Q(A) = \sum_{i,j} [y_i - f_i(A)][\mathbf{E}^{-1}]_{ij}[y_j - f_j(A)]$$

A special case is independent but *heteroskedastic* errors, for which

$$Q(A) = \sum_i \frac{[y_i - f_i(A)]^2}{\sigma_i^2}$$

This corresponds to *weighted least squares* or *minimum  $\chi^2$*  fitting of linear models (with full  $\mathbf{E}$  it's *generalized LS*)

These generalizations can be easily accommodated simply by (1) removing  $\sigma^2$  everywhere, and (2) *redefining vector dot products* using  $\mathbf{E}$  as an  $N \times N$  metric on the  $N$ -D space of  $y_i$  coordinates:

$$\vec{a} \cdot \vec{b} \equiv \sum_{ij} a_i [\mathbf{E}^{-1}]_{ij} b_j$$

All previous results are *algebraically unchanged* by this (but marginalization over  $\sigma$  is no longer relevant)

# Bayesian nonlinear curve fitting & least squares

## Setup

Data  $D = \{y_i\}$  are measurements of an underlying function  $f(x; \theta)$  at  $N$  sample points  $\{x_i\}$ . Let  $f_i(\theta) \equiv f(x_i; \theta)$ :

$$y_i = f_i(\theta) + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma_i^2)$$

We seek learn  $\theta$ , or to compare different functional forms (model choice,  $M$ ).

## Likelihood

$$\begin{aligned} p(D|\theta, M) &= \prod_{i=1}^N \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{y_i - f_i(\theta)}{\sigma_i} \right)^2 \right] \\ &\propto \exp \left[ -\frac{1}{2} \sum_i \left( \frac{y_i - f_i(\theta)}{\sigma_i} \right)^2 \right] \\ &= \exp \left[ -\frac{\chi^2(\theta)}{2} \right] \end{aligned}$$

## Posterior

For prior density  $\pi(\theta)$ ,

$$p(\theta|D, M) \propto \pi(\theta) \exp \left[ -\frac{\chi^2(\theta)}{2} \right]$$

If you have a least-squares or  $\chi^2$  code:

- Think of  $\chi^2(\theta)$  as  $-2 \log \mathcal{L}(\theta)$ .
- Bayesian inference amounts to exploration and numerical integration of  $\pi(\theta)e^{-\chi^2(\theta)/2}$ .
- If noise level is uncertain, keep the  $1/\sigma_i$  factors (dropped above!) and include noise parameters in inference (e.g., scale all  $\sigma_i$  by a parameter,  $\alpha$ )
- If any of the parameters appear *linearly*, our linear regression results show that their likelihood function—conditional on the remaining parameters—will be MVN  $\rightarrow$  analytical simplifications

# Important Case: Separable Nonlinear Models

A (linearly) separable model has parameters  $\theta = (A, \psi)$ :

- Linear amplitudes  $A = \{A_\alpha\}$
- Nonlinear parameters  $\psi$

$f(x; \theta)$  is a linear superposition of  $M$  nonlinear components  $g_\alpha(x; \psi)$ :

$$y_i = \sum_{\alpha=1}^M A_\alpha g_\alpha(x_i; \psi) + \epsilon_i$$

or

$$\vec{y} = \sum_{\alpha} A_\alpha \vec{g}_\alpha(\psi) + \vec{\epsilon}.$$

Recall: “linear/nonlinear” refers to how the predictions depend on the *parameters*, not how they depend on the predictors/covariates!

## Examples

Polynomials (simple or orthogonal);  $\psi = \emptyset$  (linear regr'n):

$$\begin{aligned}f(x) &= A_0 + A_1x + A_2x^2 + A_3x^3 \\&= A_0 + A_1x + A'_2(2x^2 - 1) + A'_3(4x^3 - 3x), \quad x \in [-1, 1] \\&= A_0g_0(x) + A_1g_1(x) + A'_2g_2(x) + A'_3g_3(x)\end{aligned}$$

Sinusoids;  $\psi = \omega$ :

$$\begin{aligned}f(x) &= A \cos(\omega x + \phi) \\&= A \cos \phi \cos \omega x - A \sin \phi \sin \omega x \\&= A_1 \cos \omega x + A_2 \sin \omega x \quad (\text{defines } A_1, A_2) \\&= A_1g_1(x, \omega) + A_2g_2(x, \omega)\end{aligned}$$

Chirps;  $\psi = (\omega, \alpha)$ :

$$\begin{aligned}f(x) &= A \cos(\alpha x^2 + \omega x + \phi), \quad \text{inst. freq.} = \omega + 2\alpha x \\&= A_1 \cos(\alpha x^2 + \omega x) + A_2 \sin(\alpha x^2 + \omega x)\end{aligned}$$

Exponentials;  $\psi = (\tau_1, \tau_2, \dots)$ :  $f(x) = A_1e^{-x/\tau_1} + A_2e^{-x/\tau_2} + \dots$

# The Jaynes-Bretthorst Algorithm

Why separable structure is important: You can marginalize over  $A$  *analytically*  $\rightarrow$  *Jaynes-Bretthorst algorithm* ( "Bayesian Spectrum Analysis & Param. Estimation" (1988), avbl. at Probability Theory As Extended Logic )

Algorithm is closely related to linear least squares, diagonalization (eigenvectors/values), and SVD

Goals:

- Estimate the nonlinear parameters  $\psi$
- Estimate amplitudes
- Compare rival models

*Supplementary material:  
Orthogonalization & the Jaynes-Bretthorst algorithm;  
Relationship to Fourier spectral analysis*

The log-likelihood is a quadratic form in  $A_\alpha$ ,

$$\mathcal{L}(A, \psi) \propto \frac{1}{\sigma^N} \exp \left[ -\frac{Q(A, \psi)}{2\sigma^2} \right]$$

$$\begin{aligned} \text{with } Q &= \left[ \vec{y} - \sum_{\alpha} A_{\alpha} \vec{g}_{\alpha} \right]^2 \\ &= \left[ \vec{y} - \sum_{\alpha} A_{\alpha} \vec{g}_{\alpha} \right] \cdot \left[ \vec{y} - \sum_{\beta} A_{\beta} \vec{g}_{\beta} \right] \\ &= y^2 - 2 \sum_{\alpha} A_{\alpha} \vec{y} \cdot \vec{g}_{\alpha} + \sum_{\alpha, \beta} A_{\alpha} A_{\beta} \eta_{\alpha\beta} \end{aligned}$$

where, as before,  $\eta_{\alpha\beta}(\psi) = \vec{g}_{\alpha}(\psi) \cdot \vec{g}_{\beta}(\psi)$



We seek to integrate out the amplitudes, but completing the square is complicated because of the nontrivial metric  $\eta_{\alpha\beta}$

Change the basis for  $\vec{f}$  from  $\vec{g}_\alpha$  to an *orthonormal basis*  $\vec{h}_\mu$ :

$$\vec{g}_\alpha = \sum_{\mu} a_{\alpha\mu} \vec{h}_\mu \quad \text{with } \vec{h}_\mu \cdot \vec{h}_\nu = \delta_{\mu\nu}$$

which implies  $\vec{h}_\mu = \sum_{\alpha} (a^{-1})_{\mu\alpha} \vec{g}_\alpha$ . Note  $a = a(\psi)$ .

Rewriting  $\vec{f}$ ,

$$\vec{f}(\theta) = \sum_{\alpha=1}^M A_{\alpha} \vec{g}_{\alpha}(\psi) = \sum_{\mu=1}^M B_{\mu}(A, \psi) \vec{h}_{\mu}(\psi)$$

with orthonormal amplitudes  $B_{\mu}(A, \psi) = \sum_{\alpha} A_{\alpha} a_{\alpha\mu}(\psi)$

Some linear algebra shows that  $\eta = aa^T$ , so we can get  $a$  from  $\eta$  via Cholesky/eigen/QR decomposition.

Now write the quadratic form in terms of the  $B$ s instead of the  $A$ s:

$$\begin{aligned}
 Q &= y^2 - 2 \sum_{\alpha} A_{\alpha} \vec{y} \cdot \vec{g}_{\alpha} + \sum_{\alpha, \beta} A_{\alpha} A_{\beta} \eta_{\alpha\beta} \\
 &= y^2 - 2 \sum_{\mu} B_{\mu} \vec{y} \cdot \vec{h}_{\mu} + \sum_{\mu} B_{\mu}^2 \\
 &= \sum_{\mu} \left[ B_{\mu} - \hat{B}_{\mu}(\psi) \right]^2 + r^2(\psi)
 \end{aligned}$$

with  $\hat{B}_{\mu}(\psi) \equiv \vec{y} \cdot \vec{h}_{\mu}(\psi)$  and the residual  $\vec{r}(\psi) \equiv \vec{y} - \sum_{\mu} \hat{B}_{\mu} \vec{h}_{\mu}$

The posterior in terms of  $B$ s is

$$p(B, \psi | D, I) \propto \frac{\pi(\psi) J(\psi)}{\sigma^N} \exp \left[ -\frac{r^2(\psi)}{2\sigma^2} \right] \exp \left[ \frac{-1}{2\sigma^2} \sum_{\mu} [B_{\mu} - \hat{B}_{\mu}(\psi)]^2 \right]$$

$J(\psi) = (\det \eta)^{1/2}$  comes from changing variables from  $A$ s to  $B$ s

Marginalize  $B$ 's analytically (*check the range!*):

$$p(\psi|D, I) \propto \frac{\pi(\psi)J(\psi)}{\sigma^{N-M}} \exp \left[ -\frac{r^2(\psi)}{2\sigma^2} \right]$$

If  $\sigma$  unknown, marginalize using  $p(\sigma|I) \propto \frac{1}{\sigma}$ .

$$p(\psi|D, I) \propto \pi(\psi) J(\psi) [r^2(\psi)]^{\frac{M-N}{2}}$$

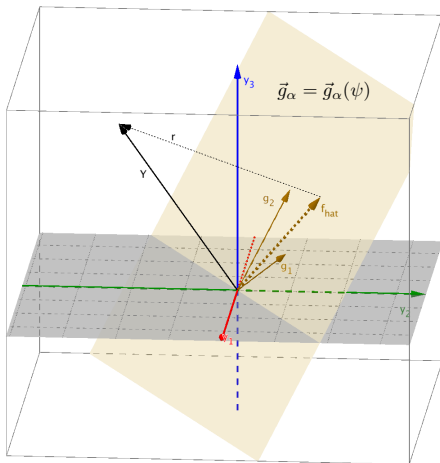
For given  $\psi$ ,  $r^2$  is just the residual sum of squares from a least squares fit to the basis functions. We can write

$$\begin{aligned} r^2(\psi) &= y^2 - \sum_{\mu} \hat{B}_{\mu}^2(\psi) \\ &= y^2 - S(\psi) \end{aligned}$$

with  $S(\psi) = \sum_{\mu} [\vec{y} \cdot \vec{h}_{\mu}(\psi)]^2$ , the sum of squared projections

## Regression geometry for separable models

The geometry is as for linear regression, but now the basis vectors (and the subspace they span) depends on the nonlinear parameters—here the model plane moves as a function of  $\psi$



## Application: Bayesian Spectrum Analysis

Adopt a sinusoid periodic signal model:

$$\begin{aligned}f(t) &= A \cos(\omega t - \phi) && \text{parameters } \omega, A, \phi \\&= A_1 \cos \omega t + A_2 \sin \omega t && \text{parameters } \omega, A_1, A_2 \\y_i &= f(t_i) + e_i && \text{Gaussian error pdfs; rms} = \sigma\end{aligned}$$

Estimate  $\omega$ :

$$\begin{aligned}p(\omega|D) &\propto \int dA_1 \int dA_2 p(\omega, A_1, A_2) \mathcal{L}(\omega, A_1, A_2) \\&\propto p(\omega) J(\omega) \exp \left[ \frac{S(\omega)}{\sigma^2} \right]\end{aligned}$$

- Equally-spaced samples:  $S(\omega) \rightarrow$  *Schuster periodogram* for large  $N$  (when  $\eta$  is nearly diagonal) — magnitude of the *discrete Fourier transform* (DFT) of the time series
- Unequally-spaced samples:  $S(\omega) \approx$  *Lomb-Scargle periodogram*