Parameter Selection Techniques:

Motivation:

1.) In some applications – e.g., neutronics – parameter dimension can be extremely large – e.g., $p = 10^6$

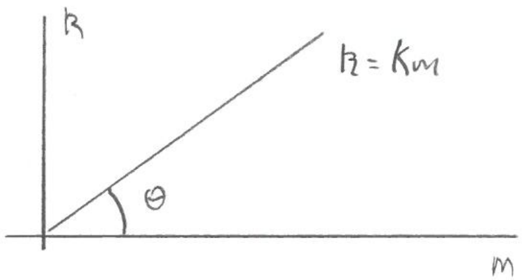
2) For many physical and biological applications, many parameters cannot uniquely be determined by data.

Example: $m \frac{d^2 z}{dt^2} + kz = 0$ **Solution:**

$z(0) = z_0, \; \frac{dz}{dt} = 0 \qquad z(t) = z_0 \cos\left(\sqrt{\frac{k}{m}}\, t\right)$

Note: Solution constant along lines $K = \frac{k}{m} \Leftrightarrow k = Km$

Admissible Parameter Space: $Q = (0,\infty) \times (0,\infty)$



- Determination of slope is equivalent to specifying Θ

Identifiable Subspace: $I(q) = \{\Theta = \arctan\left(\frac{k}{m}\right) \mid 0 < \Theta < \frac{\pi}{2}\}$

Nonidentifiable Subspace: $NI(q) = \{r = \sqrt{k^2 + m^2} \mid r > 0\}$

Note: $Q = I(q) \oplus \underbrace{NI(q)}_{\text{Orthogonal complement}}$

Definition: Consider

$$y = f(q), \quad q = [q_1, \ldots, q_p]$$

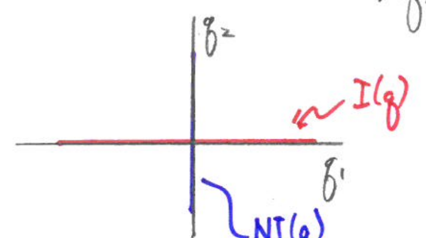
The parameters are identifiable at q^* if $f(q) = f(q^*)$ implies that $q = q^*$ for all admissible $q \in Q$. The parameters are identifiable with respect to a space $I(q)$, termed the identifiable subspace, if this holds for all $q^* \in I(q)$. The nonidentifiable subspace $NI(q)$ is the orthogonal complement of $I(q)$ wrt Q.

Examples:

1. Previous spring example
2. $q = [q_1, q_2]$ in $Q = \mathbb{R}^2$

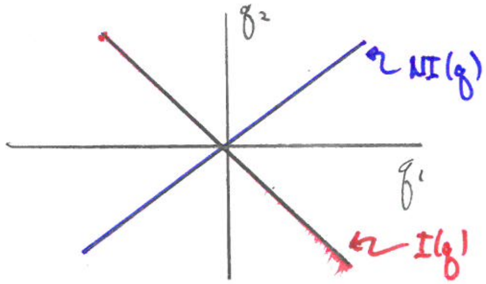
   i. Take $y = q_1$. Then $NI(q) = \{q_2 \in \mathbb{R}\}$

   $I(q) = \{q_1 \in \mathbb{R}\}$



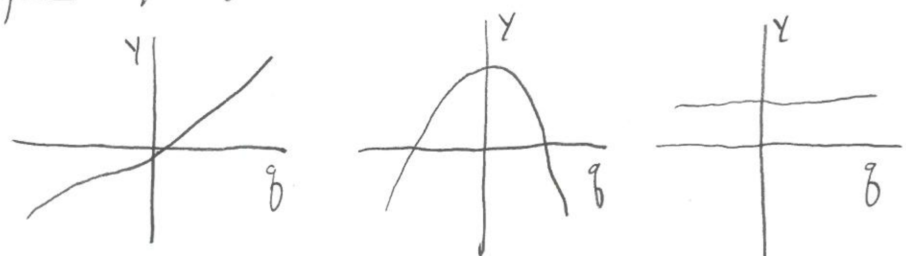
ii) $y = q_1 - q_2$ Then $NI(q) = \{(q_1, q_2) \in \mathbb{R}^2 \mid q_1 = q_2\}$

$$I(q) = \{(q_1, q_2) \in \mathbb{R}^2 \mid q_1 = -q_2\}$$



Definition: The parameters $q = [q_1, \dots, q_p]$ are noninfluential on space $NI(q)$ if $|f(q) - f(q^*)| < \varepsilon$ for all $q, q^* \in NI(q)$. The space $I(q)$ of influential parameters is the orthogonal complement of $NI(q)$ wrt Q and Euclidean inner product

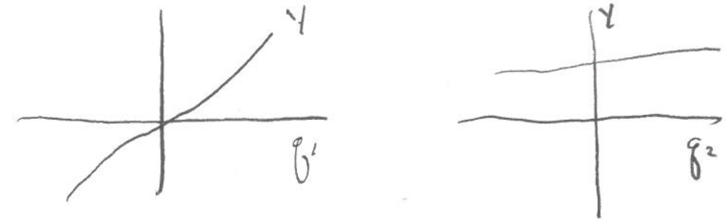
Examples: $y = f(q)$



Identifiable Un (non) - identifiable Noninfluential

Note: Non-identifiable & unidentifiable are synonymous

(ii)



Note: q_1 is more influential than q_2

Other Examples HIV and nuclear models from Lecture 1

Techniques: $y = f(q)$

1. Local sensitivity analysis – Based on derivatives $\frac{\partial y}{\partial q}$
2. Global sensitivity analysis – Quantifies how uncertainties in model outputs are apportioned to uncertainties in model input
3. Active subspace techniques based on QR or SVD

Note: 1 and 2 determine subsets of parameters whereas 3 determines subspace.

Local Sensitivity Analysis:

- We need to be able to approximate derivatives

$$s_i = \frac{\partial f}{\partial q_i}(q^*)$$

- Issues: - No quantification of uncertainties
  - Local at q^*

Example: Spring model

$$\frac{d^2z}{dt^2} + c\frac{dz}{dt} + kz = 0$$

$$z(0) = 2, \quad \frac{dz}{dt}(0) = -c$$

Displacement Observations: $y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} z(t) \\ \dot{z}(t) \end{bmatrix} = z$

Then

$$y(t) = 2e^{-ct/2} \cos\left(\sqrt{k - c^2/4} \cdot t\right)$$

Note: In general, we do not have analytic solutions!

Techniques to Compute Local Sensitivities: Section 7.3.1

0) Analytic
1) Sensitivity equations
2) Finite differences or complex step
3) Automatic differentiation

0) Analytic - Use symbolic package; Maple, Mathematica

$$\frac{\partial y}{\partial k} = \frac{-2t}{\sqrt{4k - c^2}} e^{-ct/2} \sin\left(\sqrt{k - c^2/4} \cdot t\right)$$

$$\frac{\partial y}{\partial c} = e^{-ct/2}\left[\frac{ct}{\sqrt{4k-c^2}} \sin\left(\sqrt{k - c^2/4} \cdot t\right) - t\cos\left(\sqrt{k - c^2/4} \cdot t\right)\right]$$

(iii)

Sensitivity Matrix: $q = (c, k)$

$$\chi(q) = \begin{bmatrix} \frac{\partial y}{\partial c}(t_1, q) & \frac{\partial y}{\partial k}(t_1, q) \\ \vdots & \vdots \\ \frac{\partial y}{\partial c}(t_n, q) & \frac{\partial y}{\partial k}(t_n, q) \end{bmatrix}_{n \times 2}$$

Fisher Information Matrix:

$$\mathcal{F} = \chi^T \chi \leftarrow 2 \times 2$$

Technique 1:

$$\frac{d}{dk}\left[\frac{d^2z}{dt^2} + c\frac{dz}{dt} + kz\right] = 0$$

$$\Rightarrow \frac{d^2z_k}{dt^2} + c\frac{dz_k}{dt} + z + kz_k = 0, \quad z_k = \frac{\partial z}{\partial k}$$

$$\Rightarrow \frac{d^2z_k}{dt^2} + c\frac{dz_k}{dt} + kz_k = -z, \quad z_k(0) = \frac{dz_k}{dt}(0) = 0$$

$$\frac{d^2z}{dt^2} + c\frac{dz}{dt} + kz = 0, \quad z(0) = \frac{dz}{dt}(0) = 0$$

Similarly,

$$\frac{d^2 z_c}{dt^2} + c\frac{dz_c}{dt} + Kz_c = -\frac{dz}{dt}, \quad z_c = \frac{\partial z}{\partial c}$$

$$z_c(0) = 0, \quad \frac{dz_c}{dt}(0) = -1$$

Technique 2: Finite difference

$$\frac{dy}{dK}(t) \simeq \frac{z(t, K+h_K, c) - z(t, K, c)}{h_K}$$

$$\frac{dy}{dc}(t) \simeq \frac{z(t, K, c+h_c) - z(t, K, c)}{h_c}$$

Issues: 1) Stepsizes h_K and h_c must reflect magnitudes of coefficients; e.g., $h_K = 10^{-6}|K|$

2) $\frac{\text{small}}{\text{small}}$ can be inaccurate

Solution: Complex steps - Consider

$$z = x + iy$$

$$f(z) = u(x,y) + iv(x,y)$$

For analytic f, the Cauchy-Riemann equations are

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}$$

Then

$$\frac{\partial u}{\partial x} = \lim_{h \to 0} \frac{v(x, y+h) - v(x,y)}{h}$$

$$= \lim_{h \to 0} \frac{\mathrm{Im}[f(x+i(y+h))] - \mathrm{Im}[f(x+iy)]}{h}$$

Since we have real-valued functions,

$$\mathrm{Im}(z) = y = 0$$

so

$$f(x) = u(x,0)$$

$$v(x,0) = \mathrm{Im}[f(z)] = 0$$

Thus

$$\frac{\partial f}{\partial x} = \lim_{h \to 0} \frac{\mathrm{Im}[f(x+ih)]}{h}$$

$$\Rightarrow \frac{\partial f}{\partial x} \simeq \frac{\mathrm{Im}[f(x+ih)]}{h}$$

Example: $f(x) = 3x \Rightarrow \frac{df}{dx} = 3$

Now

$$\frac{df}{dx} \simeq \frac{\mathrm{Im}[f(x+ih)]}{h} = \frac{\mathrm{Im}[3x+3ih]}{h} = \frac{3h}{h} = 3$$

<u>Technique 3</u>: Automatic differentiation - perform differentiation of basic operations - addition, subtraction, multiplication, division, composition - at compiler level.

- Good software for ODE, some for PDE

<u>Goal: Relate Sensitivities to Taylor Expansion:</u> Note that

$$f(t_i, q) \simeq f(t_i, \overset{\text{Optimal}}{q^*}) + \nabla_q f(t_i, q^*) \cdot \Delta q$$

where

$$\nabla_q f(t_i, q^*) = \left[\frac{\partial f}{\partial q_1}(t_i, q^*), \ldots, \frac{\partial f}{\partial q_p}(t_i, q^*)\right]$$

$$\Delta q = q - q^*$$

Consider the statistical model

$$v_i = f(t_i, q) + \varepsilon_i \quad , i = 1, \ldots, n,$$

where v_i are data values, and consider

$$J(q) = \frac{1}{n}\sum_{i=1}^{n} [v_i - f(t_i, q)]^2.$$

Then since $v_i \simeq f(t_i, q^*)$ at optimal q^*,

$$J(q) \simeq \frac{1}{n}\sum_{i=1}^{n} [\nabla_q f(t_i, q^*) \cdot \Delta q]^2$$

Recall

$$\chi(q^*) = \begin{bmatrix} \frac{\partial f}{\partial q_1}(t_1, q^*) & \cdots & \frac{\partial f}{\partial q_p}(t_1, q^*) \\ \vdots & & \vdots \\ \frac{\partial f}{\partial q_1}(t_n, q^*) & \cdots & \frac{\partial f}{\partial q_p}(t, q^*) \end{bmatrix}$$

Thus

$$J(q) \simeq \frac{1}{n}(\chi \Delta q)^T (\chi \Delta q)$$

or

$$J(q^* + \Delta q) \simeq \frac{1}{n} \Delta q^T \underbrace{\chi^T \chi}_{F} q$$

<u>Note:</u> Take Δq to be an eigenvector of $\chi^T \chi$ so

$$\chi^T \chi \Delta q = \lambda \Delta q$$

Thus

$$J(q^* + \Delta q) \simeq \frac{\lambda}{n} \|\Delta q\|_2^2$$

<u>Note:</u> $\lambda \simeq 0 \Rightarrow$ perturbations of $J(q^* + \Delta q) \simeq 0$

which implies that corresponding parameters are nonidentifiable

Parameter Subset Selection (PSS) Algorithm:

1. Set $n = p$ and threshold ε
2. Compute the eigenvalues $\lambda_1, \ldots, \lambda_n$ and eigenvectors $v_1, \ldots, v_n$ of $X^T X$ and order the eigenvalues by magnitude,

$$|\lambda_1| \leq |\lambda_2| \leq \cdots \leq |\lambda_n|$$

3. If $|\lambda_1| > \varepsilon$, stop
4. If $|\lambda_1| < \varepsilon$, one or more parameters is not identifiable
   - Identify component of v_1 with largest magnitude. This corresponds to least identifiable parameter.
   - Remove column of X that corresponds to this component and set $n = n-1$.