Statistics and Probability Theory

in

Civil, Surveying and Environmental Engineering

Prof. Dr. Michael Havbro Faber

Swiss Federal Institute of Technology

ETH Zurich, Switzerland

Contents of Todays Lecture

- **Basics of Reliability Analysis**

  - Short summary of previous lecture
  - The course at a glance
  - Failure events and basic random variables
  - Linear limit state functions and Normal distributed variables
  - Error propagation
  - Non-linear limit state functions
  - Monte-Carlo simulation

Summary of Previous Lecture

- Testing for goodness of fit
  - The $\chi 2$ goodness of fit test
  - The Kolmogorov-Smirnov goodness of fit test
- Model comparison

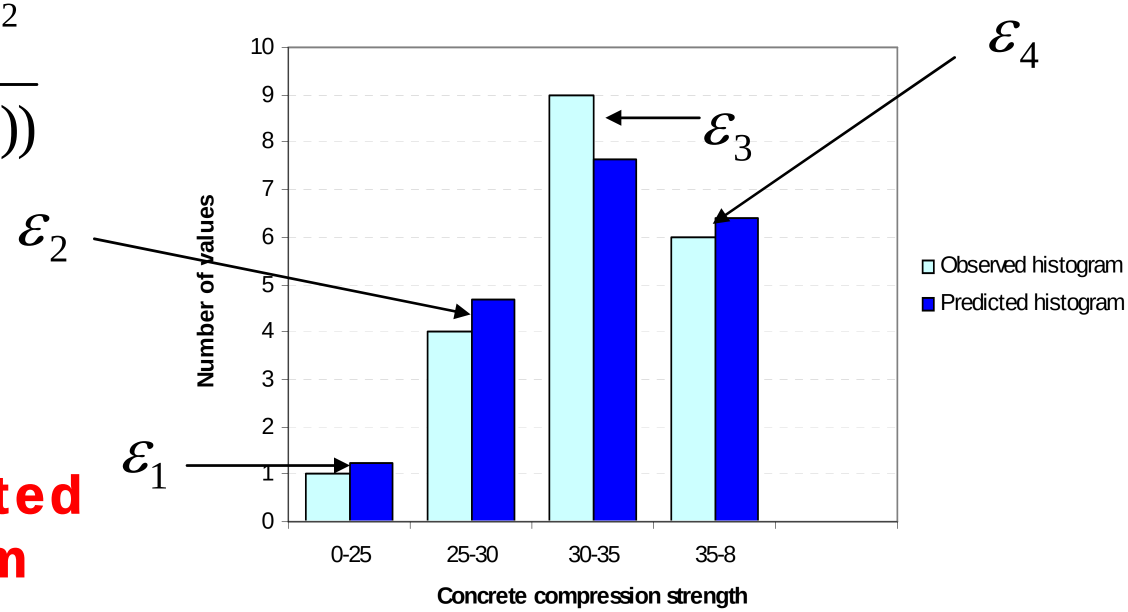
Summary of Previous Lecture

The CHI-square goodness of fit test

We test a ststistic constructed from the squared differences between the observed and the predicted histograms:

$$\varepsilon^2 = \sum_{i=1}^{k} \varepsilon_i^2 = \sum_{i=1}^{k} \frac{(N_{o,i} - N_{p,i})^2}{N_{p,i}(1 - p(x_i))}$$

$$\varepsilon_m{}^2 = \sum_{i=1}^{k} \frac{(N_{o,i} - N_{p,i})^2}{N_{p,i}}$$



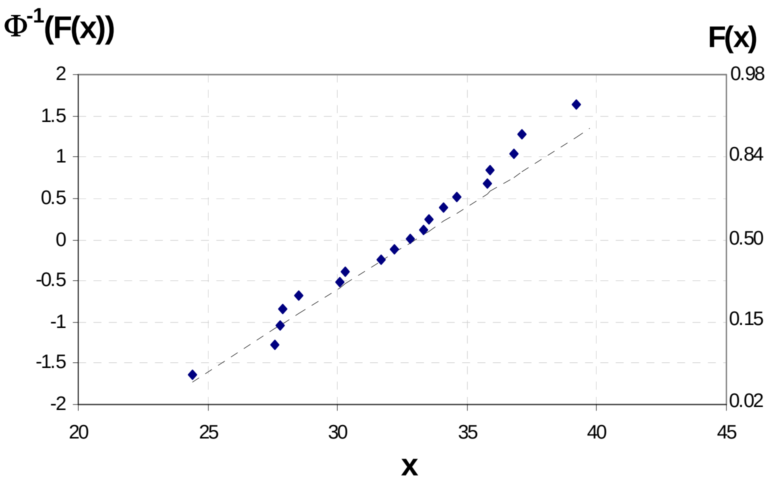
CHI-Square distributed
***k*-1 degree of freedom**

Summary of Previous Lecture

The Kolmogorov-Smirnov goodness of fit test

The observed cumulative distribution function may be calculated from:

$$F_o(x_i) = \frac{i}{n}$$



The following statistic is applied (tabularized):

$$\varepsilon_{\max} = \max_{i=1}^{n}\left[\left|F_o(x_i) - F_p(x_i)\right|\right] = \max_{i=1}^{n}\left[\left|\frac{i}{n} - F_p(x_i)\right|\right]$$

Summary of Previous Lecture

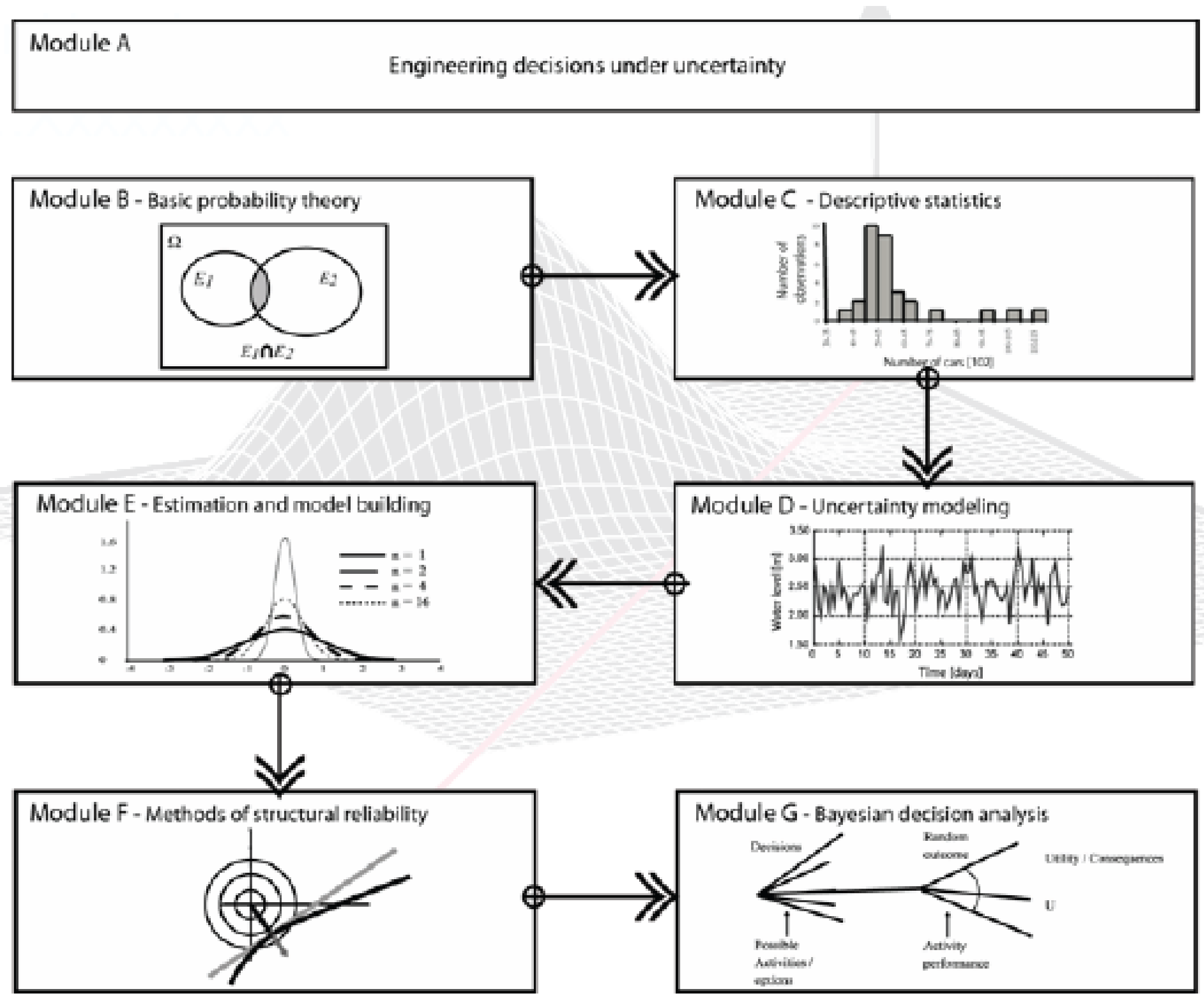
Model comparison

If testing of two different model hypothesis both fall out positive i.e. both models are plausible we can compare the goodness of fit of the two models either by

- comparing the sample statistics directly
  could be misleading / inconclusive due to different number of degrees of freedom

- comparing the sample likelihoods

The Course at a Glance



Basics of Reliability Analysis

- **Failure events and basic random variables**

By a failure event we associate in principle an event of special interest e.g. :

- Loss of functionality

- Costs

- Loss of lives

- Damage to the environment

Basics of Reliability Analysis

- **Failure events and basic random variables**

A failure event may conveniently be described in terms of a functional relationship

$$\mathbf{F} = \{ g(\mathbf{x}) \leq 0 \}$$

Such a functional relationship is denoted a limit state function

$$g(\mathbf{x})$$

Realizations of basic random variables

Basics of Reliability Analysis

- **The probability of an event**

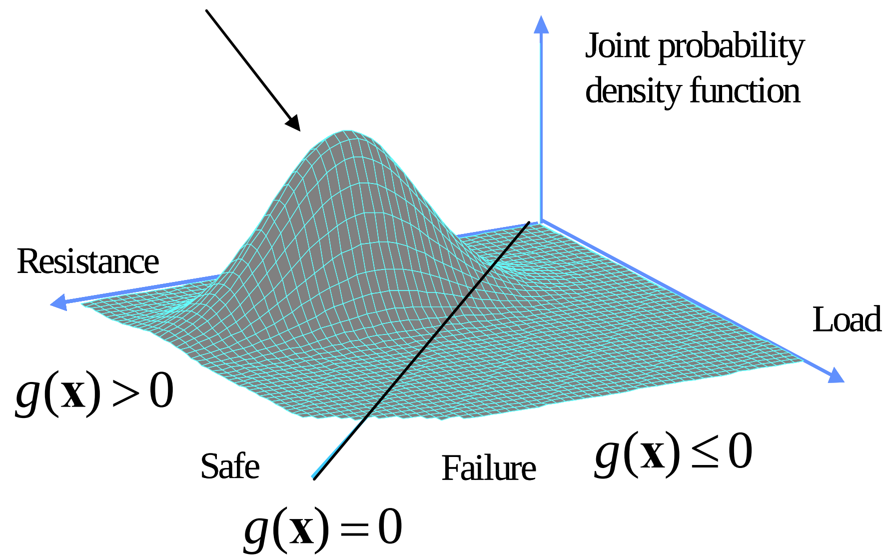
  The probability of an event e.g. a failure event can be calculated by the following integral

$$P_f = \int_{g(\mathbf{x})\leq 0} f_{\mathbf{X}}(\mathbf{x})d\mathbf{x}$$

Joint probability density function of the basic random variables X

$$g(\mathbf{x}) = r - s$$

r **: Resistance**
s **: Load**



Basics of Reliability Analysis

- **The probability of an event**

  The probability integral is in general non-trivial – can be multi-dimensional and can have a complicated integration domain

$$P_f = \int_{g(\mathbf{x}) \le 0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}$$

Classical nummerical integration techniques such as e.g. Simpson, Gauss or Schebyschev integration are not computationally efficient for dimensions larger than 5-6. Other apporaches are needed – which we will study further -

Basics of Reliability Analysis

- **Linear limit state functions and normal distributed basic variables**

First we consider the case where the limit state function is linear in the random variables and the random variables are normally distributed

$$g(x) = a_0 + \sum_{i=1}^{n} a_i x_i$$

For the case where the random variables X are normal distributed the safety margin M is also normal distributed

$$M = a_0 + \sum_{i=1}^{n} a_i X_i$$

$$\mu_M = a_0 + \sum_{i=1}^{n} a_i \mu_{X_i}$$

Correlation coefficient

$$\sigma_M^2 = \sum_{i=1}^{n} a_i^2 \sigma_{X_i}^2 + \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \rho_{ij} a_i a_j \sigma_{X_i} \sigma_{X_j}$$

Basics of Reliability Analysis

- **Linear limit state functions and normal distributed basic variables**

  The probability of failure is then determined as

$$P_F = P(g(\mathbf{X}) \leq 0) = P(M \leq 0)$$

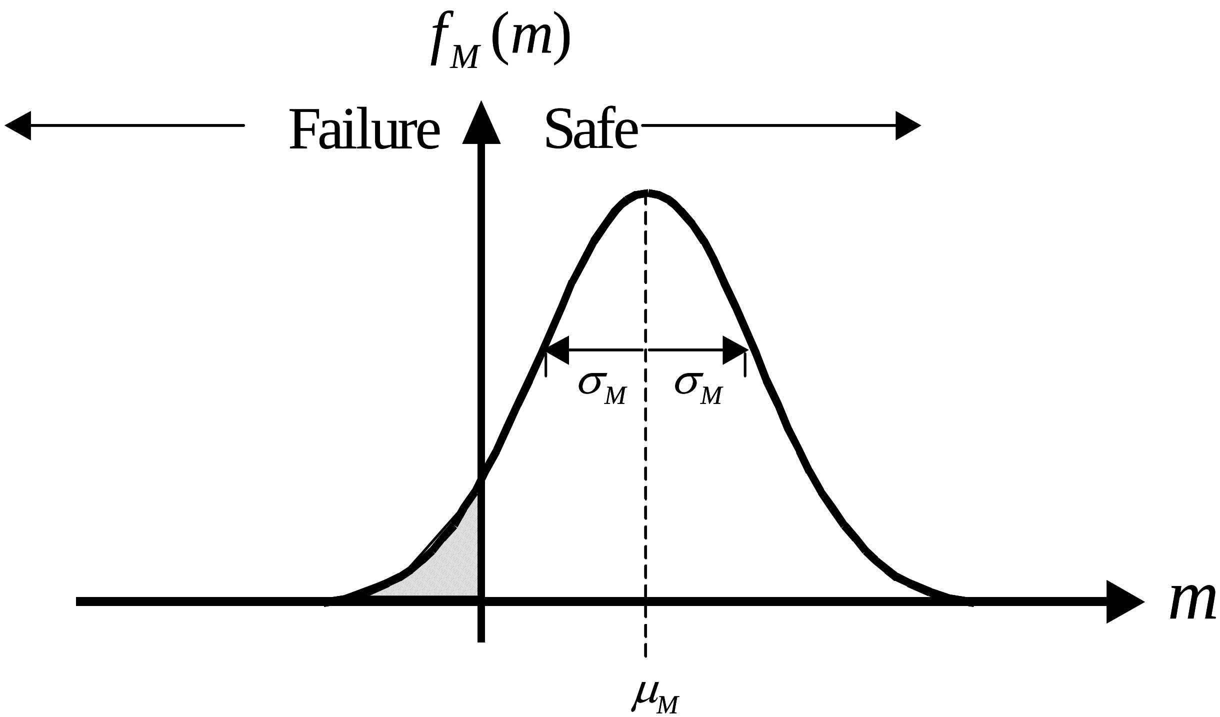
  Which reduces to the determination of the standard normal probability distribution function

$$P_F = \Phi\left(\frac{0-\mu_M}{\sigma_M}\right) = \Phi(-\beta) \qquad \textbf{with} \qquad \beta = \frac{\mu_M}{\sigma_M}$$

Reliability or safety index

Basics of Reliability Analysis

- **Linear limit state functions and normal distributed basic variables**



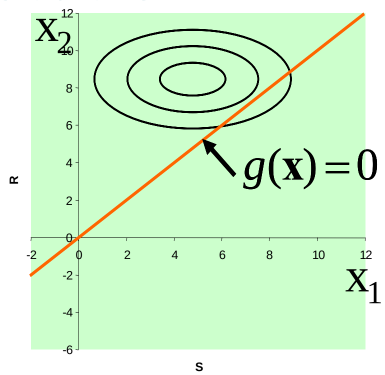
$$\beta = \frac{\mu_M}{\sigma_M}$$

The safety margin

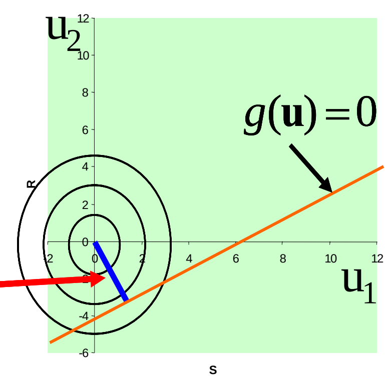
Basics of Reliability Analysis

- **Linear limit state functions and normal distributed basic variables**

  The reliability index β has a geometrical interpretation



β



$$U_i = \frac{X_i - \mu_{X_i}}{\sigma_{X_i}}$$

Smallest distance between the origin and the limit state function in standardized normal distributed space

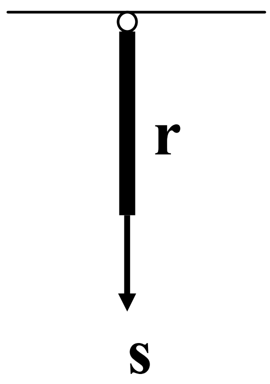
Zero mean and unit variance

Basics of Reliability Analysis

- **Linear limit state functions and normal distributed basic variables**

Example : Reliability of steel rod under tension loading

The resistance *R* and the max annual loading *S* are both assumed to be normal distributed



$$\mu_R = 350, \sigma_R = 35$$

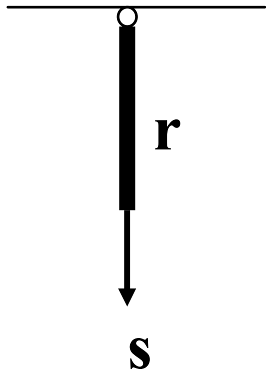
$$\mu_S = 200, \sigma_S = 40$$

Basics of Reliability Analysis

- **Linear limit state functions and normal distributed basic variables**

Example : Reliability of steel rod under tension loading

The safety margin is thus normal distributed with parameters



$$\mu_M = 350 - 200 = 150$$

$$\sigma_M = \sqrt{35^2 + 40^2} = 53.15$$

The reliability index β becomes

$$\beta = \frac{150}{53.15} = 2.84$$

$$P_F = \Phi(-2.84) = 2.4 \cdot 10^{-3}$$

Basics of Reliability Analysis

- **The error accumulation law**

  In many engineering applications the accumulation of errors is a central question

  Examples are :

  - errors due to fabrication tolerances of building components

  - errors in connection with surveying

  - errors in connection with measurements performed in the laboratory

Basics of Reliability Analysis

- **The error propagation law**

Assume that the error ε can be written as a differentiable function of random variables i.e. :

$$\varepsilon = h(\mathbf{x}) \qquad \mathbf{x} = (x_1, x_2, .., x_n)^T$$

Vector of realization of basic random variables

with parameters

$$\boldsymbol{\mu}_{\mathbf{X}} = (\mu_{X_1}, \mu_{X_2}, .., \mu_{X_n})^T \qquad Cov\left[X_i, X_j\right] = \rho_{ij}\sigma_{X_i}\sigma_{X_j}$$

Correlation coefficient

Standard deviation

The idea is to linearize *f(x)*

$$\varepsilon \cong h(\mathbf{x}_0) + \sum_{i=1}^{n}(x_i - x_{i,0}) \boxed{\left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\mathbf{x}_0}}$$

First order partial derivative taken in $\mathbf{x} = \mathbf{x}_0$

Basics of Reliability Analysis

- **The error propagation law**

If we linearize the error function around the mean value of the random variables its expected value and variance becomes :

$$\varepsilon \cong h(\boldsymbol{\mu}_{\mathbf{X}}) + \sum_{i=1}^{n} (x_i - \mu_{X_i}) \left.\frac{\partial h(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}}$$

$$E[\varepsilon] = h(\boldsymbol{\mu}_{\mathbf{X}})$$

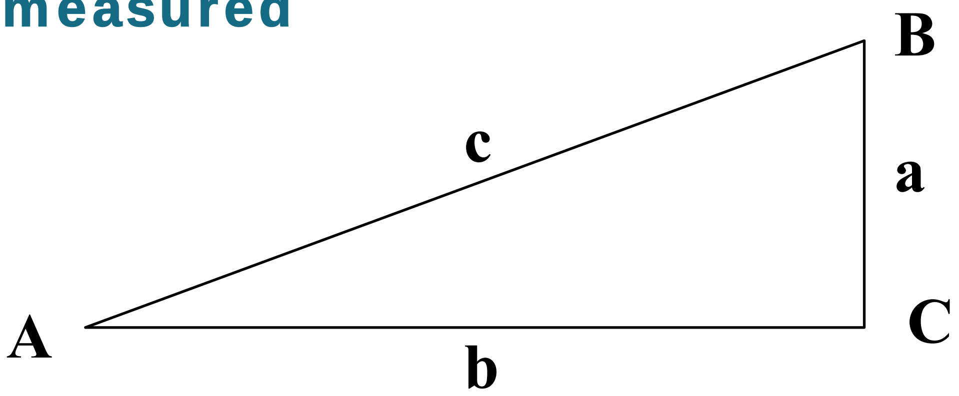
$$Var[\varepsilon] = \sum_{i=1}^{n} \left(\left.\frac{\partial h(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \right)^2 \sigma_{X_i}^2 + \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left(\left.\frac{\partial h(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \right) \left(\left.\frac{\partial h(\mathbf{x})}{\partial x_j}\right|_{\mathbf{x}=\boldsymbol{\mu}_{\mathbf{X}}} \right) \rho_{ij} \sigma_{X_i} \sigma_{X_j}$$

The mean value and the variance depends on the linearization point

Basics of Reliability Analysis

- Example : **Error propagation in measurements**

In order to estimate the length c i.e. the distance between the two points A and B the lengths a and b are measured



due to measurement uncertainty in assessing a and b also the length of c will be associated with uncertainty and it is of interest to know the probability that the length of c will exceed 13.5

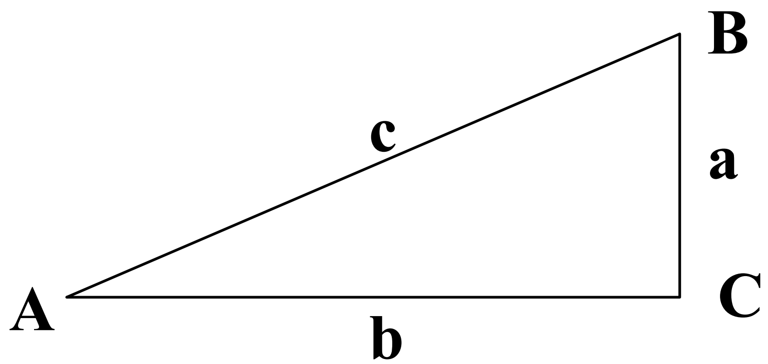
Basics of Reliability Analysis

- **Example : Error propagation in measurements**

It is assumed that a and b can be modeled as normal distributed random variables with parameters

$$\mu_a = 12.2 \quad \mu_b = 5.1$$

$$\sigma_a = 0.4 \quad \sigma_b = 0.3$$



Using that c can be given as

$$c = \sqrt{a^2 + b^2}$$

the statistical characteristics of c may be estimated through the error propagation law

Basics of Reliability Analysis

- **Example : Error propagation in measurements**

$$E[c] = \sqrt{\mu_a^2 + \mu_b^2}$$

$$Var[c] = \sum_{i=1}^{n} \left(\left. \frac{\partial h(\mathbf{x})}{\partial x_i} \right|_{\mathbf{x}=\boldsymbol{\mu}_\mathbf{X}} \right)^2 \sigma_{X_i}^2 = \frac{\mu_a}{\sqrt{\mu_a{}^2 + \mu_b{}^2}} \sigma_a^2 + \frac{\mu_b}{\sqrt{\mu_a{}^2 + \mu_b{}^2}} \sigma_b^2$$

$$E[c] = \sqrt{12.2^2 + 5.1^2} = 13.22$$

$$Var[c] = \frac{12.2}{\sqrt{12.2^2 + 5.1^2}} 0.4^2 + \frac{5.1}{\sqrt{12.2^2 + 5.1^2}} 0.3^2 = 0.1823$$

$$P_f = P(13.5 - C \leq 0) = \Phi(-\frac{(13.5 - 13.22)}{\sqrt{0.18}}) = 0.26$$

Basics of Reliability Analysis

- **Non-linear limit state functions**

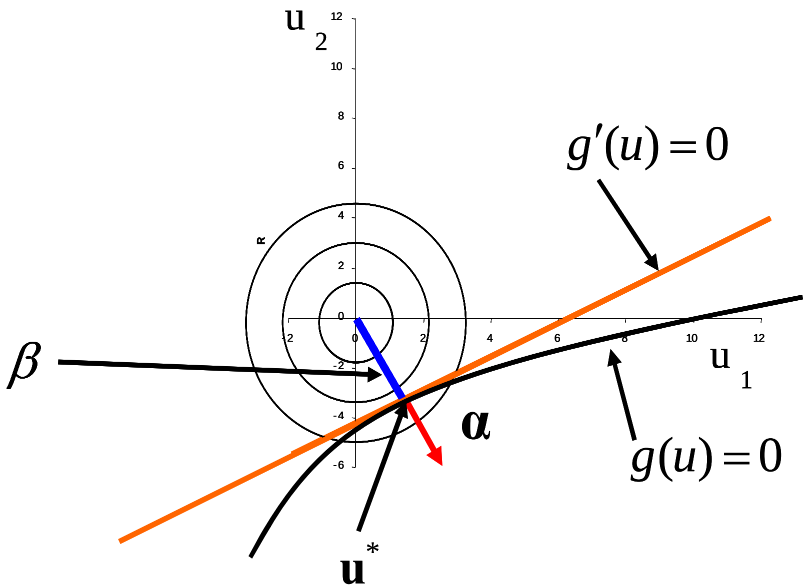
  Limit state functions are often non-linear

  As seen from the error propagation law it is possible to linearize such limit state functions but the results will depend on the linearization point and on the formulation of the limit state function

Basics of Reliability Analysis

- **Non-linear limit state functions**

Limit state functions are often non-linear



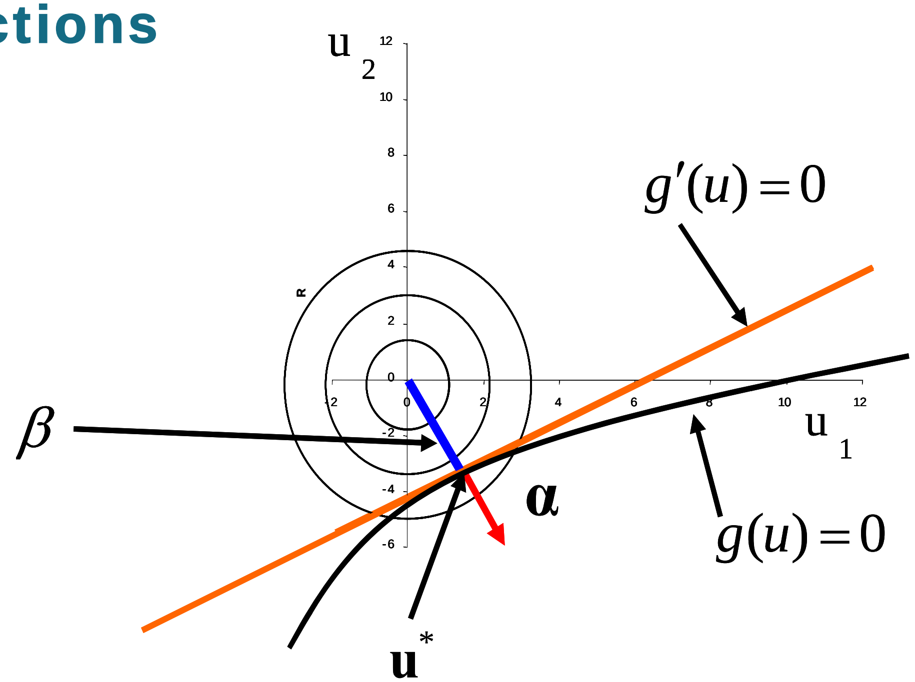
Hasofer and Lind suggested to linearize in the point where the limit state function is zero and closest to the origin in normal distributed space

Basics of Reliability Analysis

- **Non-linear limit state functions**

  The identification of the reliability index may be performed by solving an optimization problem

$$\beta = \min_{\mathbf{u}\in\{g(\mathbf{u})=0\}} \sqrt{\sum_{i=1}^{n} u_i^2}$$



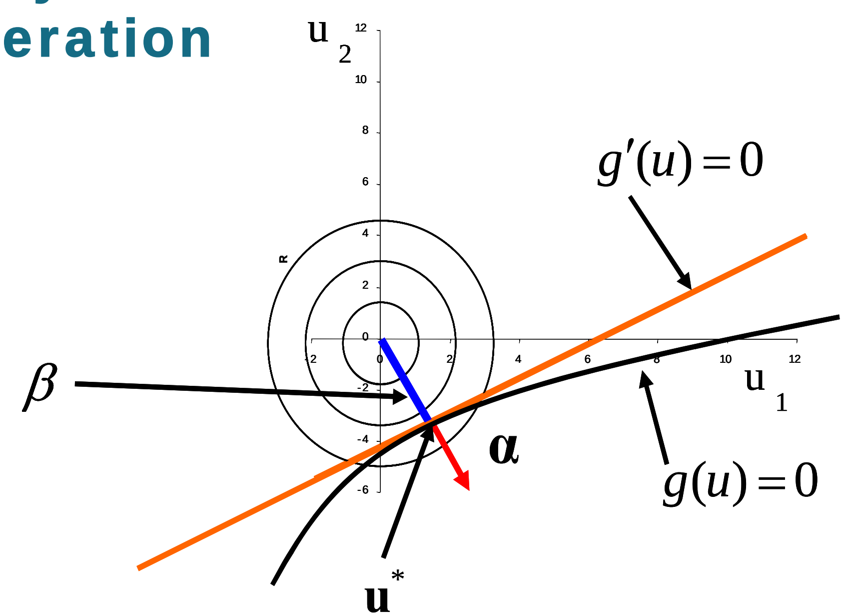
Basics of Reliability Analysis

- **Non-linear limit state functions**

  The optimization problem may be solved using the following iteration scheme

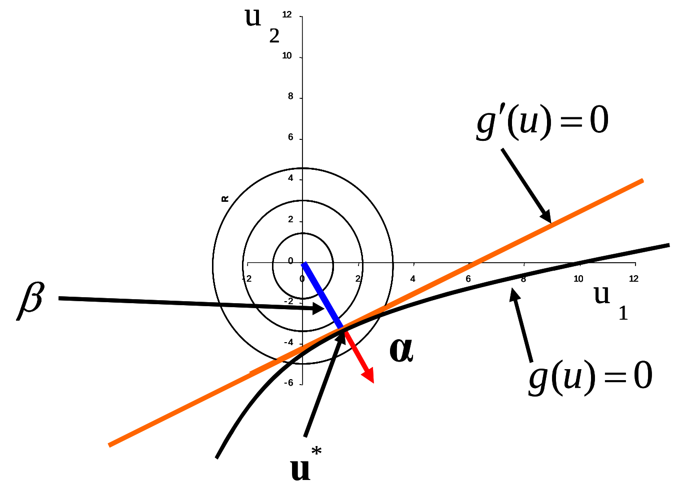
$$\alpha_i = \frac{-\dfrac{\partial g}{\partial u_i}(\beta\boldsymbol{\alpha})}{\left[\sum_{j=1}^{n}\dfrac{\partial g}{\partial u_i}(\beta\boldsymbol{\alpha})^2\right]^{1/2}}, \quad i = 1, 2, ..n$$

$$g(\beta\alpha_1, \beta\alpha_2, ...\beta\alpha_n) = 0$$



Provided that the limit state function is differentiable !

Basics of Reliability Analysis



In summary the iteration follows the following steps

1) the linearization point is chosen as $u^* = \beta\,\alpha$

2) the Normal vector to the limit state function is determined in the linearization point

$$\alpha_i = \frac{-\dfrac{\partial g}{\partial u_i}(\beta\boldsymbol{\alpha})}{\left[\sum_{j=1}^{n}\dfrac{\partial g}{\partial u_i}(\beta\boldsymbol{\alpha})^2\right]^{1/2}}, \quad i=1,2,..n$$

3) the reliability index β is calculated from

$$g(\beta\alpha_1,\beta\alpha_2,...\beta\alpha_n)=0$$

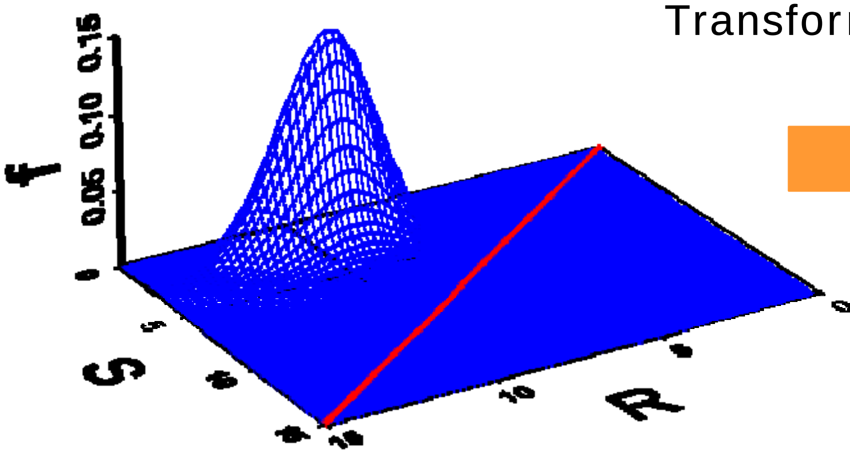
4) the new linearization point is

$$u^* = \left(\beta\alpha_1,\beta\alpha_2,...\beta\alpha_n\right)^T$$

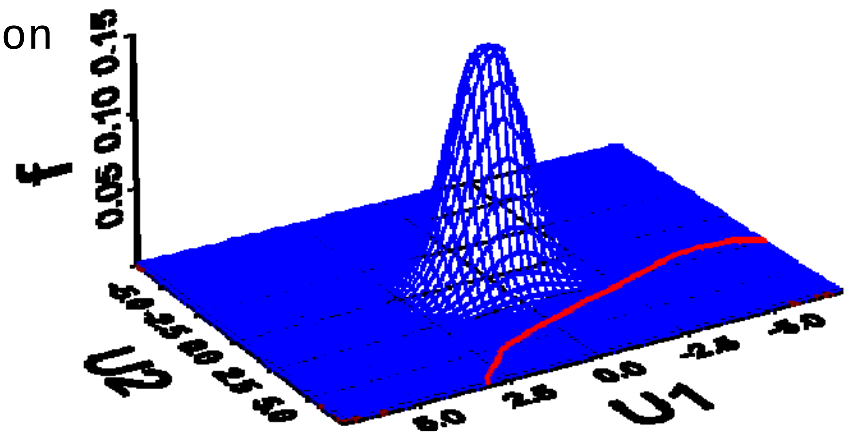
5) continue with step 2) until convergence in β

Basics of Reliability Analysis

Non-linear safety margins



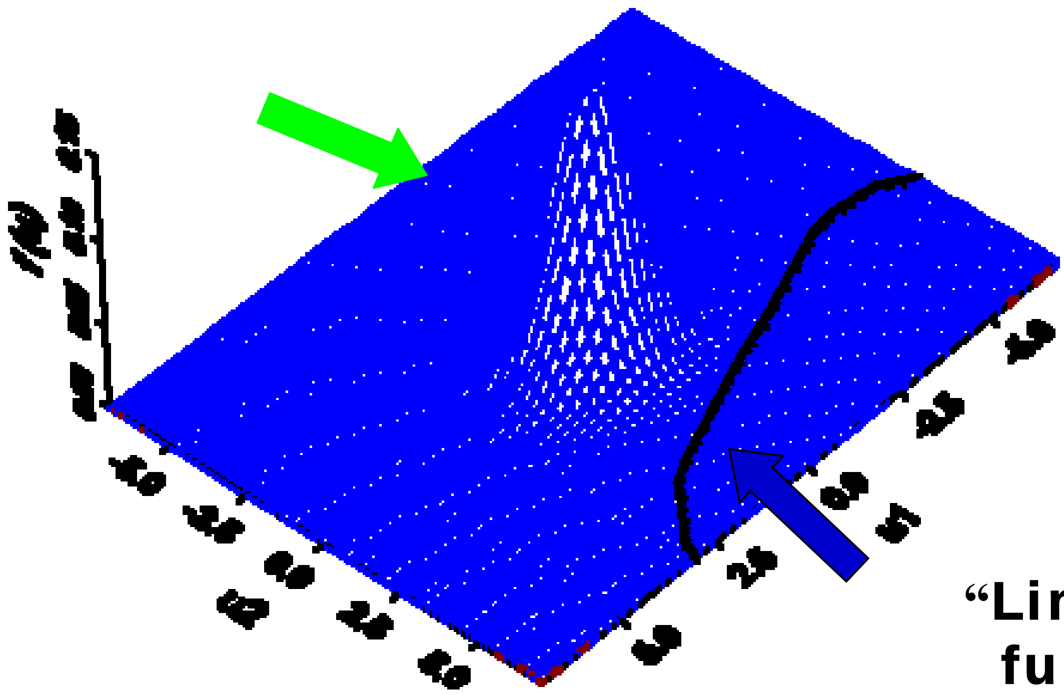
Transformation



$g(Z)$: linear	$g(U)$: non-linear
$\mu_{Z1}, \mu_{Z2} \quad \in \quad R$	$\mu_{U1} = \mu_{U2} = 0$
$\sigma_{Z1}, \sigma_{Z2} \quad \in \quad R$	$\sigma_{U1} = \sigma_{U2} = 1$

Basics of Reliability Analysis

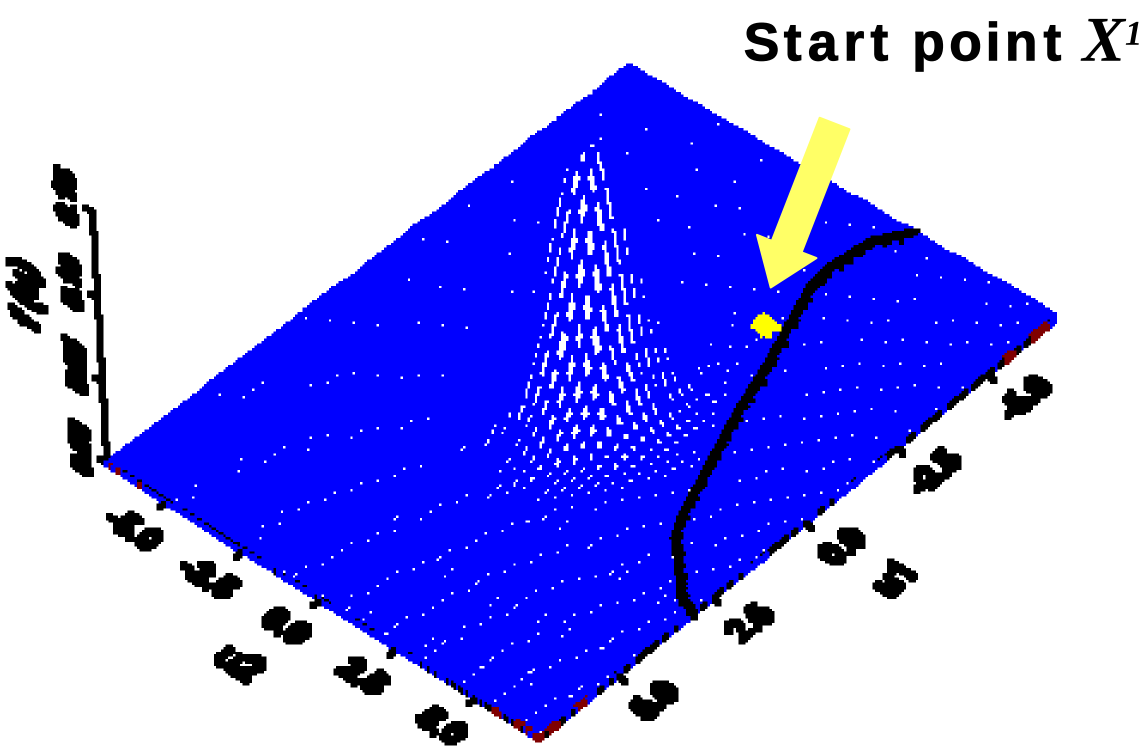
- **Non-linear safety margins**



“Limit state function”

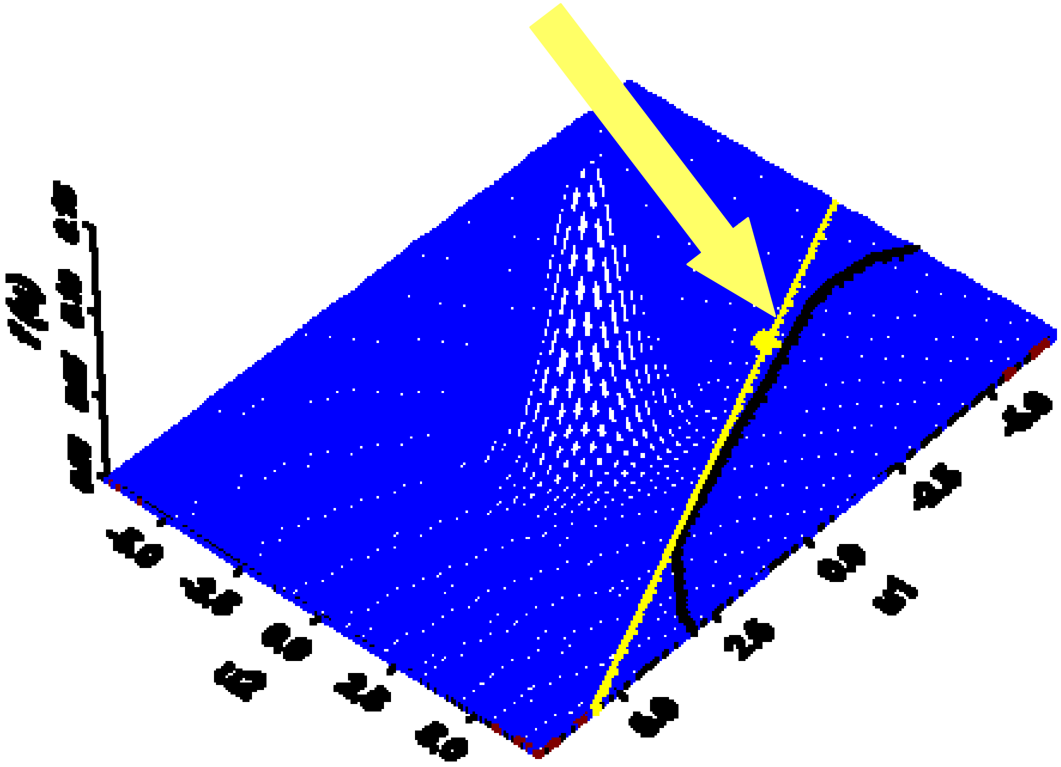
$$g(U) = R\text{-}S$$

Basics of Reliability Analysis



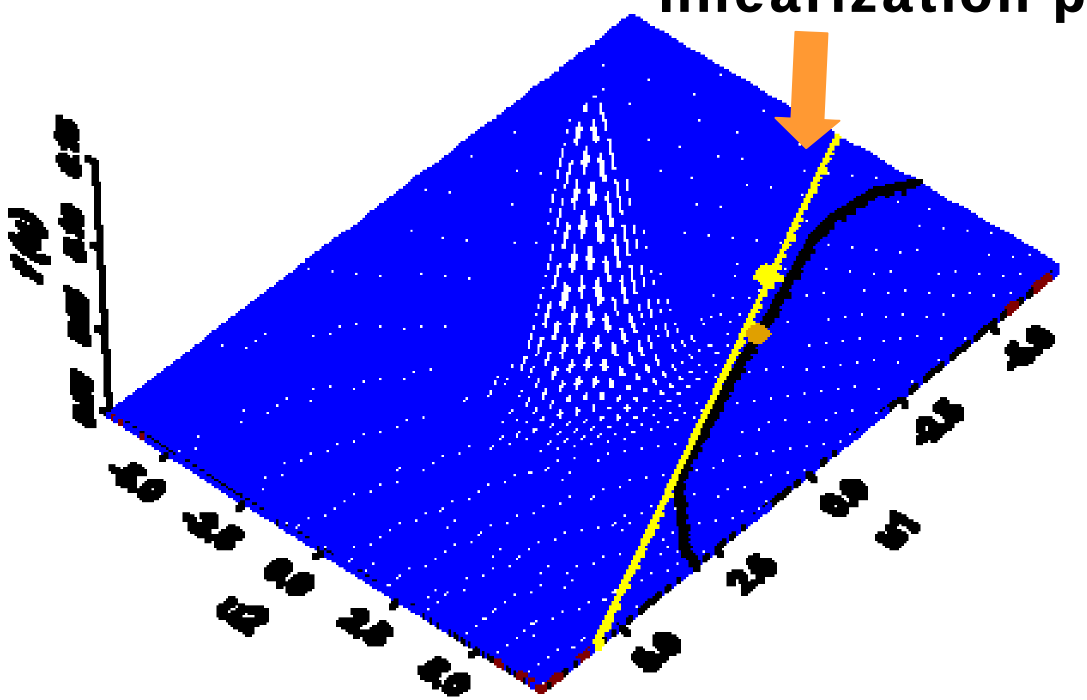
Basics of Reliability Analysis

Linearization of limit state function in X^1



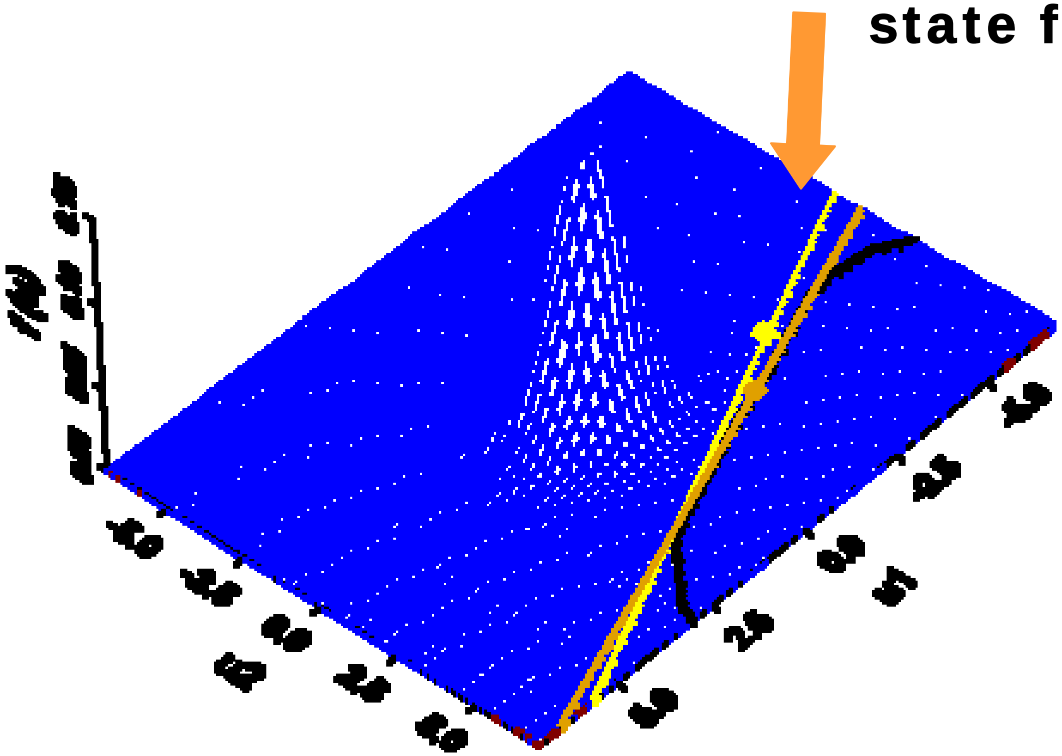
Basics of Reliability Analysis

Calculation of a new linearization point X^2



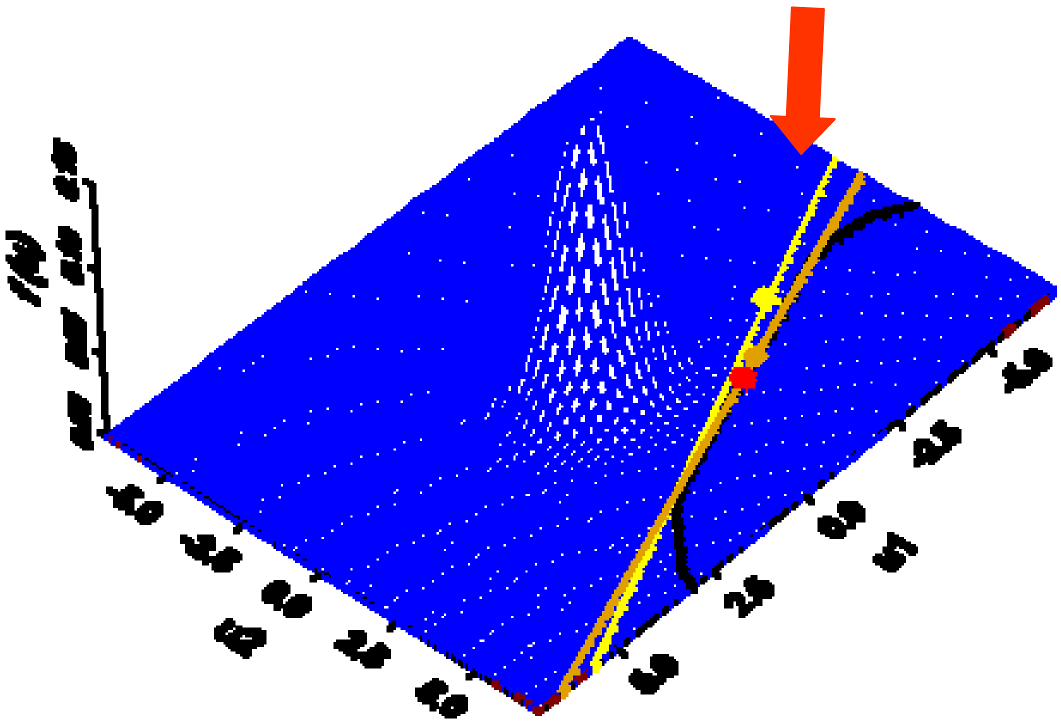
Basics of Reliability Analysis

Linearization of limit state function in X^2



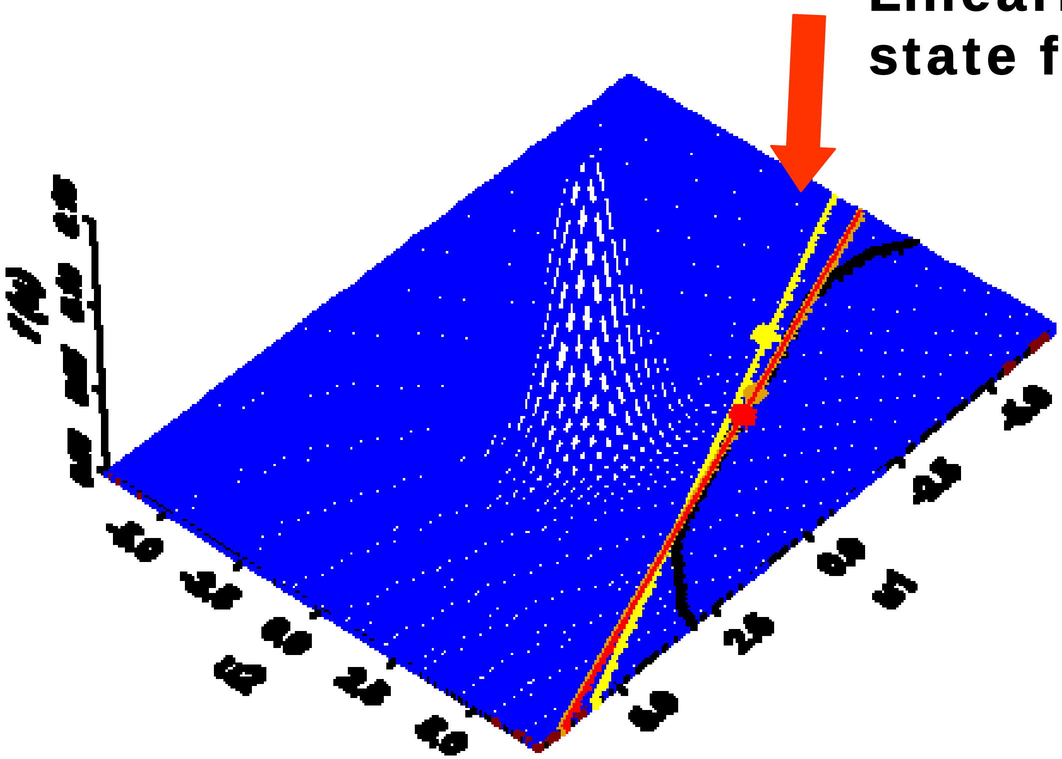
Basics of Reliability Analysis

Calculation of new linearization point X^3

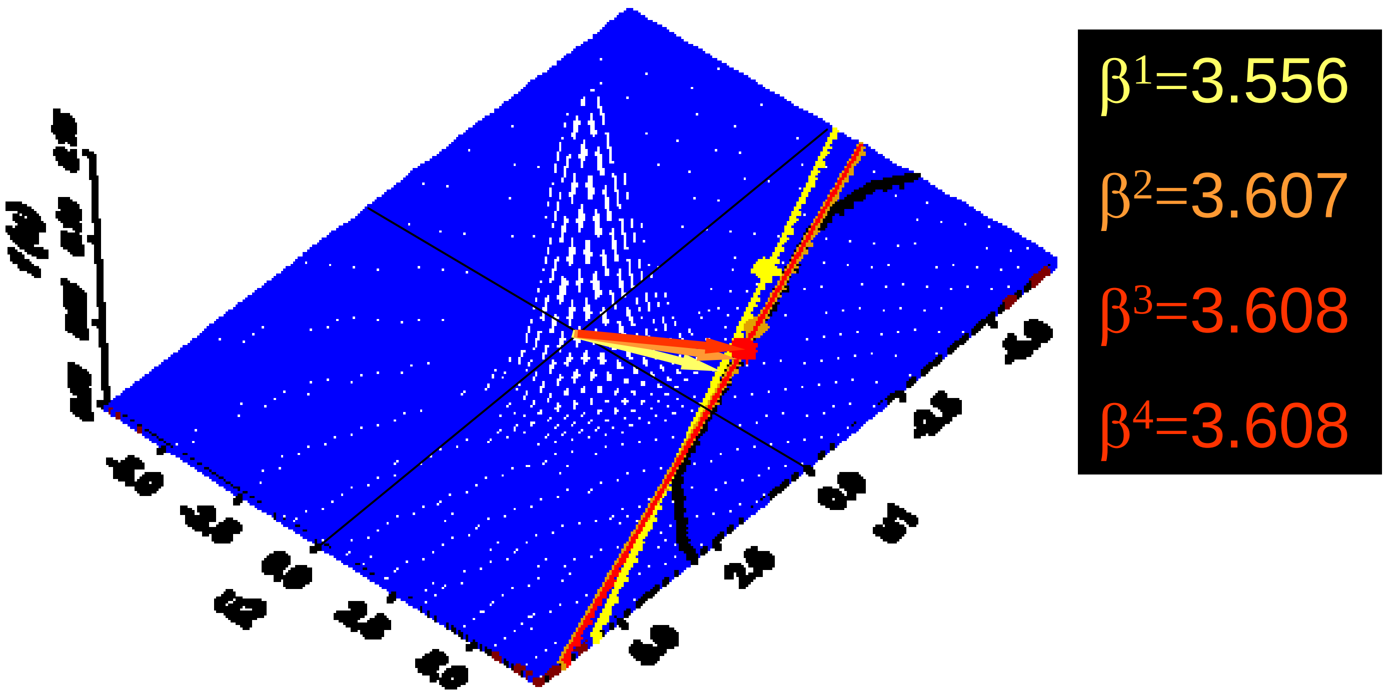


Basics of Reliability Analysis

Linearization of limit state function in X^3

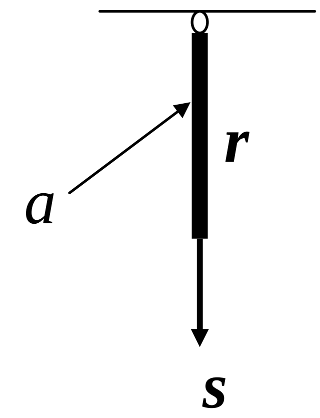


Basics of Reliability Analysis



Convergence criteria : $\Delta\beta = \left|\beta^{n+1} - \beta^{n}\right| \leq \varepsilon$

Basics of Reliability Analysis



- **Example : Reliability of steel rod**

Limit state function

Yield stress

$$g(\mathbf{x}) = r \cdot a - s$$

Load

Cross sectional area

it is assumed that R, S and A are normal distributed random variables

$$U_R = \frac{R - \mu_R}{\sigma_R} \qquad U_S = \frac{S - \mu_S}{\sigma_S} \qquad U_A = \frac{A - \mu_A}{\sigma_A}$$

$$\mu_R = 350, \sigma_R = 35$$

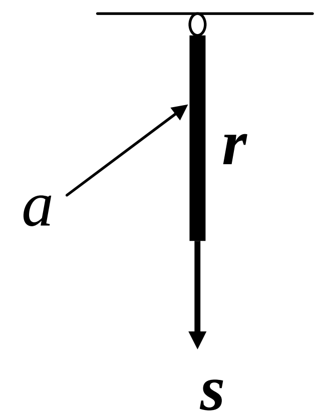
$$\mu_S = 1500, \sigma_S = 300$$

$$\mu_A = 10, \sigma_A = 2$$

Basics of Reliability Analysis

- **Example : Reliability of steel rod**

  We can now write the limit state function in terms of *u*-variables



$$\overset{\boldsymbol{r}}{}\qquad\overset{\boldsymbol{a}}{}\qquad\overset{\boldsymbol{s}}{}$$

$$g(u) = \underbrace{(u_R\sigma_R + \mu_R)}_{r}\underbrace{(u_A\sigma_A + \mu_A)}_{a} - \underbrace{(u_S\sigma_S + \mu_S)}_{s}$$

$$= (35u_R + 350)(u_A + 10) - (300u_S + 1500)$$

$$= 350\mathrm{u}_\mathrm{R} + 350u_A - 300u_S + 35u_Ru_A + 2000$$

Basics of Reliability Analysis

- **Example : Reliability of steel rod**

  The reliability index β may be found by iteration

$$\alpha_R = -\frac{1}{k}(350+35\beta\alpha_A)$$

$$\alpha_A = -\frac{1}{k}(350+35\beta\alpha_R) \qquad k=\sqrt{\alpha_R^2+\alpha_A^2+\alpha_S^2}$$

$$\alpha_S = \frac{300}{k}$$

$$\beta = \frac{-2000}{350\alpha_R+350\alpha_A-300\alpha_S+35\beta\alpha_R\alpha_A}$$

Iteration	Start	1	2	3	4	5
β	3.0000	3.6719	3.7399	3.7444	3.7448	3.7448
α_R	-0.5800	-0.5701	-0.5612	-0.5611	-0.5610	-0.5610
α_A	-0.5800	-0.5701	-0.5612	-0.5611	-0.5610	-0.5610
α_S	0.5800	0.5916	0.6084	0.6086	0.6087	0.6087

$$\alpha_i = \frac{-\frac{\partial g}{\partial u_i}(\beta\boldsymbol{\alpha})}{\left[\sum_{j=1}^{n}\frac{\partial g}{\partial u_i}(\beta\boldsymbol{\alpha})^2\right]^{1/2}}, \quad i=1,2,..n$$

$$g(\beta\alpha_1,\beta\alpha_2,...\beta\alpha_n)=0$$

$$g(u)=(u_R\sigma_R+\mu_R)(u_A\sigma_A+\mu_A)-(u_S\sigma_S+\mu_S)$$

$$=(35u_R+350)(u_A+10)-(300u_S+1500)$$
$$=350\mathrm{u}_\mathrm{R}+350u_A-300u_S+35u_Ru_A+2000$$

Basics of Reliability Analysis

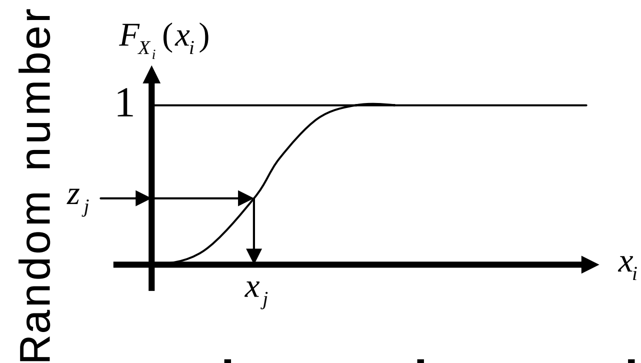
- **Monte Carlo Simulation**

  The probability integration problem may be solved by Monte Carlo simulation

$$P_f = \int_{\Omega_f = \{g(\mathbf{x}) \leq 0\}} f_{\mathbf{X}}(\mathbf{x})\, d\mathbf{x}$$

1) m **realizations of the vector** $\mathbf{X}$ **are produced**

2) for every realization the limit state function is calculated



3) the realizations for which the limit state function is equal to or less than zero are counted

$$n_f$$

Z **is a random number uniformly distributed between 0 and 1**

4) The probability of failure is estimated as

$$p_f = \frac{n_f}{m}$$

Basics of Reliability Analysis

- **Monte Carlo Simulation**

m random realizations of R and S are generated and the number of realizations n_f occuring in the failure space are counted n_f

The probability of failure p_f is then

$$p_f = \frac{n_f}{m}$$