RC 20254 (4/29/96) Revised
Engineering and Technology

IBM Research Report

Techniques for Composing a Class of Statistical Tolerance Zones

V. Srinivasan, M.A.O'Connor

IBM Research Division
T.J. Watson Research Center
Yorktown Heights, New York

F.W.Scholz

Boeing Information and Support Services
Seattle, Washington

IBM Research Division
Almaden · T.J. Watson · Tokyo · Zurich

1 Introduction

In 1994 the American Society of Mechanical Engineers issued its latest revision of the dimensioning and tolerancing standard (ASME, 1994). About three out of a total of over two hundred and thirty pages in this standard are devoted to statistical tolerancing. Previous revisions issued earlier contained no reference to statistical tolerancing at all. Though minimal, the latest attempt by ASME to address statistical tolerancing is a significant beginning. Widespread practice of statistical tolerancing within American companies, and those abroad, have forced ASME and other international bodies to take a serious look at possible codification of statistical tolerancing. Recently ISO has set up a task group towards this purpose (Srinivasan and O'Connor, 1995). Indications thus far point to an increased level of activity in understanding and codification of statistical tolerancing.

An earlier paper (Srinivasan and O'Connor, 1994) proposed possible interpretations of statistical tolerancing in the context of geometric part specifications. That study was based on our understanding of the prevailing manufacturing practices in leading companies. An important class of interpretations depended on process capability indices: C_p, C_{pk}, and C_c. This led us to define statistical tolerance zones in the C_{pk}-C_p plane as well as in the μ-σ plane to specify what populations are statistically acceptable. The focus thus far has been strictly on part specifications and their interpretations.

Since most products are assemblies of parts, it is not sufficient to consider only part specifications. Ideally, a designer should start with an assembly "budget" for allowable variation, and distribute it to the constituent parts. In reality, this problem is attacked by several iterations of tolerance analysis where part variations are composed to determine the assembly variation. In statistical tolerance analysis, this means that techniques should be found to compose the part-level specifications that are in the form of statistical tolerance zones. In this chapter, we provide such mathematical and computational techniques.

What is the use of composing part-level statistical tolerance zones? The composed zone gives us a compact geometric representation of "all" possible statistical outcomes of a critical assembly-level characteristic. If we have the composed tolerance zone for an assembly-level characteristic, we can then reason more rationally about the risk that some instances of the product may not function properly. In the absence of the composed tolerance zone, people have often resorted to heuristics to estimate the assembly-level failure rates that are hard to defend or explain, and even grossly erroneous. It is the combination of the attraction of a compact geometric representation of composed statistical tolerance zones and the possibility of subsequent risk analyses using these tolerance zones that makes our work relevant.

Section 2 describes the type of part-level statistical tolerance zones covered in this chapter. The actual task of composing the part specifications into assembly specification is addressed in Section 3. Some thoughts on the risk analysis of the assembled product based on the composition is the topic of Section 4. An example illustrating the use of the techniques is given in Section 5. The mathematical details involved in the composition is presented in the Appendix.

2 Statistical Tolerance Zones

Let x be a random variable, and LSL and USL be the lower and upper specification limits. If μ and σ are the mean and standard deviation of x, then

$$C_p = \frac{USL - LSL}{6\sigma}$$

and

$$C_{pk} = \min\left\{\frac{\mu - LSL}{3\sigma}, \frac{USL - \mu}{3\sigma}\right\} = \frac{\frac{USL-LSL}{2} - |\frac{LSL+USL}{2} - \mu|}{3\sigma}$$

are known as the process capability indices. In addition, it is useful to define

$$C_c = \frac{|\mu - \frac{LSL+USL}{2}|}{\frac{USL-LSL}{2}}$$

to quantify the mean shift from the target value of $(LSL + USL)/2$. The variable x can be derived from a collection of actual parts (or features) by Gaussian or Chebyschev fitting. In this case, x is called an *actual value*. A particular type of actual value is the *actual mating size*, which is found to be very useful for assembly analysis.

In many industrial practices a population of acceptable parts (or features) is statistically specified by

$$C_p \geq P, \;\; C_{pk} \geq K, \;\; \text{and} \;\; C_c \leq F$$

for particular values P, K, and F, or some subset of these inequalities. This leads to a set of acceptable (μ, σ) pairs, that is, a statistical tolerance zone, hereafter referred to as an STZone. It can be shown that as a subset of the μ-σ plane it is a polygonal region between the μ-axis and the graph of a function defined piecewise by finitely many non-negative linear functions with bounded domains (Srinivasan and O'Connor, 1994). Such an STZone will be said to be σ-polygonal. Table 1 summarizes the results of our survey of current part-level statistical tolerancing practices in five leading companies around the world. The names of the companies have been suppressed for reasons of confidentiality. All the STZones in the table are σ-polygonal.

One may construct a more complete statistical tolerance zone in the parametric space of the population of parts. The parameters can be, for example, all the (central) moments of the distribution. An STZone defined above may then be viewed as the intersection of the μ-σ plane with such a higher dimensional statistical tolerance zone.

When two or more parts are assembled to form a product, we would like to know the STZone for a product characteristic. When some product characteristic is a linear combination of statistically independent primitive part-level characteristics, we can easily and explicitly provide this STZone. All primitive part-level STZones will be assumed to be σ-polygonal in what follows.

3 Composing STZones

Let a linear "gap" function be

$$G = a_0 + a_1X_1 + a_2X_2 + ... + a_nX_n, \tag{1}$$

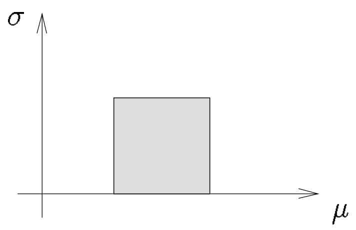
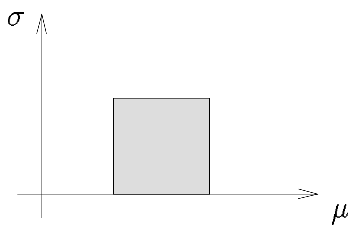
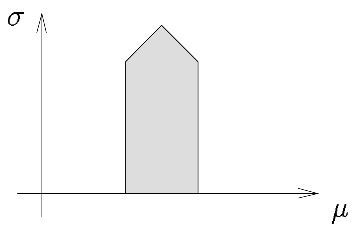
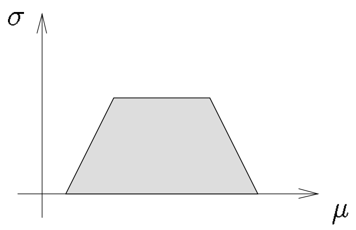
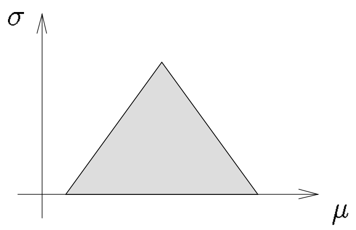
Company	Practice	Shape of STZone
A	$C_p \geq 2.0, C_{pk} \geq 1.5, C_c \leq 0.25$	
B	$C_p \geq 2.0, C_{pk} \geq 0.75C_p$	
C	$C_{pk} \geq 1.0, C_c \leq 0.20$	
D	$C_p \geq 2.0, C_{pk} \geq 1.5$	
E	$C_{pk} \geq 1.33$	

Table 1: Summary of current part-level statistical tolerancing practices in companies surveyed.

where $X_1, X_2, ..., X_n$ are independent random variables with σ-polygonal STZones. Without loss of generality, equation (1) can be rewritten in terms of $g = G - a_0$ and $x_i = a_i X_i$ as

$$g = x_1 + x_2 + ... + x_n,$$

where the x_i's are independent random variables. Hereafter, for simplicity of statement, we assume that all compositions are of this form, that is, linear with coefficients equal to unity involving only finitely many independent variables. Since x_i is a simple linear transformation of X_i, we see that the STZone for x_i is also σ-polygonal, because

$$\left\{ \begin{array}{c} \mu_i \\ \sigma_i \end{array} \right\} = \left[\begin{array}{cc} a_i & 0 \\ 0 & |a_i| \end{array} \right] \left\{ \begin{array}{c} \mu_{X_i} \\ \sigma_{X_i} \end{array} \right\} \tag{2}$$

where μ_i and σ_i are the mean and standard deviation of x_i, and μ_{X_i} and σ_{X_i} those of X_i. Our task is to find the STZone for g; we call this the problem of composition. To accomplish this, we first move from the μ-σ plane to the μ-σ^2 plane, where the composition is reduced to simple Minkowski sums. To see this, we will first define Minkowski sums.

The Minkowski sum of two sets A and B in $\mathbf{R}^n$ is defined as

$$A \oplus B = \{a + b : a \in A, b \in B\}.$$

Minkowski sums are also known as vector sums, for obvious reason. If we are summing up several independent random variables, there is a simple rule from elementary probability theory that states that the mean of the sum is the sum of the means, and the variance of the sum is also the sum of the variances. Applying this simple rule we can see that if we work in the mean-variance (that is, μ-σ^2) plane, the composition problem becomes one of computing the vector sum (that is, the Minkowski sum) of the STZones for the random variables in that plane. Minkowski sums are commutative and associative. They also distribute over unions, that is,

$$A \oplus (B \cup C) = (A \oplus B) \cup (A \oplus C).$$

For more details on Minkowski sums, see Matheron (1975) and Kaul (1993).

3.1 Composition in the μ-σ^2 Plane

Since σ is always non-negative, each (μ, σ) pair is uniquely associated with a (μ, σ^2) pair by squaring σ or choosing the positive square root of σ^2. We can thus represent any STZone in either the μ-σ plane or the μ-σ^2 plane and pass trivially from one to the other. We claim that the STZone for x_i in the μ-σ^2 plane is bounded by line segments and parabolic arcs. Indeed, if we start with an inclined line in the μ-σ plane, given by

$$\sigma = m\mu + c, m \neq 0, \infty,$$

then when transformed to the μ-σ^2 plane, it becomes the parabola

$$\sigma^2 = (m\mu + c)^2 = m^2(\mu + \frac{c}{m})^2. \tag{3}$$

A line segment in the line with non-negative σ values is transformed to a parabolic arc, which can include the apex (the point closest to the directrix) of the parabola only as an endpoint, so that the arc is an increasing arc or a decreasing arc, that is, a monotonic arc. On the other hand, a horizontal line in the μ-σ plane with a constant σ value of $\sigma_0 > 0$ is transformed to a horizontal line in the μ-σ^2 plane with a constant σ^2 value of σ_0^2, and any line segment of the line in the μ-σ plane is tranformed to a line segment of the line in the μ-σ^2 plane. It follows that the upper boundary of the STZone for x_i in the μ-σ^2 plane is composed of bounded non-negative horizontal line segments and monotonic arcs of graphs of non-negative parabolic functions. The STZone itself is the region between the μ-axis and this collection of arcs. For example, the house-shaped STZone in the μ-σ plane practiced by Company C in Table 1 leads to the region in the μ-σ^2 plane of Figure 1.

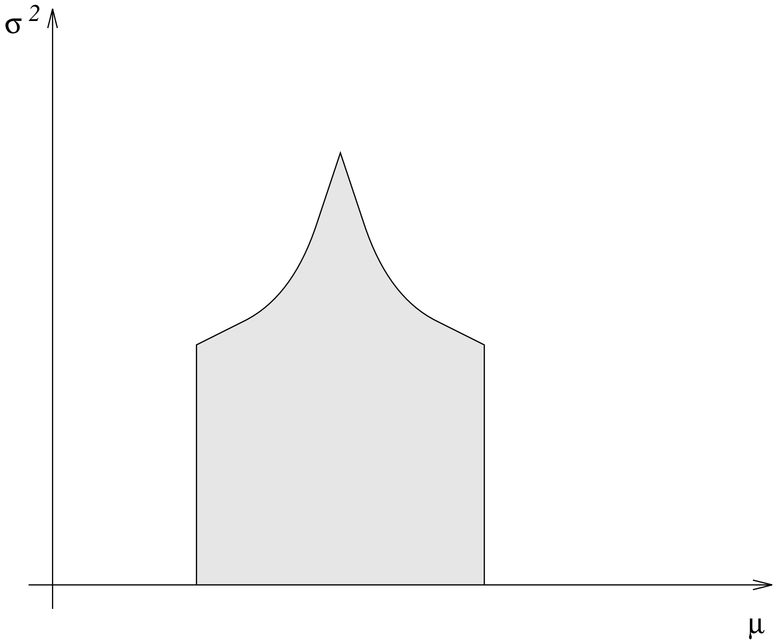


Figure 1: A statistical tolerance zone in the μ-σ^2 plane.

In analogy to the property of being σ-polygonal, an STZone will be called σ^2-parabolic, if it is represented in the μ-σ^2 plane by a region between the μ-axis and the graph of a function defined piecewise by finitely many bounded non-negative horizontal line segments and bounded monotonic subarcs of non-negative parabolic functions. In these terms we have shown that a σ-polygonal STZone will always be σ^2-parabolic.

Since the random variables are independent, we have

$$\left\{ \begin{array}{c} \mu_g \\ \sigma_g^2 \end{array} \right\} = \left\{ \begin{array}{c} \mu_1 \\ \sigma_1^2 \end{array} \right\} + \left\{ \begin{array}{c} \mu_2 \\ \sigma_2^2 \end{array} \right\} + \ldots + \left\{ \begin{array}{c} \mu_n \\ \sigma_n^2 \end{array} \right\},$$

where μ_g and σ_g are the mean and standard deviation of g. Thus the rule of composition in the μ-σ^2 plane reduces to

Lemma 1 *In the μ-σ^2 plane the STZone of g is equal to the Minkowski sum of the STZones of the x_i's.*

In general, finding the representation of the STZone of g in the μ-σ^2 plane by applying this last result requires computing the Minkowski sum of regions bounded by parabolic arcs. Computing the Minkowski sum of regions bounded by nonlinear arcs can be difficult, at best. The algebraic complexity of the arcs bounding the sum may grow (Kaul, 1993), so that in a sum involving many summands, as is common in compositions of STZones, the complexity may become prohibitive. However, in the case of sums of σ^2-parabolic zones these problems do not arise. The sums themselves are easy to obtain, and the algebraic complexity of the sum is unchanged. In particular, Corollary 1 of the appendix and the discussion following it imply the following two lemmas.

Lemma 2 *The composition of σ^2-parabolic STZones is σ^2-parabolic.*

Lemma 3 *Let A be a collection of finitely many σ^2-parabolic STZones. Let $\mathcal{P}$ be the collection of arcs formed by translating each defining arc in the upper boundary of each member of A by each endpoint of each defining arc in the upper boundary of every other member of A. The upper envelope of $\mathcal{P}$ is the upper boundary of the representation of the composition of the STZones of A in the μ-σ^2 plane.*

The first of the lemmas implies that σ^2-parabolic STZones are closed under the Minkowski sum of any finite number of them. The second implies that the arcs of the upper boundary of a Minkowski sum of σ^2-parabolic STZones merely come from translates of the arcs bounding the summands. These results permit the explicit calculation of the representation in the μ-σ^2 plane of the STZone of a composition of σ-polygonal STZones. An example that involves two STZones is shown in the top row of Figure 2. The upper boundary of the Minkowski sum consists of parabolic arcs that are mere translates of those that form the upper boundary of the summands. The number of such arcs on the upper boundary of the sum may increase in some cases, as illustrated in Figure 12, but their algebraic complexity still remains the same. To obtain the representation in the μ-σ plane we continue the analysis in the next section.

3.2 Composition in the μ-σ Plane

Translation of the parabola of (3) by (X, Y) in the μ-σ^2 plane yields the parabola

$$\sigma^2 = m^2(\mu - X + \frac{c}{m})^2 + Y. \tag{4}$$

In the μ-σ plane (4) is satisfied by a nondegenerate hyperbola, when $Y > 0$, and the degenerate hyperbola, a product of two lines,

$$0 = (m\mu - mX + c + \sigma)(m\mu - mX + c - \sigma)\ , \tag{5}$$

when $Y = 0$.

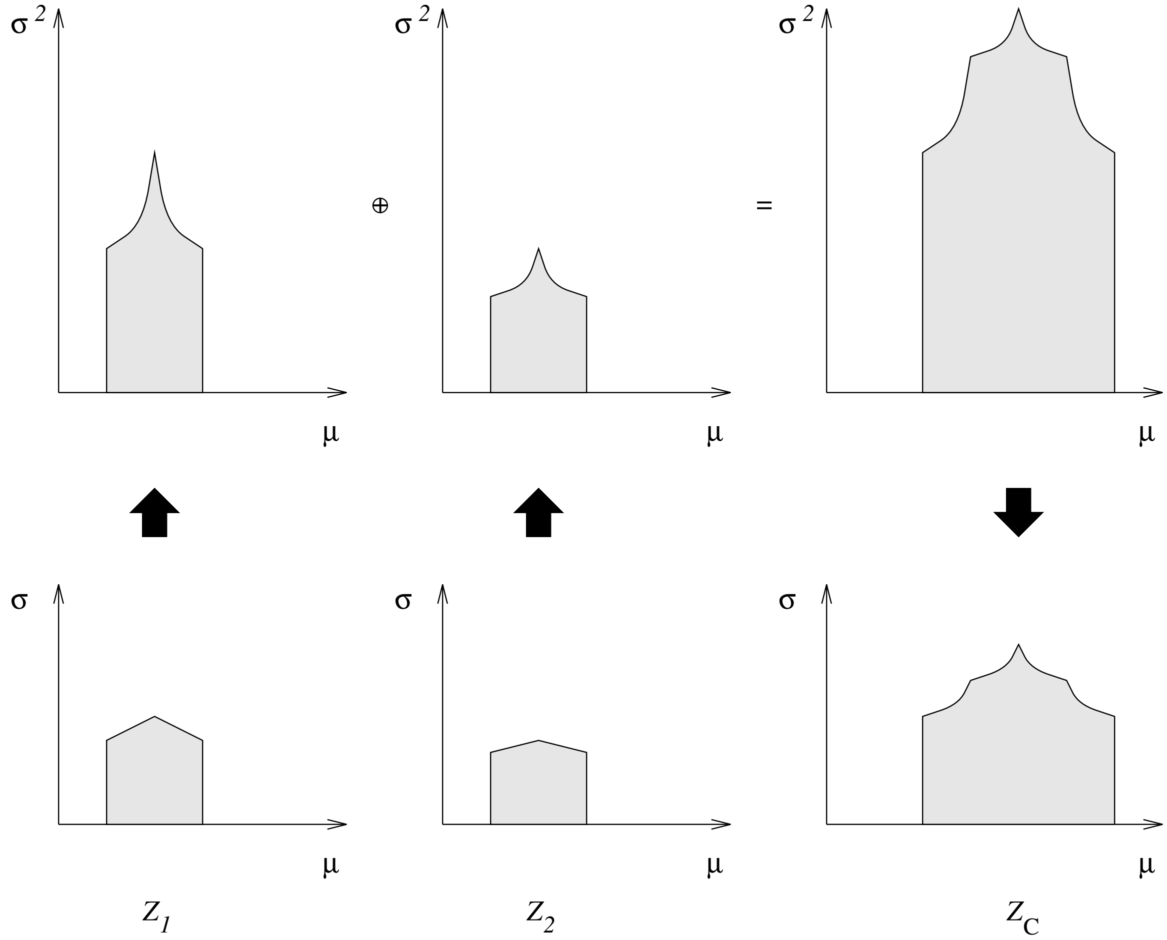


Figure 2: An illustration of composing two STZones. The bottom row shows the STZones Z_1 and Z_2 of x_1 and x_2, respectively, and also their composition in the form of the STZone Z_C of $g = x_1 + x_2$, all in the μ-σ plane. In the process of composition, we transform the STZones of x_1 and x_2 from the μ-σ plane to the μ-σ^2 plane, and obtain their Minkowski sum as shown in the top row. This sum is then transformed back to the μ-σ plane to get the desired composition.

First consider the case when $Y > 0$. If we transform the parabola of (4) in the μ-σ^2 plane to the μ-σ plane, it then becomes the upper curve of the nondegenerate hyperbola defined by (4). In this case, a monotonic arc of the parabola is transformed to a monotonic arc of the upper curve of the nondegenerate hyperbola.

On the other hand, when $Y = 0$, the parabola of (4) is mapped to the two half-lines with non-negative σ-values defined by (5). If we call these two half-lines the upper curve of the degenerate hyperbola, then a monotonic arc of the parabola becomes a line segment in the upper curve of the degenerate hyperbola.

Conversely, any monotonic arc of the upper curve of a hyperbola defined by an equation of the form of (4) with $Y \geq 0$ in the μ-σ plane transforms to a monotonic arc of a non-negative parabola of the μ-σ^2 plane. See Figure 3 for an illustration. Proceeding as before, let a subset of the μ-σ plane be called σ-hyperbolic, if it is a region between the μ-axis and the graph of a function defined piecewise by finitely many bounded horizontal non-negative line segments or bounded monotonic subarcs of the upper curves of a hyperbolas determined by equations of the form of (4) with $Y \geq 0$. In these terms we have shown

Lemma 4 *An STZone is σ^2-parabolic, if and only if it is σ-hyperbolic.*

Since a σ-polygonal STZone is always σ^2-parabolic, Lemmas 2 and 4 yield part of the characterization of the STZone for g we have been pursuing.

Proposition 1 *The STZone of a composition of σ-polygonal STZones is σ-hyperbolic.*

In fact, Lemmas 2 and 4 say more.

Proposition 2 *The STZone of a composition of σ-hyperbolic STZones is σ-hyperbolic.*

This means, loosely put, that σ-hyperbolic STZones are closed under the operation of composition of STZones. This is of more than merely theoretical interest. Since STZones often represent parts in an assembly, which in turn are often subassemblies in a larger assembly, the last result assures that no new analysis is required to treat this more general application. To complete the characterization we need only to make more explicit the piecewise-hyperbolic function which defines the STZone of g. This is accomplished by the following

Proposition 3 *Let A be a collection of finitely many σ-hyperbolic STZones. If*

$$\mu \in D \rightarrow \sqrt{(m\mu - c)^2 + d} \tag{6}$$

is a parameterization of one of the defining arcs of the upper boundary of one of the members of A, and (a,b) is an endpoint of a defining arc of the upper boundary of any other member of A, define the parameterization of a new arc by

$$\mu \in D + a \rightarrow \sqrt{(m\mu - c - ma)^2 + d + b}. \tag{7}$$

Let $\mathcal{H}$ be the collection of all possible such newly defined arcs. The upper envelope of $\mathcal{H}$ is the upper boundary of the representation of the composition of the STZones of A in the μ-σ plane.

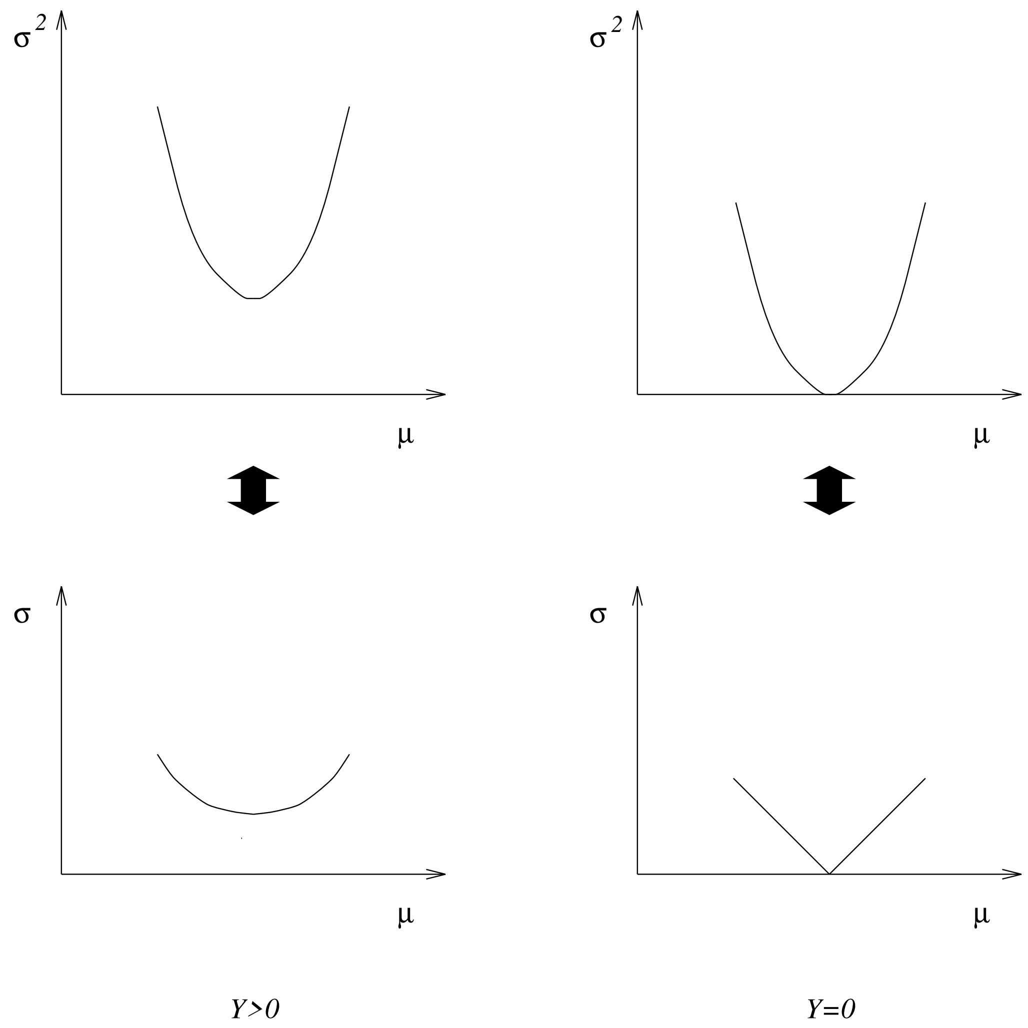


Figure 3: Transformation between μ-σ and μ-σ^2 planes. When $Y > 0$ in equation (4), a parabola in the μ-σ^2 plane goes to the upper curve of a hyperbola in the μ-σ plane (as σ is always non-negative), and conversely. When $Y = 0$, the hyperbola degenerates into two half-lines.

Proof. Passing from the μ-σ plane to the μ-σ^2 plane, applying Lemma 3, calculating the effect of the translations, and passing back from the μ-σ^2 plane to the μ-σ plane suffice to establish the claim. ■

If one of the STZones is σ-polygonal with linear parameterizations, "squaring" each of the parameterizations yields ones compatible with the statement of the proposition. If $m = 0$, the hyperbolic arcs of (6) and (7) reduce to line segments. If $d = 0$, the hyperbolic arc of (6) reduces to a line segment, and if $b = d = 0$, the hyperbolic arc of (7) reduces to a line segment. All other arcs come from nondegenerate hyperbolas.

A calculation of the upper envelope called for in the last proposition requires the intersection of hyperbolic arcs. The intersection of two hyperbolas usually requires the solution of a quartic polynomial; however, the special form of (4) reduces this complexity to that of a simple quadratic. The bottom row in Figure 2 shows an example of the composed STZone Z_C for $g = x_1 + x_2$ in the μ-σ plane.

3.3 Special Case Simplifications

In this section to this point we have considered only the most general composition problem and only from a theoretical viewpoint. Even in the general case many simplifications are possible for algorithmic purposes. For example, judicious use of the observation following Lemma 5 of the appendix that sums of increasing and decreasing functions need only use part of the vertical boundaries can decrease the resultant number of potential upper boundary pieces by a factor of four. Treating monotonic chains by a simple generalization of the lemma can add further reductions. Similar reductions are possible for pairs of increasing boundary elements.

In less general settings, prescribed design practices often allow obvious simplifications. For example, symmetric part-level STZones produce a symmetric assembly-level STZone, so that only half of the upper boundary of the STZone need be computed. More restrictive, though common, practices can lead to very explicit compositions. For example, let us assume that all parts are specified by $C_{pk} \geq K$ and $C_c \leq F$, as practiced by Company C of Table 1, with some globally fixed K and F. For ease of statement let us further assume that the nominal values of all part dimensions equal zero, so that the specification for any actual value, x_i, becomes

$$\frac{T_i - |\mu_i|}{3\sigma_i} = C_{pk} \geq K \;\; \text{and} \;\; \frac{|\mu_i|}{T_i} = C_c \leq F \; ,$$

where T_i is the USL of x_i. Finally, let us assume that $\{x_i\}_{i=1}^n$ is ordered such that $T_i \leq T_j$ for all $i \leq j$. If $g = \sum_{i=1}^n x_i$, let us say that g is a restricted type C composition defined by K, F, and $\{x_i, T_i\}_{i=1}^n$. In these terms we now have

Proposition 4 *Let g be a restricted type C composition defined by K, F, and $\{x_i, T_i\}_{i=1}^n$. Let*

$$f(0) = \frac{1}{3K}(\sum_{i=1}^n T_i^2)^{1/2}$$

and

$$f(\mu) = \frac{1}{3K}\left((1-F)^2\sum_{i=1}^{k-1} T_i^2 + T_k^2(1-\theta F)^2 + \sum_{i=k+1}^{n} T_i^2\right)^{1/2}$$

for $|\mu| = F\left(\sum_{i=1}^{k-1} T_i + \theta T_k\right)$ *with* $0 < \theta \leq 1$. *The graph of f is the upper boundary of the representation of the STZone of g in the μ-σ plane.*

For the μ coordinate of any point in Z_C, the STZone of g in the μ-σ plane, there is a unique representation of $|\mu|$ as required in this proposition, so that the result completely and explicitly yields Z_C. We provide only a sketch of the proof of the proposition here.
Let

$$D_\mu = \left\{(\mu_1, \ldots, \mu_i, \ldots, \mu_n) \ : \ 0 \leq \mu_i \leq FT_i \text{ for all } 1 \leq i \leq n \text{ and } \sum_{i=1}^{n} \mu_i = |\mu|\right\}.$$

The point, $(\mu, \tilde{\sigma})$, is in the upper boundary of Z_C, if and only if $9K^2\tilde{\sigma}^2$ is the maximum of $h(\mu_1, \ldots, \mu_n) = \sum_{i=1}^{n}(T_i - \mu_i)^2$ over D_μ. This maximum occurs at an extreme point of D_μ, that is, a point in D_μ with $\mu_i = 0$ or $\mu_i = FT_i$ for all i except possibly one with $\mu_i = \theta FT_i$. Since

$$h(\mu_1, \ldots, \mu_i + \delta, \ldots, \mu_j - \delta, \ldots, \mu_n) - h(\mu_1, \ldots, \mu_i, \ldots, \mu_j, \ldots, \mu_n) = 2\delta h_1$$

for $h_1 = \mu_i - \mu_j + \delta + T_j - T_i$, a simple case analysis, comparing extreme points differing only on coordinate pairs, and considering the form of h_1 shows that h never decreases as we pass in proper direction from extreme point to extreme point to the claim of the proposition.

4 Risk Analysis

As we pointed out in the Introduction, a major motivation for computing the composed STZone is to perform assembly-level risk analysis more rationally. In this section we outline several ways to exploit the composed STZone to perform such analyses. The main scope of this chapter is not the risk analyses, but we include them here to show the usefulness of composed STZones.

Risk analysis entails evaluating or approximating the probability that the value of some random variable from some collection of random variables will fall outside an acceptable region. One-sided risk analysis considers probabilities of the form $Pr(g \leq k)$ (or the completely analogous $Pr(g \geq k)$), and two-sided risk analysis the sum of probabilities $Pr(g \leq k) + Pr(g \geq k')$, for random variables g and fixed real values k and k'. Of course, to provide such an analysis sufficient information about the distributions of the random variables must be available.

An STZone uses only the mean and standard deviation to define a class of random variables. This offers too little distributional information to say much about risk, in general. However, by imposing mild assumptions on the class of random variables, we can change the situation. For this purpose, if we assume that each random variable in the class of random variables defined by an STZone Z is completely determined by its mean and standard deviation, so that we can parameterize the random variables of the class by $(\mu, \sigma) \in Z \to g(\mu, \sigma)$, and that any variable of the class can be transformed to any other by shifting μ and scaling σ, so that for all (μ, σ) and (μ', σ') in Z

$$Pr(g(\mu, \sigma) \leq k) = Pr\left(\frac{\sigma}{\sigma'}(g(\mu', \sigma') - \mu') + \mu \leq k\right), \tag{8}$$

then much more can be said.

Let us first consider the one-sided risk $Pr(g(\mu, \sigma) \leq k)$ associated with $(\mu, \sigma) \in Z$. We begin by observing that (8) implies immediately that

$$Pr(g(\mu, \sigma) \leq k) = Pr\left(g(\mu', \sigma') \leq \frac{\sigma'}{\sigma}(k - \mu) + \mu'\right). \tag{9}$$

It follows directly from this that for $\mu' \neq k$

$$Pr(g(\mu, \sigma) \leq k) = Pr(g(\mu', \sigma') \leq k), \text{ whenever } \sigma = \frac{\sigma'(k - \mu)}{k - \mu'},$$

so that each of the points in the intersection of a ray of the type

$$\left\{\left(\mu, \frac{\sigma'(k - \mu)}{k - \mu'}\right) \;:\; \mu > k \text{ if } \mu' > k\right\} \text{ or } \left\{\left(\mu, \frac{\sigma'(k - \mu)}{k - \mu'}\right) \;:\; \mu < k \text{ if } \mu' < k\right\}$$

with Z is associated with the same risk. Each of these rays is a nonvertical ray in the $\sigma > 0$ half-plane emanating from the point $(k, 0)$. The vertical ray emanating from $(k, 0)$ is the set of points $\{(k, \sigma) : \sigma > 0\}$. Since shifting the mean and scaling the standard deviation does not change the probability of the value of a random variable being less than the mean, (8) implies that there is a constant α such the $Pr(g(\mu, \sigma) \leq \mu) = \alpha$ for all μ and σ, and in particular $Pr(g(k, \sigma) \leq k) = \alpha$. Thus each of the rays in the half-plane emanating from $(k, 0)$ is associated with a fixed risk. If μ increases without bound with σ and k fixed, then (9) implies that the risk associated with (μ, σ) decreases to a limiting value of zero. If μ decreases without bound with σ and k fixed, then (9) implies that the risk associated with (μ, σ) increases to a limiting value of one. Hence, as we rotate the rays through $(k, 0)$ counterclockwise from the horizontal through the vertical to the horizontal again, sweeping out all the points in the half-plane in the process, the risk associated with the rays increases from zero at the horizontal through α at the vertical to one at the horizontal. We can also express this in terms of the slopes of the rays, which will be most convenient. As the slope of the ray increases from zero to the vertical and then through all negative slopes to zero again, the risk increases from zero to α to one. See Figure 4 for an illustration.

Determining the maximum one-sided risk, $Pr(g(\mu, \sigma) \leq k)$ over $(\mu, \sigma) \in Z$ is now a simple matter. If $k > \mu$ for some $(\mu, \sigma) \in Z$, then rays emanating from $(k, 0)$ with negative slopes arbitrarily close to zero must intersect Z, so the risk is one. If k equals the minimum of all μ such that there exists some $(\mu, \sigma) \in Z$, then the risk is α. If $k < \mu$ for all $(\mu, \sigma) \in Z$, we need only determine the line through $(k, 0)$ with maximum slope that intersects Z, that is, the line of support for Z through $(k, 0)$, other than the μ-axis. This can be computed easily for σ-hyperbolic STZones. By definition the hyperbolic arcs in the boundary of Z come from the upper curves of hyperbolas define by equations of the form of (4). No interior point of such an arc can be on a line

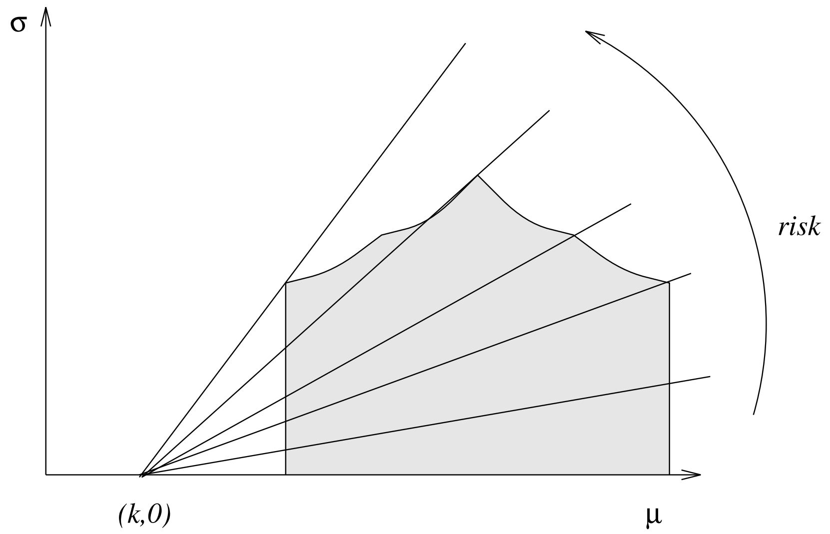


Figure 4: Illustration of isorisk contours in the STZone Z. The risk is the probability $Pr(g \leq k)$. The maximum risk is associated with a supporting line of Z passing through $(k, 0)$.

of support for Z, unless the arc is linear. If the arc is linear, and an interior point of it is on the line of support for Z, then so are the end points. Thus, only the finitely many end points of the arcs are needed to define the line of support through $(k, 0)$. If Z comes from a composition of STZones, that is $Z = Z_C$, an explicit evaluation of Z_C is not even required, as Proposition 3 enables us to obtain these end points directly. In any event, if (μ', σ') is a point defining this line of support, then $Pr(g(\mu', \sigma') \leq k)$ is the maximum one-sided risk.

We wish to stress again that the assumptions made in this development are mild. They are met by many common families of random variables. For example, assuming the random variables to be normal more than suffices. In particular, for sum aggregates, like $g = x_1 + ... + x_n$, which form Z_C, normality is a reasonable and common assumption. In the case of normality, and many others, then, a final evaluation of the maximum one-sided risk can be easily found by appeal to widely available tables.

Finding the analogous maximum two-sided risk, $Pr(g(\mu, \sigma) \leq k) + Pr(g(\mu, \sigma) \geq k')$, over $(\mu, \sigma) \in Z$, is more difficult. A simple extension of the analysis of the one-sided case shows that the two-sided risk is maximum for some (μ, σ) in the upper boundary of Z; however, the risk associated with $(\mu, \sigma) \in Z$ need no longer be monotonic in a simple parameter like the slope of a line, and the simple extension does not allow any claim that the risk must be maximized by an end point of an arc. Thus rather than the simple calculation of the maximum slope over finite set in the one-sided case, we are left with a one-dimensional nonlinear optimization problem in the two-sided case. Approximation techniques exist and may suffice in many cases, but more work on this remains.

In some situations, maximum risk analysis may be overly pessimistic. This may be particularly so, if Z is an assembly-level STZone. If individual and joint distributions can reasonably be assumed on the part-level means, then a distribution of the assembly-level means is implied. Often this assembly-level distribution can be calculated or approximated. For example, if the part-level means are independent and normal, then the assembly-level means are normal. Or if part-level means have independent, uniform distributions (or more generally, piecewise-polynomial densities), then the assembly-level distribution has a piecewise-polynomial density which can be explicitly calculated. It may even be justified to directly assume that the assembly-level means have some distribution, for example, a normal distribution by an application of the central limit theorem. In any case, if we assume that the distribution of the means in Z is known, this distribution can be used to provide a less pessimistic risk analysis. For this purpose let $Pr(\mu \in I) = p$ for some interval I, and let $Z_I = \{(\mu, \sigma) \in Z : \mu \in I\}$. If we apply the previous techniques to Z_I to obtain a maximum risk, r, over Z_I, then we can claim that with probability p the risk is no worse than r. This reduced maximum risk analysis would be useful, when very unlikely μ values far from the mean of the means determine a highly pessimistic risk assessment.

It is worthwhile to point out the distinction between the two types of probabilities p and r. The risk r concerns the fall out rate for fixed levels of (μ, σ). On the other hand, p concerns the chance that μ falls within certain limits. For one particular process this chance is taken just once. A long run frequency interpretation is only possible over many such processes.

Even the reduced maximum risk analysis in certain cases may be less useful than information about an average risk R, that is,

$$R = \int_Z R(\mu, \sigma) dF(\mu, \sigma),$$

where $R(\mu, \sigma)$ is the risk associated with (μ, σ), and dF is the joint probability measure for μ and σ over Z. It is quite possibly too much to assume full knowledge of the joint probability, but since either the one-sided or the two-sided risk is increased by increasing σ, we can bound R by

$$R \leq \int_\Delta R(\phi(\mu)) dH(\mu),$$

where Δ is the intersection of Z with the μ-axis, ϕ is a parameterization of the upper boundary of Z, and dH is the probability measure associated with μ. The simple explicit form of ϕ that can be obtained from Proposition 3 implies that at least for the one-sided risk analysis the integral can be easily approximated by standard techniques.

5 An Example

To illustrate the techniques described thus far, let us consider as a concrete example the statistical specifications of a magnetic storage product. Figure 5 shows a vertical section view of a disk-drive assembly. Four critical parts in this assembly are identified for our consideration: a magnetic disk which stores the data, an arm that swings in an angular movement to enable a magnetic head to read and write data onto the disk, and two bearings that support the disk and the arm. This figure is a grossly simplified version of an actual disk-drive cross-sectional view, but is sufficient for our purpose of illustration.

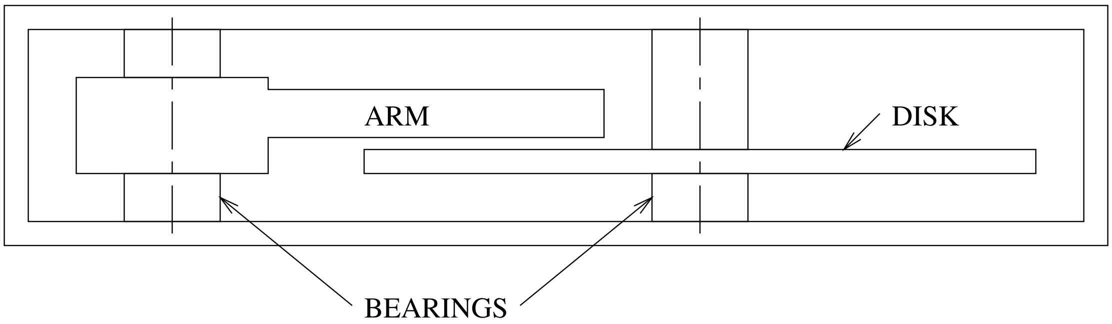


Figure 5: A disk-drive assembly indicating critical parts

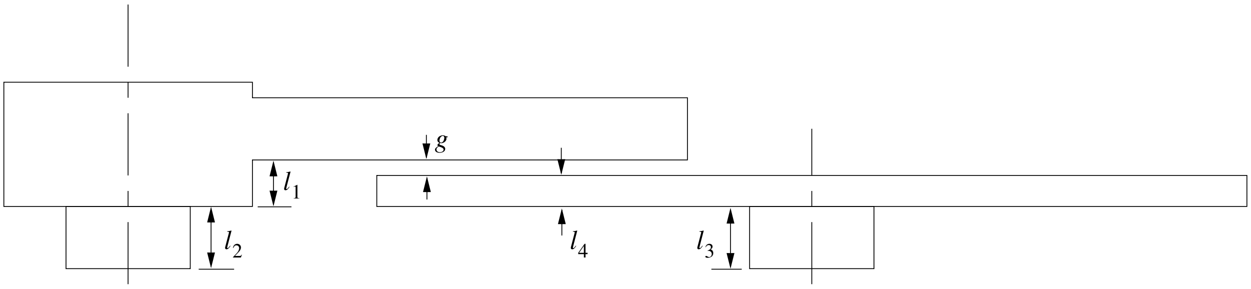


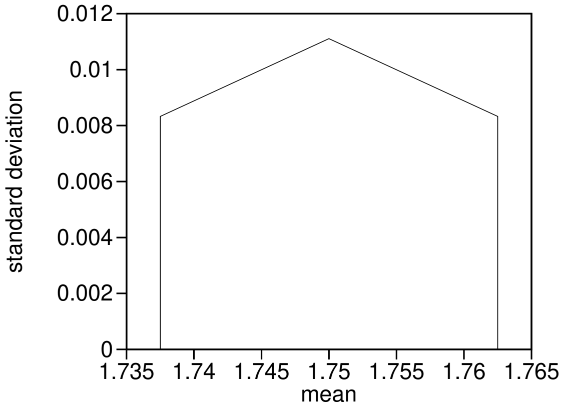
Figure 6: Critical part dimensions and an assembly characteristic in a disk-drive assembly. The critical assembly characteristic is the arm-to-disk spacing g which is related to the critical part dimensions l_1, l_2, l_3, and l_4 by the linear "gap" function $g = l_1 + l_2 - l_3 - l_4$.

A critical assembly characteristic in such a disk-drive is the arm-to-disk clearance, shown in Figure 6 as the gap g. For proper functioning of the disk-drive, this clearance g should be neither too large nor too small. It is, of course, controlled by the critical dimensions of the four parts we identified earlier. More precisely, we have an instance of the linear "gap" function (1) in the form

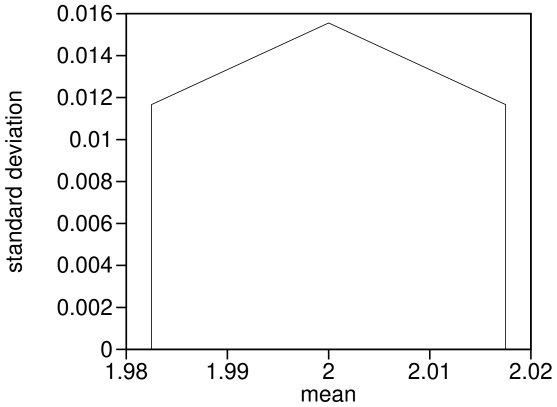
$$g = l_1 + l_2 - l_3 - l_4. \tag{10}$$

Each of the critical dimensions belongs to a different part in the disk-drive assembly, and can be justifiably assumed to be mutually independent random variables.

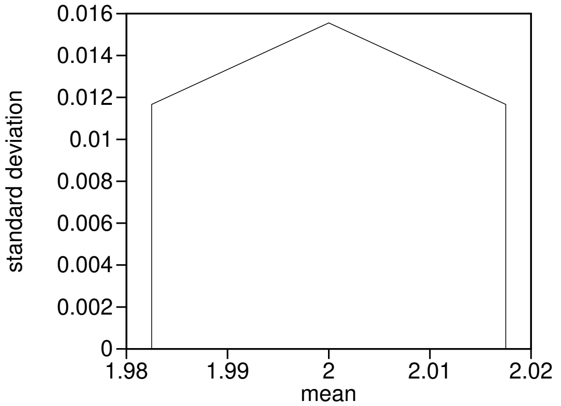
Part-level statistical specifications of these critical dimensions are summarized in Table 2. We have chosen a practice similar to that of Company C described in Table 1. For ease of explanation, the lower bound K for C_{pk} and the upper bound F for C_c have been chosen to be the same, 1.5 and 0.25 respectively, for all the four critical dimensions. STZones that correspond to these statistical specifications are plotted in Figure 7. The STZone for l_1 in Figure 7.1 is the region under the piecewise linear function of $\mu \in [1.7375, 1.7625]$. These limits for μ are obtained from the constraint that $C_c = \frac{|\mu - 1.75|}{0.05} \leq 0.25$. The other constraint that $C_{pk} = \frac{0.05 - |1.75 - \mu|}{3\sigma} \geq 1.5$ leads



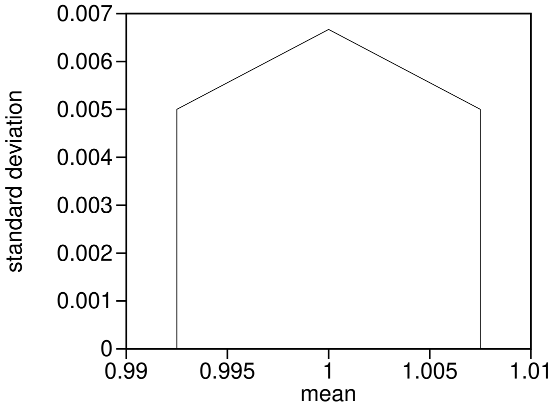
(1) STZone for l_1



(2) STZone for l_2



(3) STZone for l_3



(4) STZone for l_4

Figure 7: Plots of the STZones for the four critical dimensions.

Part	C.D.	N.D.	T.I.	USL	LSL	K	F
Arm	l_1	1.75	0.10	1.80	1.70	1.5	0.25
Arm_Bearing	l_2	2.00	0.14	2.07	1.93	1.5	0.25
Disk_Bearing	l_3	2.00	0.14	2.07	1.93	1.5	0.25
Disk	l_4	1.00	0.06	1.03	0.97	1.5	0.25

Table 2: Statistical specifications on critical dimensions of the disk-drive example. C.D. = Critical Dimension, N.D. = Nominal Dimension, T.I. = Tolerance Interval.

explicitly to

$$\sigma(\mu) = \begin{cases} \frac{\mu - 1.70}{4.5} & \text{for } \mu \in [1.7375, 1.7500] \\ \frac{1.80 - \mu}{4.5} & \text{for } \mu \in (1.7500, 1.7625] \end{cases}$$

as the piecewise linear function. Similarly Figures 7.2 and 7.3 involve the piecewise linear function

$$\sigma(\mu) = \begin{cases} \frac{\mu - 1.93}{4.5} & \text{for } \mu \in [1.9825, 2.0000] \\ \frac{2.07 - \mu}{4.5} & \text{for } \mu \in (2.0000, 2.0175] \end{cases}$$

and Figure 7.4 has

$$\sigma(\mu) = \begin{cases} \frac{\mu - 0.97}{4.5} & \text{for } \mu \in [0.9925, 1.0000] \\ \frac{1.03 - \mu}{4.5} & \text{for } \mu \in (1.0000, 1.0075] \end{cases}$$

as its piecewise linear function. Notice the change in scales in these plots, from the mean to the standard deviation axes as well as from dimension to dimension, adopted for clarity.

The next task is to compute the composed STZone for g. Since we have set the same bounds $K = 1.5$ and $F = 0.25$ for all the four variables, after rearranging the critical dimensions in the increasing order of tolerance intervals, the conditions of Proposition 4 are met. Therefore, the piecewise hyperbolic function, which is the upper boundary of the STZone of g, is given by Proposition 4 explicitly as

$$\sigma(\mu) = \begin{cases} \frac{1}{4.5}\sqrt{0.00466875 + 0.0049(1 - 0.25\frac{0.7125-\mu}{0.0175})^2} & \text{for } 0.6950 \le \mu < 0.7125 \\ \frac{1}{4.5}\sqrt{0.00681250 + 0.0049(1 - 0.25\frac{0.7300-\mu}{0.0175})^2} & \text{for } 0.7125 \le \mu < 0.7300 \\ \frac{1}{4.5}\sqrt{0.01030625 + 0.0025(1 - 0.25\frac{0.7425-\mu}{0.0125})^2} & \text{for } 0.7300 \le \mu < 0.7425 \\ \frac{1}{4.5}\sqrt{0.01230000 + 0.0009(1 - 0.25\frac{0.7500-\mu}{0.0075})^2} & \text{for } 0.7425 \le \mu < 0.7500 \end{cases}$$

and

$$\sigma(\mu) = \begin{cases} \frac{1}{4.5}\sqrt{0.01230000 + 0.0009(1 - 0.25\frac{\mu - 0.7500}{0.0075})^2} & \text{for } 0.7500 \leq \mu \leq 0.7575 \\ \frac{1}{4.5}\sqrt{0.01030625 + 0.0025(1 - 0.25\frac{\mu - 0.7575}{0.0125})^2} & \text{for } 0.7575 < \mu \leq 0.7700 \\ \frac{1}{4.5}\sqrt{0.00681250 + 0.0049(1 - 0.25\frac{\mu - 0.7700}{0.0175})^2} & \text{for } 0.7700 < \mu \leq 0.7875 \\ \frac{1}{4.5}\sqrt{0.00466875 + 0.0049(1 - 0.25\frac{\mu - 0.7875}{0.0175})^2} & \text{for } 0.7875 < \mu \leq 0.8050 \end{cases}.$$

Figure 8 shows a plot of the composed STZone for g. At the resolution in which it is plotted, the

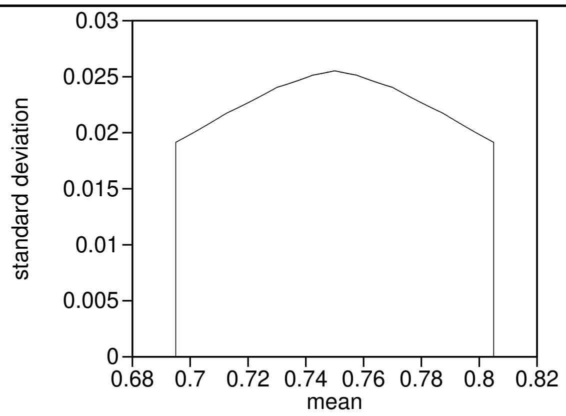
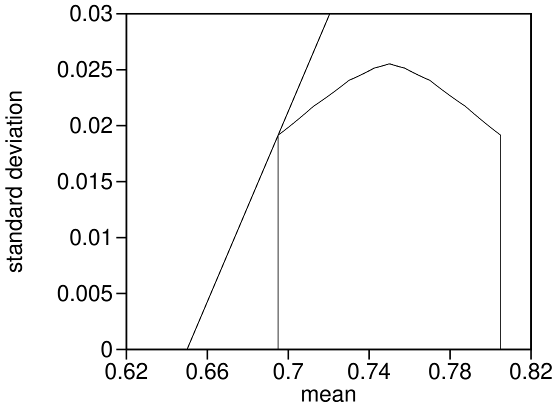


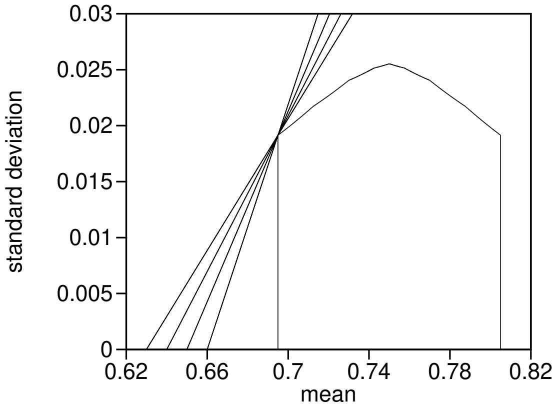
Figure 8: Composed STZone for the arm-to-disk clearance g.

curvatures of the hyperbolic arcs in the upper envelope of the STZone are barely visible.

Once we have the composed STZone risk analyses can be performed in many ways, as pointed out in a previous section. First we need to make an assumption about the distribution of the arm-to-disk clearance g, and we will assume that it is normal. We may appeal to the central limit theorem to justify that normality is a good approximation. If we are concerned about this clearance reaching below a critical value of, say, 0.65 units, we can estimate the one-sided risk as the probability $Pr(g(\mu, \sigma) \leq 0.65)$. This risk will vary as the (μ, σ) point varies within the STZone for g. However, following discussions of the previous section we know that the maximum risk is attained at the tangent point (μ, σ) on the supporting line of the STZone of g passing through $(0.65, 0)$. This tangency occurs at one of the end points of the hyperbolic arcs of the upper boundary of the STZone. Thus, since $\sigma(\mu)$ is increasing for $\mu \leq 0.75$ and decreasing thereafter, only the five endpoints $\{(0.6950, 0.01915)$, $(0.7125, 0.02174)$, $(0.7300, 0.02405)$, $(0.7425, 0.02515)$, $(0.7500, 0.02553)\}$ are possible points of tangency. Visual inspection (see Figure 9.1 and Figure 4 for comparison) indicates that and simple calculations prove that tangency occurs at (0.6950,0.01915), so that maximum risk occurs there. Using easily available approximations for the error function we



(1) One lower limit for g



(2) Several lower limits for g

Figure 9: Maximum risk analysis based on supporting lines.

find that this maximum risk is approximately 9.4 parts per thousand. If we want to estimate the maximum risk for a different critical value k of the lower limit for the arm-to-disk clearance g, we first find the supporting line that passes through $(k, 0)$. A simple calculation shows that for $0.566 < k < 0.695$ the supporting lines pass through the same tangent point $(0.695, 0.01915)$. Then the associated risk is given by $Pr(z \leq \frac{k-0.695}{0.01915})$ of a unit normal variate z. Figure 9.2 illustrates this for several values of k. Figure 10 shows a plot of the risk for a small range of k. The designer

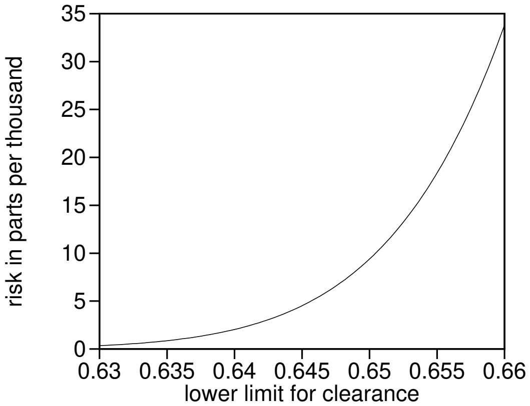


Figure 10: Variation of the maximum risk as a function of the lower limit for the arm-to-disk clearance g.

can then decide if such risks are acceptable, and if not, start the re-design process. If other types of risk analyses are needed, then the STZone for g provides a basis for them.

6 Summary and Conclusions

National and international standards committees are currently investigating how to codify the statistical specifications of part tolerances. Because of the widespread use of process capability indices (C_p, C_{pk}, and C_c) in industry, a case can be made for using them in standardizing the statistical tolerancing of parts. However, it will not be a strong case in the absence of techniques that the designers can use to infer assembly-level variation from such part specifications. In this chapter we provided the needed mathematical and computational techniques.

In developing these techniques we depended only on two major assumptions: that the assembly characteristic is a linear function of the part characteristics and that the part variations are mutually independent. With these we showed that the STZones in the μ-σ plane can be composed by first transforming them to the μ-σ^2 plane, where the composition is reduced to Minkowski sums, and then transforming the sum back to the μ-σ plane. We proved that the algebraic complexity of the Minkowski sum remains low and unchanged irrespective of the number of parts that participate in the assembly. More precisely, the STZone for the assembly-level characteristic is bounded only by line-segments and hyperbolic arcs. This enabled us to present an exact and explicit representation for the composed STZone. In a somewhat restricted, but useful, special case an explicit expression for the piecewise hyperbolic curve that forms the upper boundary of the composed STZone was given. Under some mild assumptions, the one-sided isorisk contours become line-segments in the composed STZones so that designers can reason about the chances they take with such statistical designs. We also outlined several avenues for exploring the risk analysis further. Thus a critical assembly analysis gap has been, at least partially, filled.

As we pointed out in the body of the chapter, the risk analysis may be refined further. Other advances may be attempted, if deemed important, in at least three directions: the geometric form of primitive part-level STZones may be enriched, and the assumptions of linearity and independence in the parts-to-product characteristics relation may be relaxed. The first is perhaps the easiest. If the geometric shape of the primitive part-level STZones differs from those in Table 1, but is still σ-polygonal (or, more generally, σ-hyperbolic), then all of our results apply and are sufficient to achieve the composition. STZones in the form of semi-circular disks, proposed in some literature, can be composed. For more difficult cases, linear approximations may be used to reduce them to σ-polygonal cases, which we know how to handle.

7 Acknowledgment and a Disclaimer

We gratefully acknowledge numerous colleagues in industry, and in the ASME, and ISO standards committees for valuable information and support. The opinions expressed in this chapter are, however, our own and not those of ASME, ISO, or any of their member bodies.

Appendix

In this appendix we present and prove a collection of mathematical properties of certain Minkowski sums that support the main chapter. Results presented here are more general than needed in the body of the chapter. However, we found that the proofs are no more difficult in the general case and hope that they may lead to new applications in the future. We begin with some terminology.

Recall (Theorem 4.1, Part IV of Valentine (1964)) that a closed set $S \subseteq \mathbf{R^2}$ with nonempty interior is convex, if and only if for each x in ∂S, the boundary of S, there is a line of support of S at x, that is, a line through x which defines an open half-plane that is disjoint from S.

For a piecewise-continuous function $f : Dom(f) \rightarrow \mathbf{R^2}$ with $Dom(f) \subset \mathbf{R}$ let

$$U(f) = \{(x, y) \in \mathbf{R^2} \,:\, x \in Dom(f) \text{ and } y \geq f(x)\}\,.$$

Note that the graph of f is a subset of the boundary of $U(f)$ and that the piecewise continuity of f assures that $U(f)$ will have interior points. If f is also non-negative, define $L(f)$, the region below f, as

$$L(f) = \{(x, y) \in \mathbf{R^2} \,:\, x \in Dom(f) \text{ and } 0 \leq y \leq f(x)\}\,.$$

Call a continuous function f convex, if $\overline{U(f)}$ is convex, (where bar indicates closure.) Recall that a function is monotonic if either it is an increasing function on its entire domain or it is a decreasing function on its entire domain. Let a simple function be a non-negative continuous monotonic function defined on a closed bounded interval with at most one root. For a simple function f with domain, $[a, b]$, define $V(f)$, the vertical part of the boundary of $L(f)$ as

$$V(f) = \{(a, y) \in \mathbf{R^2} \,:\, 0 \leq y \leq f(a)\} \cup \{(b, y) \in \mathbf{R^2} \,:\, 0 \leq y \leq f(b)\}\,.$$

$V(f)$ is composed of two closed vertical line segments or a point on the x-axis and a closed vertical line segment. The continuity of f implies that the interior of $L(f)$ is a nonempty open set bounded by $V(f)$, the graph of f, and those points in $L(f)$ with y-coordinate equal to 0. In these terms we have the following principal lemmas.

Lemma 5 *If f is a simple increasing function and g is a simple decreasing function, then*

$$L(f) \oplus L(g) = \{L(f) \oplus V(g)\} \cup \{V(f) \oplus L(g)\}\,.$$

Proof. Let $[a, b]$ be the domain of f and $[c, d]$ that of g. If $p \in L(f) \oplus L(g)$, then there exist $(x, y) \in L(f)$ and $(u, v) \in L(g)$ with $p = (x, y) + (u, v)$. Trivially, $p = (x + t, y) + (u - t, v)$ for all $t \in \mathbf{R}$, and in particular, for $t \geq 0$. If $t \geq 0$, then since f is increasing, $f(x + t) \geq f(x) \geq y$, whenever $x + t \in [a, b]$, so that $(x + t, y) \in L(f)$. Similarly, $(u - t, v) \in L(g)$, whenever $u - t \in [c, d]$. Thus, if we let t_0 be the minimum of $b - x$ and $u - c$, then either $(x + t_0, y) \in V(f)$ and $(u - t_0, v) \in L(g)$ or $(x + t_0, y) \in L(f)$ and $(u - t_0, v) \in V(g)$. In either case, we have shown that $p \in \{L(f) \oplus V(g)\} \cup \{V(f) \oplus L(g)\}$, so that $L(f) \oplus L(g) \subset \{L(f) \oplus V(g)\} \cup \{V(f) \oplus L(g)\}$. The opposite inclusion is trivial. ■

First, note that since $V(f)$ consists of points and vertical line segments, the computation of each Minkowski sum in the union is trivially obtained by a translation of the defining functions. Next note that the proof in fact shows that not all of $V(f)$ and $V(g)$ are required. The right part of $V(f)$ and the left part of $V(g)$ would suffice. This is illustrated in Figure 11.

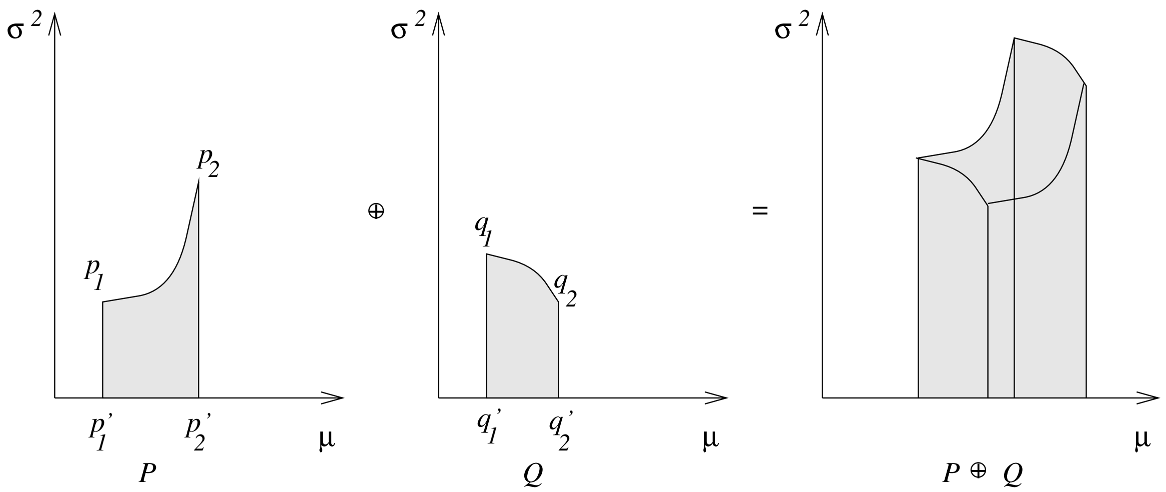


Figure 11: Minkowski sum of a region below a simple increasing function and one below a simple decreasing function. If $l(p,q)$ denotes the line-segment between points p and q, then Lemma 5 states that $P \oplus Q$ can be obtained as the union of $P \oplus l(q_1, q_1')$, $P \oplus l(q_2, q_2')$, $Q \oplus l(p_1, p_1')$, and $Q \oplus l(p_2, p_2')$. However, unioning only the first and the last of the summands will suffice in this case.

Lemma 6 *If f and g are simple convex functions, then*

$$L(f) \oplus L(g) = \{L(f) \oplus V(g)\} \cup \{V(f) \oplus L(g)\}.$$

Proof. If for $S \subset \mathbf{R^2}$, we let $S_- = \{(x,y) : (-x,y) \in S\}$, then for subsets $D, E \subset \mathbf{R^2}$, $D \oplus E = (D_- \oplus E_-)_-$. Thus we may assume that f is increasing, or otherwise we may reflect the problem around the $y-$axis. If g is decreasing, then the claim follows from the preceding lemma, so we may also assume that g is increasing.

Let $[a,b]$ be the domain of f and $[c,d]$ that of g. If $p \in L(f) \oplus L(g)$, then there exist $(x,y) \in L(f)$ and $(u,v) \in L(g)$ with $p = (x,y)+(u,v)$. Trivially, $p = (x, y+t)+(u, v-t)$ for all $t \in R$. If τ_1 is the minimum of $f(x)-y$ and v, then either $y+\tau_1 = f(x)$ and $0 \leq v-\tau_1 \leq g(u)$, or $0 \leq y+\tau_1 \leq f(x)$ and $v - \tau_1 = 0$, so that either p is the sum of $(x, f(x))$ and $(u, v+y-f(x)) \in L(g)$, or p is the sum of $(x, y+v) \in L(f)$ and $(u, 0)$. We now discuss these two cases separately.

First consider the case where p is the sum of $(x, y+v) \in L(f)$ and $(u, 0)$. If we let τ_2 be the minimum of $b-x$ and $u-c$, then either $x+\tau_2 = b$ and $c \leq u-\tau_2 \leq d$, or $a \leq x+\tau_2 \leq b$

and $u - \tau_2 = c$. Whether $x + \tau_2 = b$ or $u - \tau_2 = c$, $y + v \leq f(x) \leq f(x + \tau_2)$, since f is increasing. Thus either p is the sum of $(b, y + v)$ in $V(f)$ and $(u + x - b, 0)$ in $L(g)$, or p is the sum of $(x + u - c, y + v)$ in $L(f)$ and $(c, 0)$ in $V(g)$, establishing the claim in this case.

Now consider the case where p is the sum of $(x, f(x))$ and $(u, v + y - f(x)) \in L(g)$. For simplicity of notation let $w = v + y - f(x)$. We may assume that neither $f(x) = 0$ nor $w = 0$, since these conditions would be covered by the previous case by relabeling f and g, if necessary. We may also assume that $x \neq a, b$ and that $u \neq c, d$, or nothing remains to be demonstrated. The convexity of f implies the existence of a line of support M for $U(f)$ at $(x, f(x))$. Since M is a line of support, it could be vertical only at a or b, so that M is not vertical. We can thus parameterize M as $\{(x, \gamma(x)) : x \in \mathbf{R}\}$ for some linear function γ with $\gamma(x) \leq f(x)$ for all x in the domain of f. It follows that the intersection of M and $L(f)$ is a closed bounded line segment containing $(x, f(x))$ in its interior with end points in $V(f)$ or having a y-coordinate of 0.

Let N be the line through (u, w) parallel to M. We wish to show the existence of a half-line N^+ of N emanating from (u, w), such that N^+ intersected with $L(g)$ is a closed interval with (u, w) as one end point and a point in $V(g)$ or one with a y-coordinate of 0, as the other. If $w = g(u)$ and N is a line of support for $U(g)$ at (u, w), then an analysis identical to that just completed for $(x, f(x))$ and M verifies the existence of N^+. If $w = g(u)$, but N is not a line of support for $U(g)$ at (u, w), then at least one of the half-lines of N emanating from (u, w) is strictly below any line of support for $U(g)$ at (u, w), and hence below the graph of g. This half-line serves as N^+. Finally, if (u, w) is not in the graph of g, then it is in the interior of $L(g)$. The convexity of $U(g)$ implies that at least one of the half-lines of N emanating from (u, w) is disjoint from $U(g)$, and this half-line fulfills the requirements on N^+.

Parameterize N^+ as $N^+ : t \geq 0 \to N^+(t) = (u, w) + tv$ for some nonzero vector v, and a half-line of M as $M^- : t \geq 0 \to M^-(t) = (x, f(x)) - tv$. Note that $p = N^+(t) + M^-(t)$. Let τ_3 be the minimum t such that $N^+(\tau_3)$ is in $V(g)$ or has a y-coordinate of 0 or $M^-(\tau_3)$ is in $V(f)$ or has a y-coordinate of 0. If $N^+(\tau_3)$ is in $V(g)$, then $M^-(\tau_3)$ is in $L(f)$, or if $M^-(\tau_3)$ is in $V(f)$, then $N^+(\tau_3)$ is in $L(g)$, as claimed in the lemma. If $N^+(\tau_3)$ has a y-coordinate of 0, then $M^-(\tau_3)$ is in $L(f)$, or if $M^-(\tau_3)$ has a y-coordinate of 0, then $N^+(\tau_3)$ is in $L(g)$, which are situations we have already considered.

The opposite inclusion is trivial. ■

The parabolic arcs used to define σ^2-parabolic STZones are simple convex functions. If a σ^2-parabolic STZone can be represented in the μ-σ^2 plane by a single defining parabolic arc, call it simple. Lemmas 5 and 6 are then applicable to the Minkowski sum of simple σ^2-parabolic STZones. Figure 12 partially illustrates this. Most often, σ^2-parabolic STZones are not simple. To treat Minkowski sums of σ^2-parabolic STZones in general, however, we need only a little more.

If for a set D in $\mathbf{R^2}$ there is a set of non-negative piecewise continuous functions, $\mathcal{F}$, such that $D = \cup_{f \in \mathcal{F}} L(f)$, say that D is generated by $\mathcal{F}$. If each of the functions in $\mathcal{F}$ is simple, define $V(\mathcal{F})$ to be $\cup_{f \in \mathcal{F}} V(f)$.

Corollary 1 *If F is generated by a finite set of simple convex functions, $\mathcal{F}$, and G is generated by a finite set of simple convex functions, $\mathcal{G}$, then*

$$F \oplus G = \{F \oplus V(\mathcal{G})\} \cup \{V(\mathcal{F}) \oplus G\}.$$

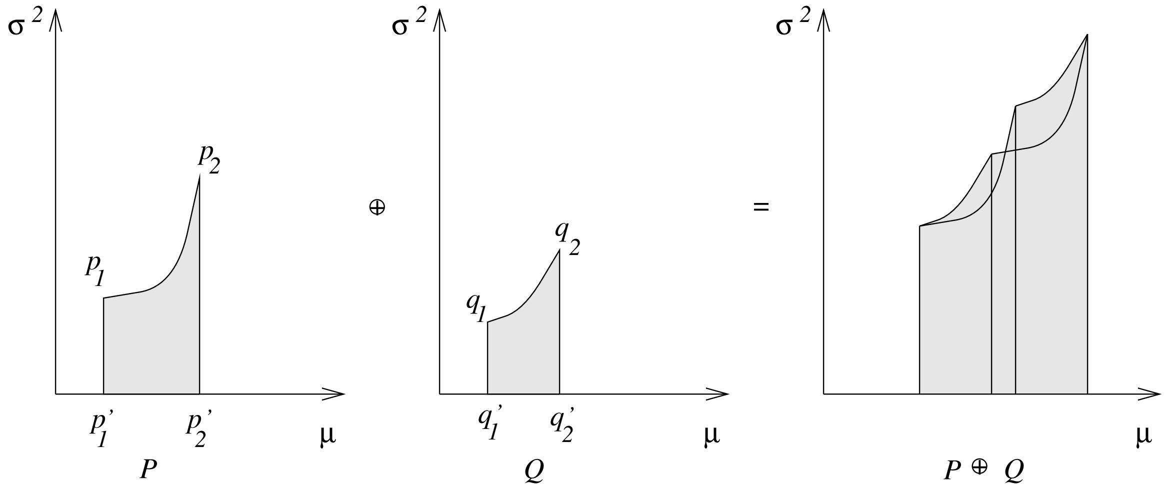


Figure 12: Minkowski sum of two simple σ^2-parabolic STZones P and Q. If $l(p, q)$ denotes the line-segment between points p and q, then Lemma 6 states that $P \oplus Q$ can be obtained as the union of $P \oplus l(q_1, q'_1)$, $P \oplus l(q_2, q'_2)$, $Q \oplus l(p_1, p'_1)$, and $Q \oplus l(p_2, p'_2)$. In some cases, one of the two vertices $p_1 + q_2$ and $p_2 + q_1$ may lie in the interior of the sum, as shown in Figure 2.

Proof. Since $\{A \cup B\} \oplus C = \{A \oplus C\} \cup \{B \oplus C\}$ for any subsets A, B, and C, the claim follows directly from Lemma 6. ■

Since $V(\mathcal{F})$ is a union of points and vertical line segments, and G is a union of regions below simple convex functions, $V(\mathcal{F}) \oplus G$ reduces to the union of the Minkowski sums of points or vertical line segments with regions below simple convex functions. Each of these Minkowski sums is easily obtained by translating a simple convex function to a new simple convex function and defining a new region below this new function. The union of this myriad of new regions that yields $F \oplus G$ is itself easy to obtain, at least conceptually, since it reduces to the region below the upper envelope of the graphs of the functions defining the regions in the union.

We close by considering the implications of this corollary on the Minkowski sum of σ^2-parabolic STZones. If we denote the set of the defining arcs of a σ^2-parabolic STZone by $\mathcal{F}$, then the STZone is generated by $\mathcal{F}$, a finite set of simple convex functions. Corollary 1 thus applies to the Minkowski sums of σ^2-parabolic STZones in the μ-σ^2 plane. The translations of the defining monotonic parabolic arcs drawn from non-negative parabolas yield arcs of the same type. The upper envelope is then defined piecewise again by arcs of the same type, determined by intersections of the arcs, yielding another σ^2-parabolic region.

References

The American Society of Mechanical Engineers, 1994, *ASME Y14.5M-1994, Dimensioning and Tolerancing*, New York.

Kaul, A., 1993, Computing Minkowski Sums, Ph.D. Thesis, Department of Mechanical Engineering, Columbia University, New York.

Matheron, G., 1975, *Random Sets and Integral Geometry*, John Wiley & Sons, New York.

Srinivasan, V., and O'Connor, M.A., 1994, "On Interpreting Statistical Tolerancing," *Manufacturing Review*, Vol. 7, No. 4, pp. 304-311.

Srinivasan, V., and O'Connor, M.A., 1995, "Towards an ISO Standard for Statistical Tolerancing," in *Proceedings of the 4th CIRP Seminar on Computer Aided Tolerancing*, University of Tokyo, Japan.

Valentine, F.A., 1964, *Convex Sets*, McGraw-Hill Book Co., New York.