FOSM, FORM, SORM

The structural reliability is the complement of the failure probability:

$$\text{Reliability} \equiv 1 - p_f \tag{1}$$

The reliability problem, i.e., the problem of computing p_f, has two ingredients: random variables that describe the uncertainty and one or more limit-state functions that define failure. If the problem has only one limit-state function then the problem is referred to as a component reliability problem. Otherwise, it is a system reliability problem. It is important to note that in modern reliability analysis the failure event is not necessarily structural collapse or some other event that is easily associated with the word failure. Rather, limit-state functions are defined to compute the probability of a range of events, such as the monetary loss exceeding some selected threshold.

Limit-state Functions

Since the early 80's, when "allowable stress design" was replaced by "limit-state design," structural engineers have become familiar with many "limit-states." Limit-states related to strength are often called ultimate limit-states (ULS), while those related to deflections and vibrations are called serviceability limit-states (SLS). More generally, a limit-state function defines what the engineer considers to be failure. In other words, in reliability analysis the limit-state function defines the event for which the probability is sought. Denoting the limit-state function by $g(\mathbf{x})$, where $\mathbf{x}$ is the vector of realizations of the random variables, the syntax is:

$$\begin{aligned} &g(\mathbf{x}) < 0\text{: Failure} \\ &g(\mathbf{x}) > 0\text{: Safe} \\ &g(\mathbf{x}) = 0\text{: The limit-state surface} \end{aligned} \tag{2}$$

One example of a limit-state function is

$$g = u_0 - u(\mathbf{x}) \tag{3}$$

where u_0 is a threshold value and u is a response value that depends on the outcome of the random variables. A reliability analysis with this limit-state function yields and estimate of the probability that u exceeds the threshold u_0.

The same reliability problem can be posed by different but equivalent limit-state functions, which all satisfy Eq. (2). One example is the basic reliability problem; equivalent limit-state functions for this problem include

$$g = R - S \quad \Leftrightarrow \quad g = 1 - \frac{S}{R} \quad \Leftrightarrow \quad g = \ln\left(\frac{R}{S}\right) \tag{4}$$

All these limit-state functions are negative when S exceeds R, otherwise they are positive. Therefore, because they all correctly identify the failure event, they are called "equivalent limit-state functions."

The Basic Reliability Problem

A simple but fundamental problem in reliability analysis is the "basic reliability problem," which is defined by the limit-state function

$$g = R - S \tag{5}$$

where R, from the French word *résistance*, is a random variable that represents resistance, i.e., capacity, and S, from the French word *sollicitation*, is a random variable that represents load, i.e., demand. It is observed that g is negative when S exceeds R, hence this limit-state function appropriately defines the failure event. Under the assumption that the resistance and load are statistically independent, in which case the joint PDF is $f(r,s)=f(r)f(s)$, the problem specializes to

$$p_f = \int_{R<S} \int f_R(r) f_S(s) dr\, ds = \int_0^\infty \int_0^s f_R(r) f_S(s) dr\, ds = \int_0^\infty F_R(s) f_S(s) ds \tag{6}$$

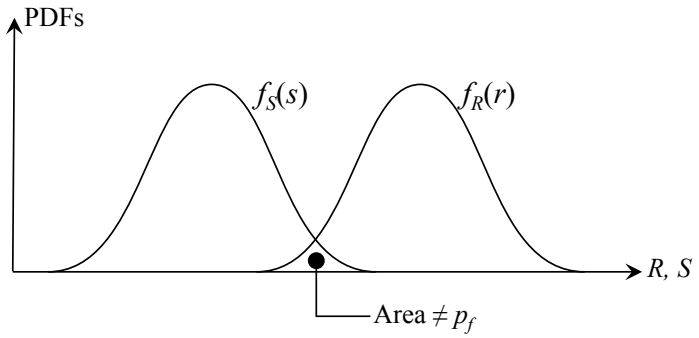

Figure 1: Visualization of the basic reliability problem by individual PDFs.

Eq. (6) shows that the failure probability for the basic reliability problem is not the overlapping area between the two marginal PDFs shown in Figure 1. Nor is it the overlapping area between the resistance-PDF and the load-CDF. Rather, the failure probability is the solution to an integral, i.e., the one in Eq. (6). To visualize this integral, it is more instructive to view the plane stretched by the two random variables, as shown in Figure 2.

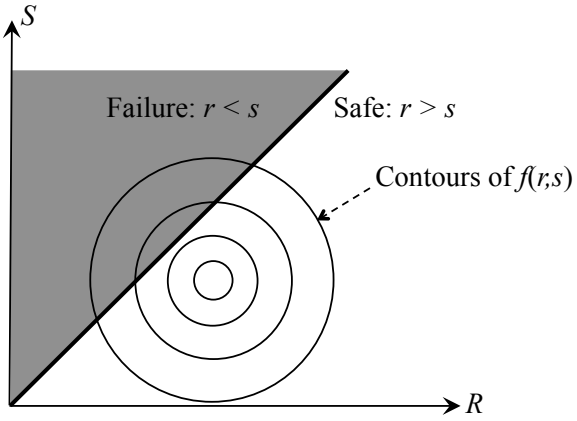

Figure 2: Visualization of the basic reliability problem in the random variable space.

Now consider the situation where R and S have the Normal probability distribution. In this case, g is also Normal because it is a linear function of R and S. In fact, the mean of g is:

$$\mu_g = \mu_R - \mu_S \tag{7}$$

and the variance of g is:

$$\sigma_g^2 = \nabla g^T \mathbf{\Sigma}_{RS} \nabla g = \sigma_R^2 + \sigma_S^2 \tag{8}$$

This implies that the failure probability is computed from the standard normal CDF, Φ:

$$p_f = \mathrm{P}(g \le 0) = \Phi\left(\frac{g - \mu_g}{\sigma_g} \right) = \Phi\left(-\frac{\mu_g}{\sigma_g} \right) = \Phi(-\beta) \tag{9}$$

where the reliability index, β, is defined as

$$\beta \equiv \frac{\mu_g}{\sigma_g} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \tag{10}$$

This reliability index, in the context of the basic reliability problem, is interpreted in Figure 3. It is observed that β is the number of standard deviations from the failure domain (grey shaded) to the mean. Clearly, the greater this distance is the greater the safety is. In other words, a high reliability index implies a low failure probability. This relationship is visualized in Figure 4, where β is plotted against p_f in the range of typical β-values. The exact relationship in Eq. (9) is shown by a solid line, and the gross approximation $p_f = 10^{-\beta}$ is shown as a dashed line. The approximation is not used in practice because it is observed in Figure 4 that it is only a valid estimate around $\beta = 3$.

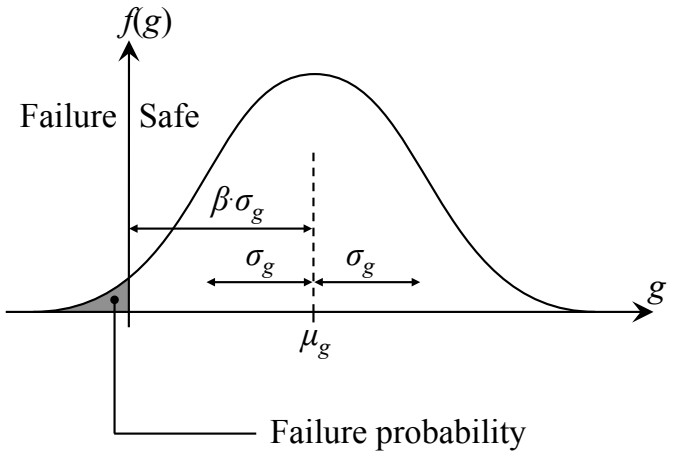

Figure 3: Normal probability distribution for the basic limit-state function.

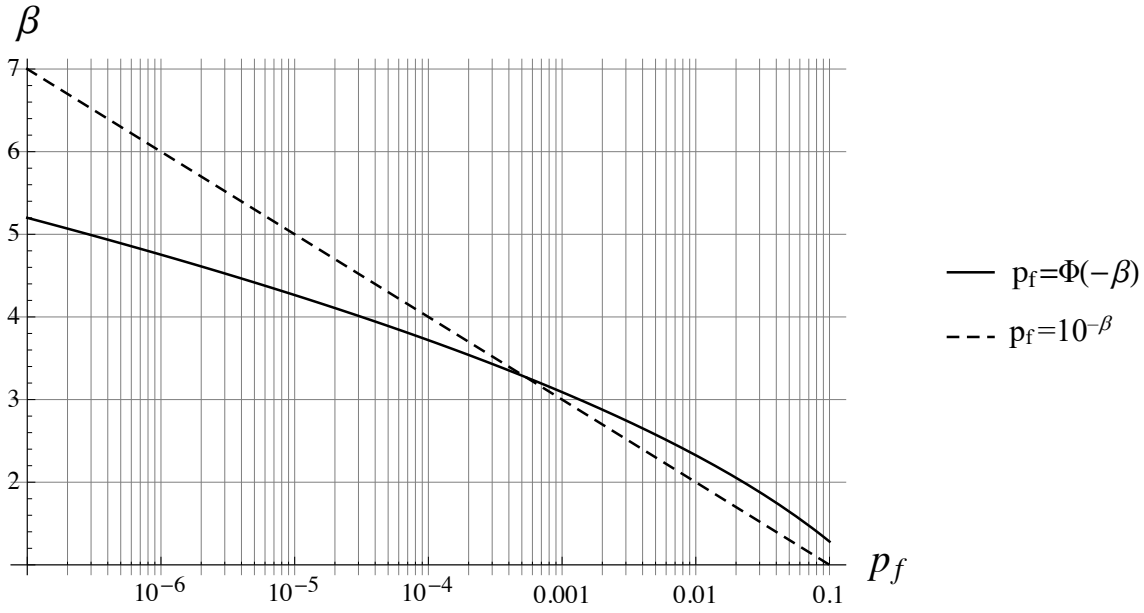

Figure 4: Relationship between β and p_f.

The General Reliability Problem

A generic reliability problem with only one limit-state function is referred to as the "component reliability problem," which reads:

$$p_f = \mathrm{P}(g \le 0) = \int_{g \le 0} \cdots \int f(\mathbf{x}) d\mathbf{x} \tag{11}$$

where $f(\mathbf{x})$ is the joint PDF for the random variables. Problems in which the joint state of more than one limit-state function defines failure are called "system reliability problems," and they are addressed in another document. Eq. (11) shows that the component reliability problem amounts to integrating the probability density in the region of the **x**-space where the limit-state function is negative. All reliability methods address this problem. Unfortunately, it is impossible to solve the multi-fold integral in Eq. (11) analytically, except for a few special cases. However, methods like FORM, SORM, and sampling provides approximate solutions. Because it is possible to solve Eq. (11) analytically for only a few special cases a host of reliability methods have been developed to solve the component reliability problem in an approximate manner. Some of these are listed below and described in other documents.

- The mean-value, first-order, second-moment (MVFOSM) method employs only second-moment information about the random variables together with a first-order approximation of the limit-state function about the mean of the random variables. This method is highly efficient, but it is associated with several disadvantages: 1) It may lead to different results for equivalent limit-state functions, referred to as the invariance problem; 2) The linearization of the limit-state function may lead to inaccuracies when

the limit-state function is nonlinear; 3) Information about the distribution type of the random variables is ignored.

- The first-order reliability method (FORM) remedies the invariance problem of MVFOSM and incorporates information about the distribution type of the random variables. These advantages come at a higher computational cost, accompanied by the possibility of non-convergence. FORM maintains the linearization of the limit-state function and may thus lead to inaccuracies when the limit-state function is nonlinear.
- The second-order reliability method (SORM) is equivalent to FORM except that the limit-state function is approximated by a second-order function. To some extent this remedies possible inaccuracies when the limit-state function is nonlinear, but more computations are necessary compared with FORM.
- Sampling methods evaluate the limit-state function at many realizations of the random variables and yield an approximate value of the failure probability. Several sampling methods are available, ranging from mean-centered Monte Carlo sampling to adaptive sampling schemes. A particularly appealing sampling method when FORM as been carried out is importance sampling. Based on information from the FORM analysis this approach provides an accurate and robust estimate of the failure probability with far fewer samples than, say, Monte Carlo sampling.
- Response surface methods, including the neural network approach, are three-step approaches. First, the limit-state function is evaluated at many pre-selected realizations of the random variables. Next, some interpolation function, typically a second-order function or a neural network, is utilized to approximate the limit-state function. Finally, a reliability method, like FORM or sampling, is employed in conjunction with the approximate function to estimate the failure probability.

Although it is somewhat artificial in practical use, several textbooks distinguish between Level I, II, III, and IV reliability methods:

- Level I methods: In these methods only one value associated with each uncertain parameter enters the analysis. For example, design code equations that utilize one characteristic value for each parameter represent Level I methods. Level I methods do not yield a failure probability.
- Level II methods: Two values for each uncertain parameter enter. Typically, the mean and standard deviation for each parameter, in addition to possible correlation between the parameters. According to this definition, second-moment methods are Level II methods. Unless the random variables are normally distributed, Level II methods do not yield a failure probability.
- Level III methods: At this level the joint probability distribution for the parameters are available and a failure probability is produced.
- Level IV methods: These go beyond Level III methods by including cost-benefit considerations that yield the target reliability. This approach is suitable for projects with limited precedence or dramatic failure consequences, such as offshore oil and gas installations and nuclear power plants.

MVFOSM

Consider a situation in which the only information we have about the random variables is their second-moments, i.e., means, standard deviations, and correlation coefficients. Under these circumstances, analysis of functions provides the mean and standard deviation of the limit-state function, i.e., μ_g and σ_g. When those values are substituted into Eq. (10) we have the mean-value first-order second-moment reliability method, MVFOSM:

$$\beta = \frac{\mu_g}{\sigma_g} \tag{12}$$

The validity of β as a proxy for failure probability is understood from Eq. (9). The quantity β can also be interpreted by first imagining the PDF of the limit-state function. Unless g is a linear function of normal random variables, or some other simple case, this PDF cannot be established analytically. Nevertheless, this is a pedagogically useful thought-construct. Obviously, the PDF of g has mean μ_g and standard deviation σ_g. Furthermore, g=0 separates the failure outcomes from the safe outcomes. By multiplying Eq. (12) by σ_g it becomes clear that β is the number of standard deviations from the mean to the failure region. The more standard deviations the failure domain is away from the mean the safer. In other words, the higher reliability index the smaller is the failure probability. Conversely, small values of the reliability index indicate that the failure domain is closer to the mean, which implies a higher failure probability. In addition to this explanation, a geometric interpretation of β in the space of random variables, which will be useful in the derivation of other reliability methods, is presented later in this document.

Invariance Problem

Formulas for the computation of μ_g and σ_g are known from the analysis of functions. If the limit-station function is linear then μ_g and σ_g are exact. If it is nonlinear then first-order approximations are available:

$$\mu_g \approx g(\mathbf{M}_X) \tag{13}$$

$$\sigma_g \approx \sqrt{\nabla g(\mathbf{M}_X)^T \boldsymbol{\Sigma}_{XX} \nabla g(\mathbf{M}_X)} \tag{14}$$

It is the potential inaccuracy in these two equations that gives rise to the "invariance problem" of the MVFOSM method. It is called the invariance problem because the MVFOSM may provide different results for equivalent limit-state functions. Two limit-state functions are equivalent if they share the surface g=0 in the space of random variables. Consider the basic reliability problem as an example. For the limit-state function g=R-S the MVFOSM reliability index is

$$\beta = \frac{\mu_g}{\sigma_g} = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 - 2\rho_{RS}\sigma_R\sigma_S + \sigma_S^2}} \tag{15}$$

For the equivalent limit-state function g=ln(R/S) the solution is

$$\beta = \frac{\mu_g}{\sigma_g} = \frac{\ln\left(\mu_R/\mu_S\right)}{\sqrt{\sigma_R^2/\mu_R^2 - (2\rho_{RS}\sigma_R\sigma_S)/(\mu_R\mu_S) + \sigma_S^2/\mu_S^2}} \tag{16}$$

In general, Eqs. (15) and (16) yield different results. For example, for μ_R=30, μ_S=20, σ_R=5, σ_S=10, and ρ_{RS}=0.5 Eq. (15) yields β=1.15 while Eq. (16) yields β=0.92. This exemplifies the invariance problem. The root of the problem is that the linearization of the limit-state function is made at the mean. For a nonlinear limit-state functions this implies that the limit-state surface, i.e., g=0, is different for the linearized function and the actual function. It would be better to carry out the linearization somewhere at the limit-state surface g=0, which is shared by all equivalent limit-state functions. This solution gives rise to FORM and SORM, which are described in other documents.

Geometric Interpretation of β

To better understand the reliability index and to assist the development of FORM, a geometric interpretation of the second-moment reliability index is made. First consider a linear limit-state function of the form

$$g(\mathbf{X}) = a + \mathbf{b}^T\mathbf{X} \tag{17}$$

Furthermore, transform the function into the space of standard variables, i.e., transform $\mathbf{X}$ into a vector of variables with zero means and unit covariance matrix:

$$\mathbf{X} = \mathbf{M}_X + \mathbf{D}_X\mathbf{L}\mathbf{Y} \tag{18}$$

Substitution of Eq. (18) into Eq. (17) and naming the limit-state function in the standard space G yields

$$\begin{aligned} G(\mathbf{Y}) &= a + \mathbf{b}^T\mathbf{M}_X + \mathbf{b}^T\mathbf{D}_X\mathbf{L}\mathbf{Y} \\ &= c + \mathbf{d}^T\mathbf{Y} \end{aligned} \tag{19}$$

where c and $\mathbf{d}$ have been defined. Now carry out MVFOSM with this limit-state function:

$$\beta = \frac{\mu_G}{\sigma_G} = \frac{c}{\sqrt{\mathbf{d}^T\mathbf{d}}} = \frac{c}{\|\mathbf{d}\|} \tag{20}$$

Compare this result with the geometry formula for the distance from a point to a plane. Indeed, the distance from $\mathbf{Y}=\mathbf{0}$ (the origin in the $\mathbf{Y}$-space) to the limit-state plane $G(\mathbf{Y})=0$ is:

$$\Delta = \left|\frac{G(\mathbf{0})}{\|\nabla G\|}\right| = \left|\frac{c}{\|\mathbf{d}\|}\right| \tag{21}$$

We conclude that the reliability index β is the distance from the origin to the limit-state surface in the space of standardized random variables. This shows an important appeal of the standard space: distances can be measured. Conversely, the original random variables usually have a variety of units and a distance in that space is not a meaningful concept.

Normal vs. Lognormal

The MVFOSM method yields a reliability index, but it does not provide a failure probability. This changes in the special case where the limit-state function is linear and the random variables are normally distributed. In this case, the probability distribution of the limit-state function is also normal. Furthermore, the second-moment analysis yields the

precise mean, μ_g, and standard deviation, σ_g. Consequently, the standard normal CDF is employed to obtain the failure probability:

$$p_f = P(g \le 0) = \Phi\left(\frac{0-\mu_g}{\sigma_g}\right) = \Phi(-\beta) \tag{22}$$

A note is here added in regards to lognormal random variables. In another document it is described that if Y is a normal random variable and

$$Y = \ln(X) \tag{23}$$

then X is said to have the lognormal distribution. Now consider a linear limit-state function with lognormal random variables. Taking the left-most limit-state function in Eq. (4) as an example, and taking the natural logarithm of it, we essentially face the right-most limit-state function in Eq. (4):

$$g = \ln(R) - \ln(S) \equiv Y_R - Y_S \tag{24}$$

where the normal random variables Y_R and Y_S are defined. This means that the solution in Eq. (15) applies, with means and standard deviations replaced by those of the underlying normal random variables. Expressions for those are given in the document on random variables.

MVFOSM Importance Vector

Ranking of the random variables according to relative importance often provides valuable insight. The most important random variables should be subjected to particular scrutiny, while unimportant random variables may be omitted from the analysis. An importance vector from MVFOSM, i.e., a vector with components that measure the relative importance the corresponding random variable is derived by considering the linearized limit-state function at the mean point:

$$g(\mathbf{x}) \approx g(\mathbf{M}_\mathbf{x}) + \nabla g(\mathbf{M}_\mathbf{x})^T \cdot \left(\mathbf{x} - \mathbf{M}_\mathbf{x}\right) \tag{25}$$

The variance of the linearized limit-state function is

$$\begin{aligned} \mathrm{Var}\left[g(\mathbf{x})\right] &\approx \nabla g^T \cdot \mathbf{\Sigma}_{\mathbf{xx}} \cdot \nabla g \\ &= \left(\nabla g_1 \sigma_1\right)^2 + \left(\nabla g_2 \sigma_2\right)^2 + \cdots + \left(\nabla g_n \sigma_n\right)^2 + \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \nabla g_i \nabla g_j \sigma_i \sigma_j \rho_{ij} \end{aligned} \tag{26}$$

where n is the number of random variables. It is observed that the direct contribution of random variable x_i to the total variance is $\left(\nabla g_i \cdot \sigma_i\right)^2$. For this reason, the following vector is considered an importance vector in MVFOSM:

$$\boldsymbol{\omega} = -\nabla g^T \mathbf{D} \tag{27}$$

The greater absolute value of $\omega_i = \nabla g_i \cdot \sigma_i$ the greater importance of the corresponding random variable x_i. The sign of ω_i also matters:

- *Positive* ω_i means that x_i acts like a *load* variable
- *Negative* ω_i means that x_i acts like a *resistance* variable

FORM

This material was first described to me in a course taught by my PhD thesis supervisor Professor Armen Der Kiureghian at the University of California at Berkeley. In 2005 he made an excellent description available in Chapter 14 of the CRC Engineering Design Reliability Handbook edited by Nikolaidis, Ghiocel and Singhal, published by the CRC Press in Boca Raton, Florida. FORM is an abbreviation of the first-order reliability method, which remedies the invariance problem of MVFOSM. The remedy is to approximate the limit-state function somewhere on the limit-state surface, i.e., at a point where g=0 instead of at the mean. The limit-state surface, which separates the failure domain from the safe domain, is shared by all equivalent limit-state functions.

Properties of the Standard Normal Space

Given the prologue, two questions appear: Which point on the limit-state surface to select, and thereafter: How to obtain the failure probability. The answer to both questions is found in the "standard normal space." This is a space of uncorrelated standard normal random variables, in which realizations are denoted $\mathbf{y}$ and the joint PDF is

$$\varphi(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^n}} \cdot \exp\left(-\frac{1}{2}\mathbf{y}^T\mathbf{y}\right) \tag{28}$$

The transformation from the original $\mathbf{x}$-space to the standard normal $\mathbf{y}$-space is adopted for two reasons:

1. The joint PDF in the standard normal space is rotationally symmetric and it decays in the radial and tangential directions. Consequently, the point on the limit-state surface that is closest to the origin is the point in the failure domain with highest probability density. As a result, the point closest to the origin is an appealing point for approximating the limit-state function, because that is where a significant portion of the failure probability density is located.
2. In the standard normal space it is possible to develop a formula for the probability content outside a hyper-plane, which is used in FORM, and outside a hyper-parabolid, which is used in SORM. The probability content outside a hyper-plane is

$$p_f = \Phi(-\beta) \tag{29}$$

where β is the distance from the origin to the closest point on the hyper-plane.

The limit-state function is denoted $g(\mathbf{x})$ in the original space, and it is denoted by capital letter $G(\mathbf{y})$ in the standard normal space.

The FORM Linearization

The essence of FORM is to linearize the limit-state function $G(\mathbf{y})$ at the design point, denoted $\mathbf{y}^*$, which is the point on the limit-state surface $G(\mathbf{y})$=0 that is closest to the origin. To establish a few key concepts, consider the linear Taylor series expansion around the design point:

$$G(\mathbf{y}) \approx G(\mathbf{y}^*) + \nabla G(\mathbf{y}^*)^T \cdot (\mathbf{y} - \mathbf{y}^*) \tag{30}$$

where the first term on the right-hand side is zero because $\mathbf{y}^*$ is on the limit-state surface. In FORM analysis it is common to replace the gradient vector by its negative and normalized version, called the alpha-vector:

$$\boldsymbol{\alpha} = -\frac{\nabla G(\mathbf{y})}{\|\nabla G(\mathbf{y})\|} \tag{31}$$

In these notes all vectors are column vectors, and combination of Eq. (31) and Eq. (30) yields:

$$G \approx -\left\|\nabla G(\mathbf{y}^*)\right\| \cdot \boldsymbol{\alpha}^T (\mathbf{y} - \mathbf{y}^*) = \left\|\nabla G(\mathbf{y}^*)\right\| \cdot (\boldsymbol{\alpha}^T \mathbf{y}^* - \boldsymbol{\alpha}^T \mathbf{y}) \tag{32}$$

$\boldsymbol{\alpha}$ is a unit vector, thus the dot product between the parallel vectors $\boldsymbol{\alpha}$ and $\mathbf{y}^*$ is the length of $\mathbf{y}^*$. This is the distance from the origin to the design point, which is defined as the reliability index, β. Substitution into Eq. (32) yields

$$G \approx \left\|\nabla G(\mathbf{y}^*)\right\| (\beta - \boldsymbol{\alpha}^T \mathbf{y}) \tag{33}$$

which expresses the linearized limit-state function in FORM.

Finding the Design Point

From the previous it is understood that FORM consists of a transformation from the original space of random variables $\mathbf{x}$ to the standard normal space of random variables $\mathbf{y}$, where the limit-state function is linearized at the point closest to the origin. This point is called the design point, or the most probable failure point. From above it is understood that it is the solution to the following optimization problem:

$$\mathbf{y}^* = \arg\min \left\{ \|\mathbf{y}\| \;\middle|\; G(\mathbf{y}) = 0 \right\} \tag{34}$$

where $\mathbf{y}^*$ is the design point coordinates and $G(\mathbf{y})$ is the limit-state function in the standard normal space. The value of $G(\mathbf{y})$ is determined by transforming realizations of $\mathbf{y}$ into realizations of $\mathbf{x}$ and evaluating the original limit-state function $g(\mathbf{x})$. In the context of optimization analysis it is common to write the problem in Eq. (34) in terms of the dot product of $\mathbf{y}$ with itself, multiplied by a convenience factor ½:

$$\mathbf{y}^* = \arg\min \left\{ \frac{1}{2}\|\mathbf{y}\|^2 \;\middle|\; G(\mathbf{y}) = 0 \right\} \tag{35}$$

Eqs. (34) and (35) are a nonlinear equality-constrained optimization problem. The inequality-constrained version

$$\mathbf{y}^* = \arg\min \left\{ \|\mathbf{y}\| \;\middle|\; G(\mathbf{y}) \le 0 \right\} \tag{36}$$

is an equivalent problem unless the limit-state function is negative at the origin in the standard normal space, which implies an unusually high failure probability. Upon solving the optimization problem the distance from the origin in the standard normal space to the point $\mathbf{y}^*$ is the reliability index β:

$$\beta = \left\| \mathbf{y}^* \right\| \tag{37}$$

The generic search scheme to determine the design point, i.e., solving Eqs. (34), (35), or (36), is

$$\mathbf{y}_{m+1} = \mathbf{y}_m + s_m \cdot \mathbf{d}_m \tag{38}$$

where m is the iteration counter, s_m is the step size at the m^{th} iteration, and $\mathbf{d}_m$ is the search direction at the m^{th} iteration. Typically, but not necessarily, the search starts at the origin of the standard normal space, and algorithms to determine s_m and $\mathbf{d}_m$ are described shortly.

Convergence Criteria

Like all other optimization algorithms, only local convergence can be generally proven. Two convergence criteria must be satisfied. First, the trial point must be close to the limit-state surface:

$$\left| \frac{G(\mathbf{y}_m)}{G_0} \right| \leq e_1 \tag{39}$$

where G_0 is a scaling factor that is usually selected as the start-value of the limit-state function and e_1 is a threshold that is usually selected around 10^{-3}. Second, the trial point is required to be the point on the limit-state surface that is closest to the origin. For that to be the case it must be a gradient projection point, i.e., the gradient of the limit-state function at the trial point should point at the origin. This is the case if $\mathbf{y}_m$ is parallel to the gradient vector. The length of the $\mathbf{y}$-vector in the direction of the $\boldsymbol{\alpha}$-vector defined in Eq. (31) is the dot product $\boldsymbol{\alpha}^T\mathbf{y}$. As shown in Figure 5, the difference between the vector $(\boldsymbol{\alpha}^T\mathbf{y})\boldsymbol{\alpha}$ and $\mathbf{y}$ is therefore a reasonable convergence criterion:

$$\left\| \mathbf{y}_m - \left(\boldsymbol{\alpha}_m^T \mathbf{y}_m\right)\boldsymbol{\alpha}_m \right\| \leq e_2 \tag{40}$$

where e_2 is a threshold that is usually selected around 10^{-3}.

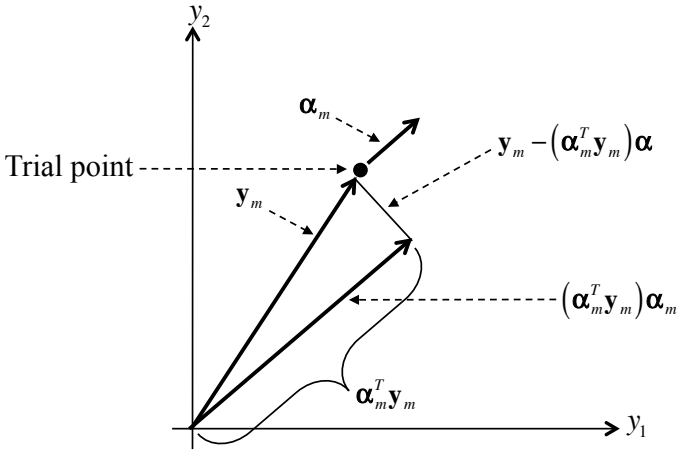

Figure 5: Second convergence criterion in FORM.

Because the deviation between the two vectors in Eq. (40) is measured as a distance instead of an angle, the convergence criterion gets stricter with increasing distance from the origin. This is a problem that is remedied either by a scaling, for example by the norm of $\mathbf{y}$:

$$\left\| \frac{\mathbf{y}_m}{\|\mathbf{y}_m\|} - \left(\boldsymbol{\alpha}_m^T \frac{\mathbf{y}_m}{\|\mathbf{y}_m\|} \right) \boldsymbol{\alpha}_m \right\| \leq e_2 \tag{41}$$

or by comparing angles instead. Fundamental vector geometry gives

$$\cos(\theta) = \frac{\boldsymbol{\alpha}^T \mathbf{y}}{\|\mathbf{y}\|} \tag{42}$$

because $\boldsymbol{\alpha}$ is a unit vector. Consequently, an angle-based second convergence criterion is

$$1 - \frac{\boldsymbol{\alpha}_m^T \mathbf{y}_m}{\|\mathbf{y}_m\|} \leq e_2 \tag{43}$$

With reference to optimization theory, this convergence criterion represents a Karush-Kuhn-Tucker optimality condition. To understand this, consider the Lagrangian function for the optimization problem in Eq. (34):

$$L(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 + \lambda \cdot G(\mathbf{y}) \tag{44}$$

where the factor ½ is insignificant except for convenience of later expressions. The first optimality condition is that $G(\mathbf{y})=0$. The second requires the gradient of the Lagrangian function to be zero:

$$\nabla L(\mathbf{y}) = \mathbf{y} + \lambda \cdot \nabla G(\mathbf{y}) = 0 \tag{45}$$

Consider the possible solution

$$\lambda = \frac{\|\mathbf{y}\|}{\|\nabla G(\mathbf{y})\|} \tag{46}$$

Substitution into Eq. (45) yields

$$\frac{\mathbf{y}}{\|\mathbf{y}\|} + \frac{\nabla G(\mathbf{y})}{\|\nabla G(\mathbf{y})\|} = 0 \tag{47}$$

Indeed, this equation is satisfied if the trial point is a gradient projection point, in which case $\mathbf{y}$ is parallel to the gradient vector and Eq. (47) is the sum of two vectors with opposite directions and unit length. It is observed that Eq. (47) is equivalent to Eq. (43) when Eq. (47) is multiplied by $\mathbf{y}^T / \|\mathbf{y}\|$ and adopting Eq. (31). Hence, Eq. (43) represents a Karush-Kuhn-Tucker optimality condition and is appropriate as a convergence criterion.

iHLRF Algorithm

This state-of-the-art algorithm for finding the design point has evolved from its conception together with FORM in the 1970s. Hasover and Lind initiated the development in 1974 and Rackwitz and Fiessler extended it in 1978. Further improvements were made in the 1990s

by Der Kiureghian and students to arrive at the algorithm's current name and form. For pedagogical purposes, start by considering the linearized limit-state surface at a point $\mathbf{y}_m$ in the standard normal space, similar to the linearization in Eq. (30):

$$G(\mathbf{y}) \approx G(\mathbf{y}_m) + \nabla G(\mathbf{y}_m)^T \cdot (\mathbf{y} - \mathbf{y}_m) \tag{48}$$

The limit-state surface, which in this case is a hyper-plane, is expressed by setting Eq. (48) equal to zero:

$$G(\mathbf{y}_m) + \nabla G(\mathbf{y}_m)^T \cdot (\mathbf{y} - \mathbf{y}_m) = 0 \tag{49}$$

According to fundamental geometry, the distance from the origin to the closest point on the limit-state surface equals the function value at the origin divided by the norm of the gradient vector:

$$\Delta = \frac{G(\mathbf{0})}{\left\|\nabla G(\mathbf{y}_m)\right\|} = \frac{G(\mathbf{y}_m) - \nabla G(\mathbf{y}_m)^T \cdot \mathbf{y}_m}{\left\|\nabla G(\mathbf{y}_m)\right\|} = \frac{G(\mathbf{y}_m)}{\left\|\nabla G(\mathbf{y}_m)\right\|} + \boldsymbol{\alpha}^T \cdot \mathbf{y}_m \tag{50}$$

If the actual limit-state function was linear then Eq. (50) would represent the sought distance, i.e., the reliability index β, and the design point coordinates would be obtained by multiplying the distance by the negative normalized gradient vector, called the alpha-vector:

$$\mathbf{y}_{m+1} = -\Delta \cdot \frac{\nabla G(\mathbf{y}_m)}{\left\|\nabla G(\mathbf{y}_m)\right\|} \equiv \Delta \cdot \boldsymbol{\alpha} \tag{51}$$

For non-linear limit state functions this is not the final solution, but Eq. (51) is nevertheless useful. It is employed as a trial point at which a new linearization is performed. In other words, Eq. (51) is a recursive algorithm. Combining Eqs. (50) and (51) yields:

$$\mathbf{y}_{m+1} = \left(\frac{G(\mathbf{y}_m)}{\left\|\nabla G(\mathbf{y}_m)\right\|} + \boldsymbol{\alpha}^T \cdot \mathbf{y}_m \right) \boldsymbol{\alpha} \tag{52}$$

Furthermore, a search algorithm of the form in Eq. (38) is obtained by extracting the search direction vector $\mathbf{d}_m$ from Eq. (52):

$$\mathbf{d}_m = \mathbf{y}_{m+1} - \mathbf{y}_m = \left(\frac{G(\mathbf{y}_m)}{\left\|\nabla G(\mathbf{y}_m)\right\|} + \boldsymbol{\alpha}^T \cdot \mathbf{y}_m \right) \boldsymbol{\alpha} - \mathbf{y}_m \tag{53}$$

The step size s_m in Eq. (38) is unity by default, which characterizes the original HLRF algorithm and gives rise to potential convergence problems. This issue is analogue to the potential convergence problems in the elementary Newton root-finding algorithm. The unit step size may be far too long under certain circumstances. Rather, in accordance with optimization theory the line search problem of selecting a step size is recognized as an optimization problem in itself. However, rather than carry out this new optimization analysis the Armijo rule is adopted:

$$s_m = b^k \tag{54}$$

where $0<b<1$ is a constant usually equal to ½ and k is an integer that is sequentially increased from zero until the step size is accepted. Effectively, for b=½, Armijo's rule cuts the step size in half until a step size is accepted. The criterion for accepting a step size is expressed in terms of a merit function. The step size must satisfy

$$m(\mathbf{y}_{m+1}) - m(\mathbf{y}_m) \leq a \cdot s_m \cdot \left(\nabla m(\mathbf{y}_m)^T \mathbf{d}_m\right) \tag{55}$$

where $a>0$ is typically set equal to ½ and the merit function reads

$$m(\mathbf{y}) = \frac{1}{2} \cdot \|\mathbf{y}\|^2 + c \cdot |G(\mathbf{y})| \tag{56}$$

with gradient

$$\nabla m(\mathbf{y}) = \mathbf{y} + c \cdot \nabla G(\mathbf{y}) \cdot \text{sgn}\left(G(\mathbf{y})\right) \tag{57}$$

The parameter c must satisfy the condition

$$c \geq \frac{\|\mathbf{y}\|}{\|\nabla G(\mathbf{y})\|} \tag{58}$$

for the search direction to be a descent direction of the merit function. It may be implemented as

$$c = \gamma \cdot \frac{\|\mathbf{y}_m\|}{\|\nabla G(\mathbf{y}_m)\|} + \eta \tag{59}$$

where $\gamma>1$ and $\eta\geq 0$ with typical values γ=2 and η=10 . Other algorithms are available to solve Eqs. (34), (35), and (36). They include the Gradient Projection algorithm, the Polak-He algorithm, as well as a sequential quadratic programming (SQP) approach.

Gradient Computations

The gradient of the limit-state function, i.e.,

$$\nabla G(\mathbf{y}) \equiv \frac{\partial G(\mathbf{y})}{\partial \mathbf{y}} \tag{60}$$

is a vector that appear frequently in the previous derivations. In particular, this gradient vector is an important ingredient in the iHLRF design point search algorithm. The iHLRF algorithm is therefore called a gradient-based algorithm, along with many other optimization schemes. The evaluation of Eq. (60) requires attention because the limit-state function is not expressed in the standard normal **y**-space but in the original **x**-space. Furthermore, in modern reliability analysis the limit-state function is often expressed in terms of responses **u** from a numerical analysis like the finite element method. Under these circumstances the gradient vector in Eq. (60) is evaluated by the chain rule of differentiation:

$$\frac{\partial G}{\partial \mathbf{y}} = \frac{\partial g}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{y}} \tag{61}$$

The following comments are provided about each factor:

- $\partial g/\partial \mathbf{u}$ is usually trivial because the limit-state function is some algebraic expression in terms of the responses $\mathbf{u}$.
- $\partial \mathbf{u}/\partial \mathbf{x}$ is a vector of response sensitivities from the numerical analysis like the finite element method. Its computation is a subject in the computational stiffness method.
- $\partial \mathbf{x}/\partial \mathbf{y}$ is the inverse of the Jacobian matrix that is computed in the document on probability transformations.

The FORM Analysis Procedure

In light of the previous derivations, the FORM algorithm is summarized as follows:

1. Set m=1
2. Select a starting point in the standard normal space: $\mathbf{y}_m$
3. Transform into the original space: $\mathbf{y}_m \rightarrow \mathbf{x}_m$
4. Evaluate the limit-state function $G(\mathbf{y}_m)=g(\mathbf{x}_m)$
5. Evaluate the gradient of the limit-state function according to Eq. (61)
6. Set the scaling factor for the first convergence criterion: $G_0=G(\mathbf{y}_m)$
7. Check convergence according to Eqs. (39) and (43)
8. If convergence is not achieved, compute $\mathbf{y}_{m+1}$ according to Eq. (38) with search direction and step size defined by, e.g., the iHLRF algorithm
9. Repeat from Step 3 but skip Step 6

Upon converging, compute the reliability index according to Eq. (37) and compute the failure probability from Eq. (29).

Suitability of the FORM Linearization

FORM is a powerful and efficient reliability method that provides physical insight in addition to the probability estimate. It works well for a host of practical problems. However, the analyst should be aware of several potential pitfalls:

- FORM is based on the assumption that the limit-state function is not highly nonlinear, particularly in the vicinity of the design point. This is not a problem with most practical problems rooted in structural analysis, but caution must be exercised, especially in structural dynamics problems.
- When gradient-based design point search algorithms are applied the limit-state function must be continuously differentiable, i.e., the components of the gradient vector must be continuous. This requirement is violated when *max* and *if* statements are incorporated in the definition of the limit-state function. Gradient discontinuities also appear if the limit-state function is defined in terms of responses from a structural analysis that is associated with sudden yielding of elements.
- It has been contended that the failure probability density in the vicinity of the design point is dispersed for high-dimensional problems, i.e., when the number of random variables is in the hundreds or thousands. In that case it is claimed that nonlinearity in the limit-state function will have a significant impact on the FORM estimate of the failure probability because little probability density exist near the design point. However, this concern may be overstated for two reasons. First, the number of random variables must be significant for this problem to appear. Second and more importantly, it is highly unlikely that the nonlinearity in all axis direction is significant and

simultaneously convex or concave, which is the circumstance under which the FORM estimate would deteriorate.

Reliability Sensitivities with respect to Constants

Derivatives of p_f and β with respect to distribution parameters or other parameters of the limit-state function are useful for several reasons. They are employed in "importance measures" that rank the relative importance of different parameters, and they are essential in reliability-based optimization algorithms that are gradient-based. The problem of computing derivatives with respect to parameters of the limit-state function is first addressed. For example, we may seek the derivative of the threshold u_0 in the limit-state function

$$g = u_0 - u(\mathbf{x}) \tag{62}$$

A reliability analysis with this limit-state function essentially yields a point on the complementary cumulative distribution function of the response u, and the derivative $\mathrm{d}p_f/\mathrm{d}u_0$, with a minus sign in front, is therefore a point on the probability density function of u. Typically, u is computed with an array of models and many random variables collected in $\mathbf{x}$. A similar example is differentiation carried out with respect to parameter(s) buried inside the computation of u, which is also covered by the following derivations. First, it is recognized that computation of $\mathrm{d}\beta/\mathrm{d}u_0$ immediately gives $\mathrm{d}p_f/\mathrm{d}u_0$ because

$$p_f = \Phi(-\beta) \tag{63}$$

Thus

$$\frac{\partial p_f}{\partial u_0} = \frac{\partial}{\partial u_0}\Phi(-\beta) = \frac{\partial \beta}{\partial u_0}\frac{\partial}{\partial \beta}\Phi(-\beta) = -\frac{\partial \beta}{\partial u_0}\cdot\varphi(\beta) \tag{64}$$

Next, the problem of computing $\mathrm{d}\beta/\mathrm{d}u_0$ is addressed by recognizing that the reliability index is directly related to the design point coordinates $\mathbf{y}^*$ and applying the chain rule of differentiation:

$$\frac{\partial \beta}{\partial u_0} = \frac{\partial \beta}{\partial \mathbf{y}^*}^{\mathrm{T}}\frac{\partial \mathbf{y}^*}{\partial u_0} \tag{65}$$

The reliability index can be written as $\beta=\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{y}^*$, hence the first factor reads:

$$\frac{\partial \beta}{\partial \mathbf{y}^*} = \boldsymbol{\alpha} \tag{66}$$

Thus, we temporarily arrive at the following expression for the derivative of the reliability index:

$$\frac{\partial \beta}{\partial u_0} = -\frac{\nabla G^{\mathrm{T}}}{\|\nabla G\|}\frac{\partial \mathbf{y}^*}{\partial u_0} \tag{67}$$

To proceed it is necessary to take a detour. In particular, differentiate the limit-state function, G, at the design point, where $G=0$:

$$\frac{\partial G}{\partial u_0} = \frac{\partial G}{\partial \mathbf{y}^*}\frac{\partial \mathbf{y}^*}{\partial u_0} + \left.\frac{\partial G}{\partial u_0}\right|_{\mathbf{y}^*} = \nabla G^T \frac{\partial \mathbf{y}^*}{\partial u_0} + \left.\frac{\partial G}{\partial u_0}\right|_{\mathbf{y}^*} = 0 \tag{68}$$

Then solve for the conditional derivative:

$$\left.\frac{\partial G}{\partial u_0}\right|_{\mathbf{y}^*} = -\nabla G^T \frac{\partial \mathbf{y}^*}{\partial u_0} \tag{69}$$

Now compare Eqs. (67) and (69). The right-hand sides of those two equations are equal, barring the norm of the gradient. Hence, it is concluded that the sought derivative of the reliability index is actually almost equal to the derivative of the limit-state function of the design point. In fact:

$$\frac{\partial \beta}{\partial u_0} = \frac{1}{||\nabla G||}\cdot\left.\frac{\partial G}{\partial u_0}\right|_{\mathbf{y}^*} = \frac{1}{||\nabla G||}\cdot\left.\frac{\partial g}{\partial u_0}\right|_{\mathbf{x}^*} \tag{70}$$

Reliability Sensitivities with respect to Distribution Parameters

Another problem related to reliability sensitivities is to compute the derivative of p_f and β with respect to the mean and standard derivation of the random variables. Once again it is understood that $\mathrm{d}\beta/\mathrm{d}\mu$ and $\mathrm{d}\beta/\mathrm{d}\sigma$ is what we seek, where μ and σ are the mean and standard deviation of a random variable, respectively, because $\mathrm{d}p_f/\mathrm{d}\mu$ and $\mathrm{d}p_f/\mathrm{d}\sigma$ follow directly from Eq. (64). It turns out that

$$\frac{\partial \beta}{\partial \theta} = -\frac{\nabla G^T}{||\nabla G||}\frac{\partial \mathbf{y}^*}{\partial \theta} \tag{71}$$

where θ is the distribution parameter and the last factor in the right-hand side is essentially the derivative of the probability transformation at the design point.

FORM Importance Vectors

Importance vectors are related to reliability sensitivity measures, but they serve a different purpose. Specifically, importance vectors are intended to reveal the relative "importance" of different parameters. The derivatives in the previous section cannot generally be utilized for that purpose, because the derivatives have different units, depending on the units of the intervening random variables. Four importance vectors are available in FORM. First, consider the linearized limit-state function at the design point in the standard normal space in Eq. (30) and obtain its variance by using the technique from the document on functions of random variables:

$$\begin{aligned}\mathrm{Var}[\hat{G}] &= \nabla \mathrm{G}^T \nabla \mathrm{G} \\ &= \left(-||\nabla \mathrm{G}||\cdot\boldsymbol{\alpha}\right)^T\left(-||\nabla \mathrm{G}||\cdot\boldsymbol{\alpha}\right) \\ &= ||\nabla \mathrm{G}||^2\cdot\left(\alpha_1^2+\alpha_2^2+\cdots+\alpha_n^2\right)\end{aligned} \tag{72}$$

This shows that the individual components of the alpha-vector serve as indicators of the importance, i.e., contribution to the total variance, of the corresponding random variables. As a result, the alpha-vector is the primary importance vector for the random variables **y**, i.e., the random variables in the standard normal space. The higher absolute value of α_i,

the more important random variable number *i* is. Furthermore, the sign of the components of the alpha-vector tells whether the random variable is a "load variable" or a "resistance variable:"

$$\begin{aligned} &\alpha_i > 0: \quad y_i \text{ is a load variable} \\ &\alpha_j < 0: \quad y_j \text{ is a resistance variable} \end{aligned} \tag{73}$$

Although $\boldsymbol{\alpha}$ is a good importance vector for the random variables in the standard normal space, **y**, it may not be an accurate importance measure for the original random variables, **x**, when correlation is present. For that reason the following measure is employed to rank the random variables in the original space (Der Kiureghian 2005):

$$\boldsymbol{\gamma} = \frac{\boldsymbol{\alpha}^T \mathbf{J}_{\mathbf{y}^*,\mathbf{x}^*} \hat{\mathbf{D}}}{\left\| \boldsymbol{\alpha}^T \mathbf{J}_{\mathbf{y}^*,\mathbf{x}^*} \hat{\mathbf{D}} \right\|} \tag{74}$$

where the Jacobian matrix is recognized and $\hat{\mathbf{D}}$ is a diagonal matrix with the standard deviations of the "equivalent normal" random variables $\hat{\mathbf{x}}$ on the diagonal. The covariance matrix of $\hat{\mathbf{x}}$ is (Der Kiureghian 2005):

$$\hat{\boldsymbol{\Sigma}}_{\hat{\mathbf{x}}\hat{\mathbf{x}}} = \mathbf{J}^{-1}_{\mathbf{y}^*,\mathbf{x}^*} \mathbf{J}^{-T}_{\mathbf{y}^*,\mathbf{x}^*} \tag{75}$$

Two additional importance vectors can be considered, essentially reliability sensitivities normalized by the standard deviation of the random variables:

$$\boldsymbol{\delta} = \frac{\partial \beta}{\partial \mu_i} \cdot \sigma_i = \nabla_{\boldsymbol{\mu}} \beta^T \mathbf{D}_X \tag{76}$$

and

$$\boldsymbol{\eta} = \frac{\partial \beta}{\partial \sigma_i} \cdot \sigma_i = \nabla_{\boldsymbol{\sigma}} \beta^T \mathbf{D}_X \tag{77}$$

SORM

Essentially, the second-order reliability method (SORM) is an extension of FORM. Specifically, a hyper-paraboloid is employed to approximate the limit-state surface instead of a hyper-plane. The exact expression for the probability mass outside a hyper-paraboloid with distance β from the origin in the uncorrelated standard normal space is

$$p_f = \varphi(\beta) \cdot \mathrm{Re}\left(i \cdot \sqrt{\frac{2}{\pi}} \cdot \int_0^{i\omega} \frac{1}{s} \cdot e^{\left(\frac{(s+\beta)^2}{2}\right)} \cdot \prod_{i=1}^{n-1} \frac{1}{\sqrt{1+\kappa_i s}} ds \right) \tag{78}$$

where *n* is the number of random variables. An alternative expression is the asymptotic approximation

$$p_f \approx \Phi(-\beta) \cdot \prod_{i=1}^{n-1} \frac{1}{\sqrt{1+\psi(\beta)\cdot\kappa_i}} \tag{79}$$

in which the FORM probability is multiplied by a correction factor that depends upon the principal curvatures of the limit-state surface at the design point, κ_i and

$$\psi(\beta) = \frac{\varphi(\beta)}{\Phi(-\beta)} \tag{80}$$

The key question in SORM is how to determine the curvatures, κ_i. There are several options, some of which are outlined in the following sections.

Curvature-based SORM

Suppose the second-derivatives of the limit-state function are not too difficult to compute analytically. Then consider the second-order Taylor series expansion of the limit-state function in the standard normal space:

$$G(\mathbf{y}) \approx \underbrace{G(\mathbf{y}^*)}_{=0} + \nabla G(\mathbf{y}^*)^T(\mathbf{y}-\mathbf{y}^*) + \frac{1}{2}\cdot(\mathbf{y}-\mathbf{y}^*)^T \underbrace{\left[\frac{\partial^2 G}{\partial y_i \partial y_j}\right]}_{=\mathbf{H}} (\mathbf{y}-\mathbf{y}^*) \tag{81}$$

where the Hessian matrix, $\mathbf{H}(\mathbf{y})$, of second-derivatives is identified. Take notice that $\mathbf{H}$ is indeed here defined in the standard normal $\mathbf{y}$-space of random variables. Transformation of the Hessian in the original $\mathbf{x}$-space to the $\mathbf{y}$-space is addressed later. Substitution of the FORM-linearization of the limit-state function into Eq. (81), i.e., writing the linear Taylor series expansion as $G(\mathbf{y}) \approx \|\nabla G\|(\beta - \boldsymbol{\alpha}^T\mathbf{y})$ yields

$$G(\mathbf{y}) \approx \left\|\nabla G(\mathbf{y}^*)\right\|(\beta - \boldsymbol{\alpha}^T\mathbf{y}) + \frac{1}{2}\cdot(\mathbf{y}-\mathbf{y}^*)^T\mathbf{H}(\mathbf{y}-\mathbf{y}^*) \tag{82}$$

To construct a hyper-paraboloid at the design point in the standard normal space the coordinate system is first rotated so that one of the random variable axes align with the alpha-vector. Usually, the last random variable is arbitrarily selected for this purpose. This is shown for the simple example of two random variables in Figure 6. The tilde is used to identify the rotated coordinate system.

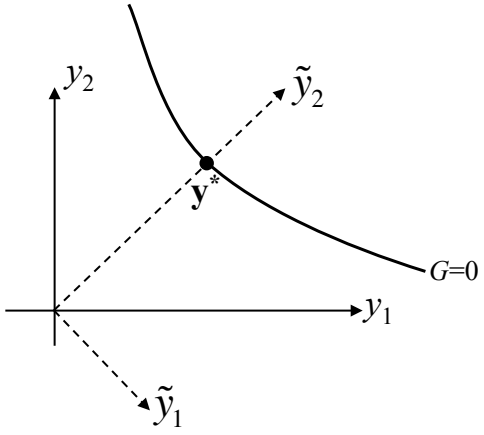

Figure 6: Rotation of the standard normal space to construct hyper-paraboloid.

Mathematically, the rotation of the coordinate system is expressed by means of a rotation matrix, **P**:

$$\tilde{\mathbf{y}} = \mathbf{P}\mathbf{y} \tag{83}$$

where **P** is an *n*-dimensional square matrix with the alpha-vector as its last row, where *n* is the number of random variables. P satisfies the property $\mathbf{P}^{\mathrm{T}}\mathbf{P}=\mathbf{I}$ so that

$$\mathbf{y} = \mathbf{P}^T\tilde{\mathbf{y}} \tag{84}$$

The other rows of **P** are orthogonal to the alpha-vector and orthogonal to each other. They serve the purpose of ensuring a proper Cartesian coordinate system of the rotated system. Algorithms like the Gram-Schmidt procedure are employed to establish the set of orthogonal vectors. Substitution of Eq. (83) into the Taylor approximation in Eq. (81) yields

$$\begin{aligned} G(\tilde{\mathbf{y}}) &\approx \left\|\nabla G(\mathbf{y}^*)\right\|\left(\beta - \boldsymbol{\alpha}^T\left(\mathbf{P}^T\tilde{\mathbf{y}}\right)\right) + \frac{1}{2}\cdot\left(\mathbf{P}^T\tilde{\mathbf{y}} - \mathbf{P}^T\tilde{\mathbf{y}}^*\right)^T\mathbf{H}\left(\mathbf{P}^T\tilde{\mathbf{y}} - \mathbf{P}^T\tilde{\mathbf{y}}^*\right) \\ &= \left\|\nabla G(\mathbf{y}^*)\right\|\left(\beta - \tilde{y}_n\right) + \frac{1}{2}\cdot\left(\tilde{\mathbf{y}} - \tilde{\mathbf{y}}^*\right)^T\mathbf{P}\mathbf{H}\mathbf{P}^T\left(\tilde{\mathbf{y}} - \tilde{\mathbf{y}}^*\right) \end{aligned} \tag{85}$$

where $\tilde{y}_n \equiv \boldsymbol{\alpha}^T\left(\mathbf{P}^T\tilde{\mathbf{y}}\right)$. To further simplify, the limit-state function is scaled by the norm of the gradient at the design point, so that

$$\tilde{G}(\tilde{\mathbf{y}}) \approx \beta - \tilde{y}_n + \frac{1}{2}\cdot\left(\tilde{\mathbf{y}} - \tilde{\mathbf{y}}^*\right)^T\mathbf{A}\left(\tilde{\mathbf{y}} - \tilde{\mathbf{y}}^*\right) \tag{86}$$

where $\tilde{G}(\tilde{\mathbf{y}}) \equiv G(\tilde{\mathbf{y}})/\left\|\nabla G(\mathbf{y}^*)\right\|$ and $\mathbf{A} \equiv \mathbf{P}\mathbf{H}\mathbf{P}^T/\left\|\nabla G(\mathbf{y}^*)\right\|$. The design point coordinates in the rotated coordinate system is

$$\tilde{\mathbf{y}}^* = \left\{\begin{array}{cccc} 0 & 0 & \cdots & \beta \end{array}\right\}^T \tag{87}$$

That allows the matrix multiplication in Eq. (86) to be reduced to

$$\tilde{G}(\tilde{\mathbf{y}}) \approx \beta - \tilde{y}_n + \frac{1}{2}\cdot\tilde{\mathbf{y}}_{\text{cut}}^T\mathbf{A}_{\text{cut}}\tilde{\mathbf{y}}_{\text{cut}} \tag{88}$$

where the subscript "cut" indicates removal of the last vector element and the last matrix row and column. Finally, $\mathbf{A}_{\text{cut}}$ is diagonalized by computing its eigenvalues and Eq. (88) is rewritten as

$$\tilde{G}(\tilde{\mathbf{y}}) \approx \beta - \tilde{y}_n + \frac{1}{2}\cdot\sum_{i=1}^{n-1}\kappa_i\cdot\tilde{y}_i^2 \tag{89}$$

where κ_i are the *n*-1 eigenvalues of $\mathbf{A}_{\text{cut}}$, which are the principal curvatures of the hyper-paraboloid. These are the curvatures that enter into Eq. (79) to compute the failure probability from SORM. A positive curvature, κ_i, implies an outwards-curving limit-state surface in the $\tilde{y}_i, \tilde{y}_n$ coordinate system. If the curvature is negative, it cannot be so large that points on the limit-state surface are closer to the origin than the design point. That would invalidate the design point from FORM. In other words, the absolute value of

negative curvatures cannot exceed the curvature of a circle around the origin with curvature $1/\beta$:

$$\kappa_i > -\frac{1}{\beta} \tag{90}$$

An important task above is to calculate the Hessian. The limit-state function is often defined in terms of numerical responses collected in **u**, which depends upon the input variables **x**. In that case the Hessian is obtained by invoking the product rule and chain rule of differentiation:

$$\begin{aligned}\frac{\partial^2 G}{\partial y_i \partial y_j} &= \frac{\partial}{\partial y_i}\left(\frac{\partial G}{\partial y_j}\right) = \frac{\partial}{\partial y_i}\left(\frac{\partial g}{\partial u_k}\frac{\partial u_k}{\partial x_m}\frac{\partial x_m}{\partial y_j}\right) \\ &= \left(\frac{\partial u_p}{\partial x_q}\frac{\partial x_q}{\partial y_i}\underbrace{\frac{\partial^2 g}{\partial u_p \partial u_k}}_{A}\right)\frac{\partial u_k}{\partial x_m}\frac{\partial x_m}{\partial y_j} + \frac{\partial g}{\partial u_k}\left(\frac{\partial x_q}{\partial y_i}\underbrace{\frac{\partial^2 u_k}{\partial x_q \partial x_m}}_{B}\right)\frac{\partial x_m}{\partial y_j} + \frac{\partial g}{\partial u_k}\frac{\partial u_k}{\partial x_m}\left(\underbrace{\frac{\partial^2 x_m}{\partial y_i \partial y_j}}_{C}\right)\end{aligned} \tag{91}$$

where summation over repeated indices is implied. The matrix identified by *A*, not to be confused with the matrix **A** above, is readily obtained because the limit-state function is usually an algebraic function of the responses **u**. The matrix identified by *B* contains the response sensitivities from the numerical response algorithm. The matrix identified by *C* is obtained by differentiating the probability transformation twice. Suppose the Nataf transformation is used, then the **x**-variables are linked with **z**-variables:

$$F(x_i) = \Phi(z_i) \tag{92}$$

The first differentiation yields:

$$\begin{aligned}&\frac{\partial}{\partial z_j}\Big(F(x_i) = \Phi(z_i)\Big) \\ &\Rightarrow \frac{\partial}{\partial x_k}\frac{\partial x_k}{\partial z_j}F(x_i) = \frac{\partial}{\partial z_j}\Phi(z_i) \\ &\Rightarrow \frac{\partial F(x_i)}{\partial x_k}\frac{\partial x_k}{\partial z_j} = \frac{\partial \Phi(z_i)}{\partial z_j} \\ &\Rightarrow f(x_i)\frac{\partial x_i}{\partial z_i} = \varphi(z_i) \\ &\Rightarrow \frac{\partial x_i}{\partial z_i} = \frac{\varphi(z_i)}{f(x_i)}\end{aligned} \tag{93}$$

where the result is a diagonal matrix because each marginal probability distribution varies only with its own random variable. The second differentiation yields:

$$\begin{aligned}
&\frac{\partial}{\partial z_j}\left(\frac{\partial x_i}{\partial z_i}=\frac{\varphi(z_i)}{f(x_i)}\right)\\
&\Rightarrow \frac{\partial^2 x_i}{\partial z_i \partial z_j}=\frac{\partial \varphi(z_i)}{\partial z_j}\cdot\frac{1}{f(x_i)}+\frac{\partial}{\partial z_j}\left(\frac{1}{f(x_i)}\right)\cdot\varphi(z_i)\\
&\Rightarrow \frac{\partial^2 x_i}{\partial z_i \partial z_j}=\frac{\partial \varphi(z_i)}{\partial z_j}\cdot\frac{1}{f(x_i)}+\frac{\partial}{\partial x_j}\left(\frac{1}{f(x_i)}\right)\cdot\frac{\partial x_j}{\partial z_j}\cdot\varphi(z_i)\\
&\Rightarrow \frac{\partial^2 x_i}{\partial z_i \partial z_j}=\frac{\partial \varphi(z_i)}{\partial z_j}\cdot\frac{1}{f(x_i)}-\frac{1}{f(x_i)^2}\cdot\frac{\partial f(x_i)}{\partial x_j}\cdot\frac{\partial x_j}{\partial z_j}\cdot\varphi(z_i)
\end{aligned} \tag{94}$$

The result is zero for $i \neq j$ because each marginal probability distribution varies only with its own random variable. This means that Eq. (94) is a square diagonal tensor, with dimension equal to the number of random variables, just like Eq. (93) was a square diagonal matrix. In turn, the **z**-variables are linked with the **y**-variables via the Cholesky decomposition of the modified correlation matrix:

$$z_i = L_{ij} y_j \tag{95}$$

Differentiation yields

$$\frac{\partial}{\partial y_k}\left(z_i = L_{ij} y_j\right) \Rightarrow \frac{\partial z_i}{\partial y_k} = L_{ij}\frac{\partial y_j}{\partial y_k} \Rightarrow \frac{\partial z_i}{\partial y_j} = L_{ij} \tag{96}$$

The second differentiation yields zero. We are now ready to substitute the derivatives of the probability transformation into the expressions for the gradient and Hessian. Using the above equations we obtain the gradient vector in the **y**-space as follows:

$$\frac{\partial G}{\partial y_i} = \frac{\partial g}{\partial u_j}\cdot\frac{\partial u_j}{\partial x_k}\cdot\frac{\partial x_k}{\partial z_l}\cdot\frac{\partial z_l}{\partial y_i} = \frac{\partial g}{\partial u_j}\cdot\frac{\partial u_j}{\partial x_k}\cdot\left[\frac{\varphi(z)}{f(x)}\right]_{kl}\cdot L_{li} \tag{97}$$

where square brackets are used to denote a diagonal matrix. Next we reconsider Eq. (91), now including the intermediate **z**-variables:

$$\begin{aligned}
\frac{\partial G}{\partial y_i \partial y_j} &= \frac{\partial}{\partial y_j}\left(\frac{\partial g}{\partial u_k}\cdot\frac{\partial u_k}{\partial x_l}\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial z_m}{\partial y_i}\right)\\
&= \frac{\partial}{\partial y_j}\left(\frac{\partial g}{\partial u_k}\right)\cdot\frac{\partial u_k}{\partial x_l}\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial z_m}{\partial y_i}+\frac{\partial g}{\partial u_k}\cdot\frac{\partial}{\partial y_j}\left(\frac{\partial u_k}{\partial x_l}\right)\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial z_m}{\partial y_i}\\
&+\frac{\partial g}{\partial u_k}\cdot\frac{\partial u_k}{\partial x_l}\cdot\frac{\partial}{\partial y_j}\left(\frac{\partial x_l}{\partial z_m}\right)\cdot\frac{\partial z_m}{\partial y_i}+\frac{\partial g}{\partial u_k}\cdot\frac{\partial u_k}{\partial x_l}\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial}{\partial y_j}\left(\frac{\partial z_m}{\partial y_i}\right)\\
&= \left(\frac{\partial^2 g}{\partial u_n \partial u_k}\cdot\frac{\partial u_n}{\partial x_o}\cdot\frac{\partial x_o}{\partial z_p}\cdot\frac{\partial z_p}{\partial y_j}\right)\cdot\frac{\partial u_k}{\partial x_l}\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial z_m}{\partial y_i}+\frac{\partial g}{\partial u_k}\cdot\left(\frac{\partial^2 u_k}{\partial x_l \partial x_n}\cdot\frac{\partial x_n}{\partial z_o}\cdot\frac{\partial z_o}{\partial y_j}\right)\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial z_m}{\partial y_i}\\
&+\frac{\partial g}{\partial u_k}\cdot\frac{\partial u_k}{\partial x_l}\cdot\left(\frac{\partial^2 x_l}{\partial z_m \partial z_n}\cdot\frac{\partial z_n}{\partial y_j}\right)\cdot\frac{\partial z_m}{\partial y_i}+\frac{\partial g}{\partial u_k}\cdot\frac{\partial u_k}{\partial x_l}\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial^2 z_m}{\partial y_i \partial y_j}
\end{aligned} \tag{98}$$

The last term is zero because the second differentiation of **z**=**Ly** is zero. To further simplify, consider explicit limit-state functions where g depends directly on **x** without any **u**-variables. Eq. (98) then reads:

$$\begin{aligned}\frac{\partial G}{\partial y_i \partial y_j} &= \frac{\partial^2 g}{\partial x_l \partial x_n}\cdot\frac{\partial x_n}{\partial z_o}\cdot\frac{\partial z_o}{\partial y_j}\cdot\frac{\partial x_l}{\partial z_m}\cdot\frac{\partial z_m}{\partial y_i}+\frac{\partial g}{\partial x_l}\cdot\frac{\partial^2 x_l}{\partial z_m \partial z_n}\cdot\frac{\partial z_n}{\partial y_j}\cdot\frac{\partial z_m}{\partial y_i} \\ &= \frac{\partial^2 g}{\partial x_l \partial x_n}\cdot\left[\frac{\varphi(z)}{f(x)}\right]_{no}\cdot L_{oj}\cdot\left[\frac{\varphi(z)}{f(x)}\right]_{lm}\cdot L_{mi} \\ &+\frac{\partial g}{\partial x_l}\cdot\left[\frac{\partial\varphi(z)}{\partial z}\cdot\frac{1}{f(x)}-\frac{1}{f(x)^2}\cdot\frac{\partial f(x)}{\partial x}\cdot\frac{\partial x}{\partial z}\cdot\varphi(z)\right]_{lmn}\cdot L_{nj}\cdot L_{mi}\end{aligned} \tag{99}$$

Caution must be exercised when conducting matrix multiplications to evaluate Eq. (99). The first term has factors with indices *ln*, *no*, *oj*, *lm*, *mi*. Here one can first multiply the three first matrices *ln*, *no*, *oj* giving *lj*. The last to matrices *lm*, *mi* are multiplied to give *li*. Hence, the final value of the first term is the *lj* matrix transposed times the *li* matrix. The second term has factors with indices *l*, *lmn*, *nj*, *mi*. The product of the first two is an *mn* matrix with values $\left(\partial g/\partial x\right)\cdot\left(\partial^2 x/\partial z^2\right)$ along the diagonal. Then what remains in the second term is *mn*, *nj*, *mi*. To multiply this out with matrix products the last of them, *mi*, is transposed and placed first to give the matrix product *im*, *mn*, *nj*. These calculations are demonstrated for the CalREL limit-state function on the example-webpage. In summary, the procedure for curvature-fitted SORM is:

1. Have an analytical expression for the limit-state function
2. Differentiate it twice to obtain gradient and Hessian
3. Establish the **A**-matrix use an orthogonalization algorithm
4. Perform eigenvalue analysis to obtain principal curvatures
5. Update the failure probability using Eq. (79)

This approach is straightforward but it is limited to explicit limit-state functions, or at least limit-state functions for which second-derivative computations are possible. Obtaining those derivatives can be computationally expensive, and the results may be prone to numerical noise in the calculation of the limit-state function.

Gradient-based SORM

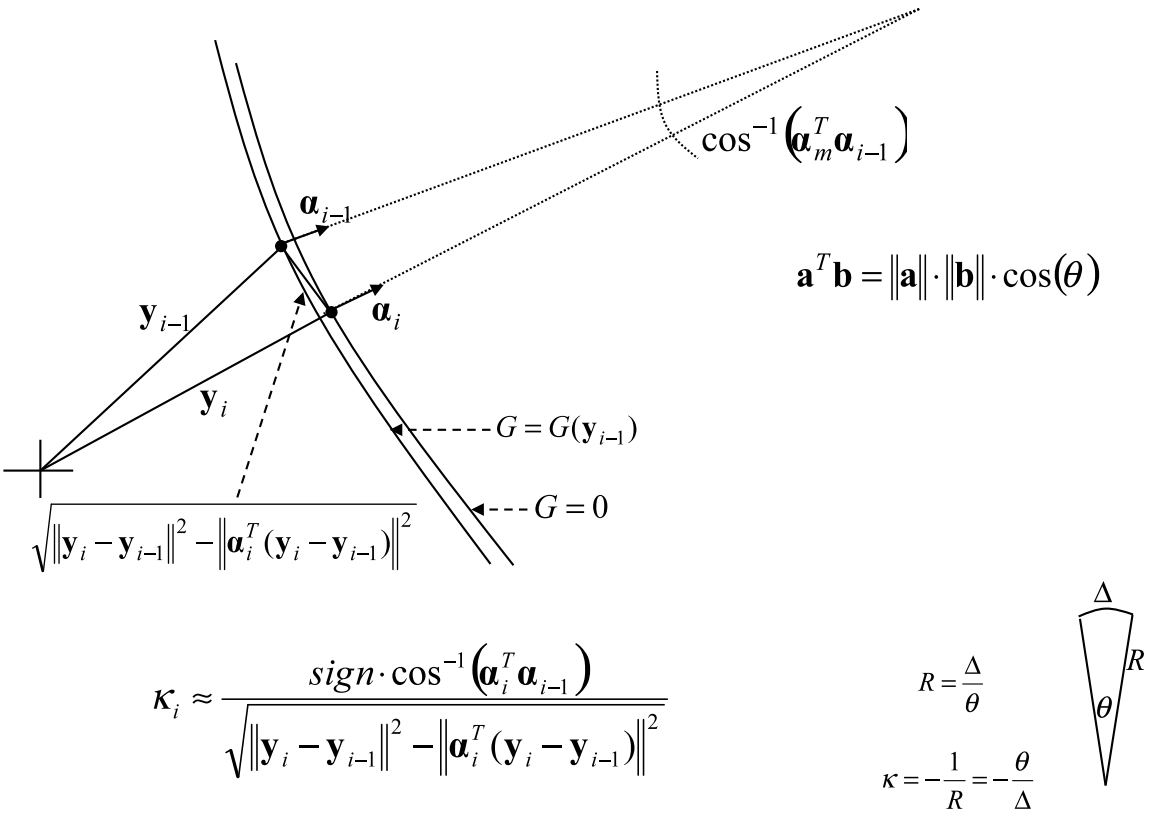

Figure 7: Geometrical considerations for gradient-based SORM.

References

Der Kiureghian, A. (2005). "First- and second-order reliability methods." *Engineering Design Reliability Handbook*, E. Nikolaidis, D. M. Ghiocel, and S. Singhal, eds., CRC Press.