CHAPTER III
METHODOLOGY

3.1 Identifying Different Regions According To The Presence of Undetected Gross Errors

Let's consider the case of two gross errors present in the system. Let θ_1, θ_2 be two random values of gross errors in measurements i1 and i2, respectively. Assume also that the MIMT strategy (i.e. serial elimination strategy) is used to detect gross errors. The maximum power MT test statistics for the two measurements i1 and i2 are given by:

$$Z_1^{MP} = \frac{\left|W_{i1i1}\theta_1 + W_{i1i2}\theta_2\right|}{\sqrt{W_{i1i1}}} \; ; \; Z_2^{MP} = \frac{\left|W_{i2i1}\theta_1 + W_{i2i2}\theta_2\right|}{\sqrt{W_{i2i2}}} \qquad (3.1)$$

where $\boldsymbol{W} = \boldsymbol{A}^T(\boldsymbol{ASA}^T)^{-1}\boldsymbol{A}$:variance-covariance matrix of $\boldsymbol{d} = \boldsymbol{S}^{-1}\boldsymbol{a}$. The threshold value for the MT test statistics is ξ (which is usually the value of 1.96).
Diagonal elements of matrix $\boldsymbol{W}$ (i.e. W_{i1i1}, W_{i2i2}) are positive. $\boldsymbol{W}$ is also a symmetric matrix (that is, $W_{i1i2} = W_{i2i1}$). If W_{i1i2}, W_{i2i1} are also positive, the different regions are identified as follows:

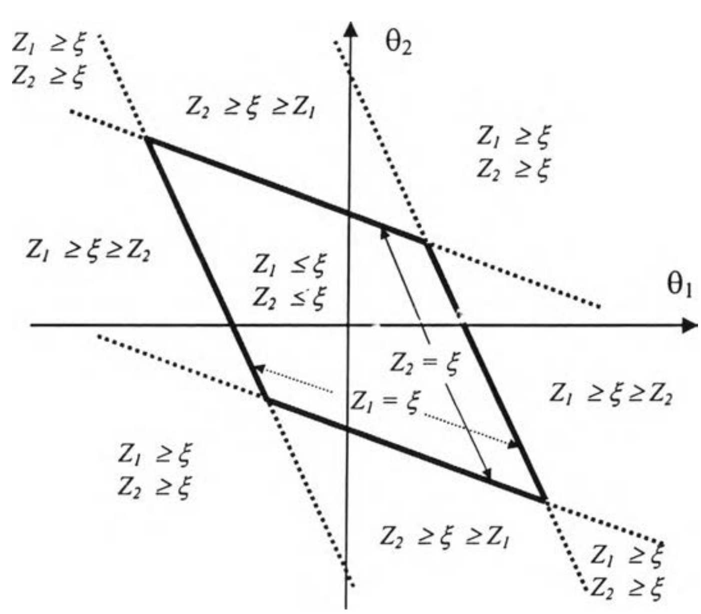

Figure 3.1.a Different regions when two gross errors are present in the system.

When Z_1, $Z_2 \leq \xi$: none of the two gross errors is detected because they are too small. In turn when Z_1, $Z_2 \geq \xi$ at least one of the two gross errors is detected. Which one is detected (or detected first) depends on which test statistic is greater, Z_1 *or* Z_2. We now identify the corresponding regions:

$Z_1 = Z_2$ when

$$\frac{W_{i1i1}\theta_1 + W_{i1i2}\theta_2}{\sqrt{W_{i1i1}}} = \frac{W_{i2i1}\theta_1 + W_{i2i2}\theta_2}{\sqrt{W_{i2i2}}} \quad \text{(in parts I, III of the plane)}$$

or $$\frac{(W_{i1i1}\theta_1 + W_{i1i2}\theta_2)}{\sqrt{W_{i1i1}}} = -\frac{(W_{i2i1}\theta_1 + W_{i2i2}\theta_2)}{\sqrt{W_{i2i2}}} \quad \text{(in parts II, IV of the plane)}$$

Then

$$\left(\sqrt{W_{i1i1}} - \frac{W_{i2i1}}{\sqrt{W_{i2i2}}}\right)\theta_1 = \left(\sqrt{W_{i2i2}} - \frac{W_{i1i2}}{\sqrt{W_{i1i1}}}\right)\theta_2 \quad \text{(Line 1)}$$

or $$\left(\sqrt{W_{i1i1}} + \frac{W_{i2i1}}{\sqrt{W_{i2i2}}}\right)\theta_1 = -\left(\sqrt{W_{i2i2}} + \frac{W_{i1i2}}{\sqrt{W_{i1i1}}}\right)\theta_2 \quad \text{(Line 2)}$$

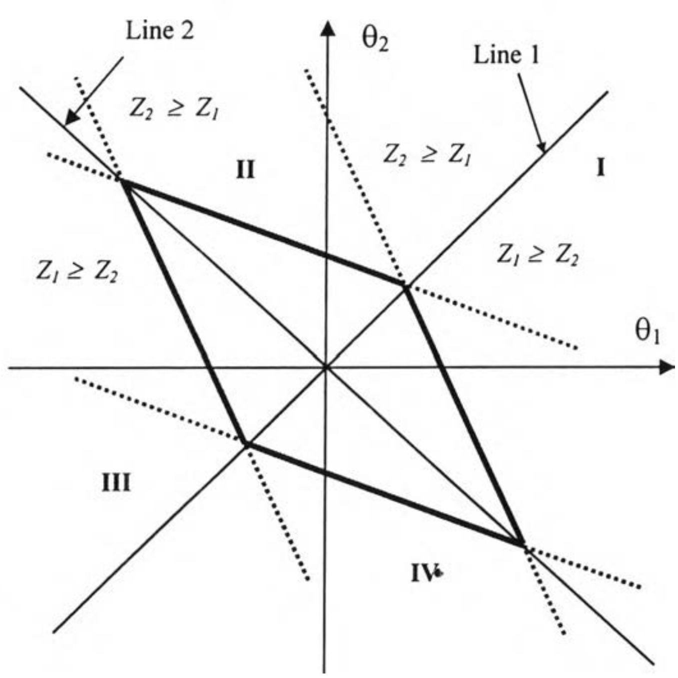

Figure 3.1.b Different regions when two gross errors are present in the system.

Let's consider the cases $Z_1 \geq Z_2 \geq \xi$ and $Z_1 \geq \xi \geq Z_2$, in these cases gross error θ_1 is detected first and the corresponding measurement i1 is eliminated. Then at the next stage of the serial elimination strategy, the test statistic for the measurement i2 is:

$$Z_2' = \frac{|W'_{i2i2}\theta_2|}{\sqrt{W'_{i2i2}}} \tag{3.2}$$

Then gross error θ_2 is also detected if

$$Z_2' = \frac{|W'_{i2i2}\theta_2|}{\sqrt{W'_{i2i2}}} \geq \xi \text{ or } |\theta_2| \geq \frac{\xi}{\sqrt{W'_{i2i2}}} = \delta_2 \tag{3.3}$$

Where W' is the updated matrix W after measurement i1 has been eliminated.

Consequently gross error θ_2 is undetected (i.e. only gross error θ_1 is detected) if $Z_2' < \xi \ or \ |\theta_2| < \delta_2$.

Note that we have

$$\delta_2 = \frac{\xi}{\sqrt{W'_{i2i2}}} \geq \frac{\xi}{\sqrt{W_{i2i2}}} \tag{3.4}$$

This expression stems from the fact that $W_{i2i2} \geq W'_{i2i2}$ (see appendix)

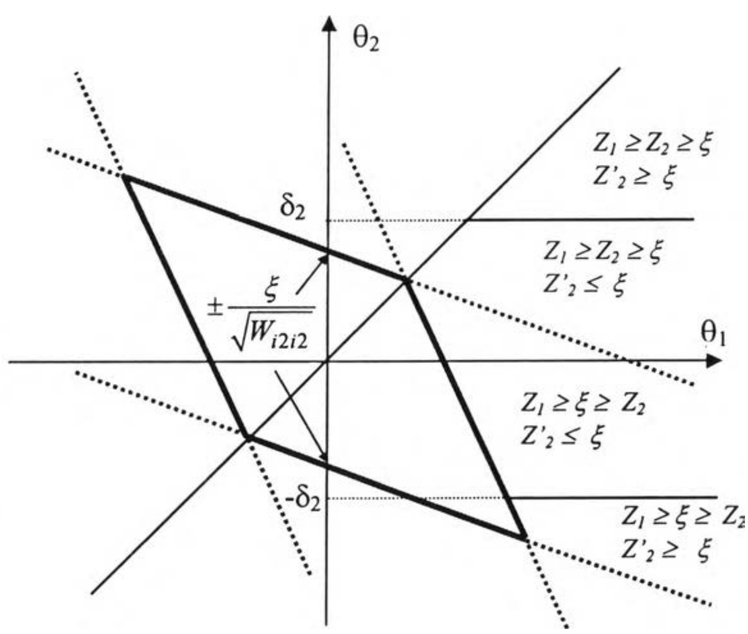

Figure 3.1.c Different regions when two gross errors are present in the system.

Performing the same analysis for other cases, we obtain a typical figure with different regions as follows:

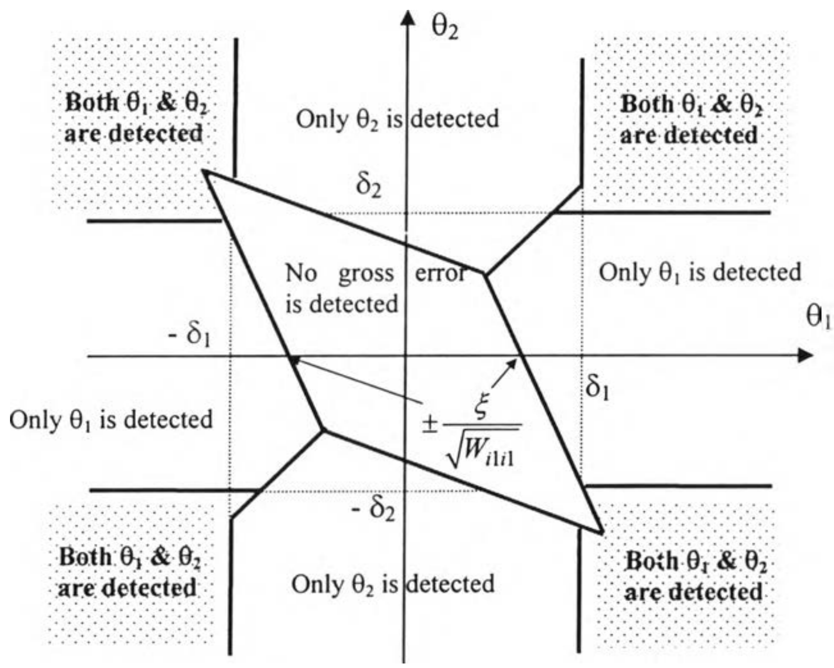

Figure 3.1.d Different regions when two gross errors are present in the system.

where the value δ_1 is given by:

$$Z_1' = \frac{|W''_{i1i1}\,\theta_1|}{\sqrt{W''_{i1i1}}} \geq \xi \;\Rightarrow\; |\theta_1| \geq \frac{\xi}{\sqrt{W''_{i1i1}}} = \delta_1 \tag{3.5}$$

W" is the updated matrix *W* after measurement i2 is eliminated; Z'_1 is the test statistic for measurement i1 after gross error θ_2 has been eliminated by eliminating the corresponding measurement i2.

Similarly, we also have:

$$\delta_1 = \frac{\xi}{\sqrt{W'_{i1i1}}} \geq \frac{\xi}{\sqrt{W_{i1i1}}} \tag{3.6}$$

If W_{i1i2} and W_{i2i1} are negative, we can identify different regions as follows:

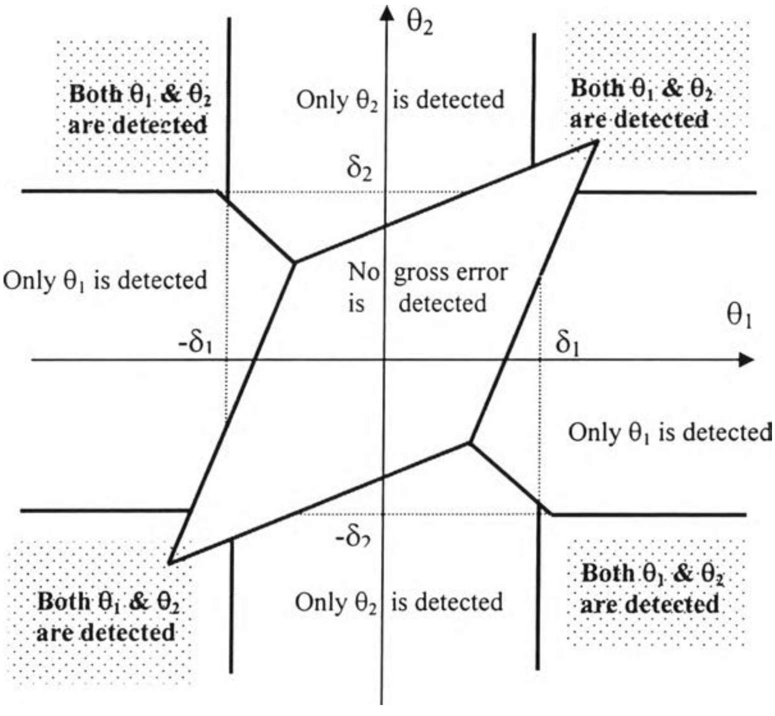

Figure 3.1.e Different regions when two gross errors are present in the system.

Clearly, in case two gross errors are present, there are 4 (= 2^2) regions as shown above. The shape of these regions changes with the values of Wij. Then, in case three gross errors are present, there are 8 (=2^3) regions in the three dimensional graph corresponding to 8 possibilities that can happen and so on. In each region, only one possibility can occur and the corresponding integrand function is used in that region.

Having identified the difference regions, we propose two methods to evaluate the integral expressions as follows. Two approaches for calculating the probability (P) and the financial loss (DEFL) are: (i) letting the integrand function change accordingly to different regions in the whole sample space of variables (in approximation method) and (ii): considering them as sum of separate integrals (in Monte Carlo method).

3.2 Methods For Calculating Integral Expressions For The Financial Loss And The Probability

Under simplified assumptions (negligible process variation and normal distributions), the general expressions given by Bagajewicz (2004b) for the probability and the financial loss in the presence of two gross errors reduce to the following expressions:

The probability P is given by:

$$P\left(\hat{m}_p \geq m_p^* \middle| i1,i2\right)$$
$$= \frac{\Phi_{i1,i2}^2}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\begin{array}{l}\frac{1}{2} \quad \text{(when both gross errors are detected)} \\ +\left\{\begin{array}{ll}\frac{1}{2}erfc\left\{-\frac{\hat{\delta}_p^{i1,i2}(\theta_1,\theta_2)}{\sqrt{2}\hat{\sigma}_p}\right\} & \hat{\delta}_p^{i1,i2}(\theta_1,\theta_2)\leq 0 \\ 1-\frac{1}{2}erfc\left\{\frac{\hat{\delta}_p^{i1,i2}(\theta_1,\theta_2)}{\sqrt{2}\hat{\sigma}_p}\right\} & \hat{\delta}_p^{i1,i2}(\theta_1,\theta_2)>0\end{array}\right\} \\ \text{(when no gross errors is detected)} \\ +\left\{\begin{array}{ll}\frac{1}{2}erfc\left\{-\frac{\hat{\delta}_p^{i2}(\theta_2)}{\sqrt{2}\hat{\sigma}_p}\right\} & \hat{\delta}_p^{i2}(\theta_2)\leq 0 \\ 1-\frac{1}{2}erfc\left\{\frac{\hat{\delta}_p^{i2}(\theta_2)}{\sqrt{2}\hat{\sigma}_p}\right\} & \hat{\delta}_p^{i2}(\theta_2)>0\end{array}\right\} \\ \text{(when only gross errors 1 is detected)} \\ +\left\{\begin{array}{ll}\frac{1}{2}erfc\left\{-\frac{\hat{\delta}_p^{i1}(\theta_1)}{\sqrt{2}\hat{\sigma}_p}\right\} & \hat{\delta}_p^{i1}(\theta_1)\leq 0 \\ 1-\frac{1}{2}erfc\left\{\frac{\hat{\delta}_p^{i1}(\theta_1)}{\sqrt{2}\hat{\sigma}_p}\right\} & \hat{\delta}_p^{i1}(\theta_1)>0\end{array}\right\} \\ \text{(when only gross errors 2 is detected)}\end{array}\right]\frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 \qquad (3.7)$$

where $\hat{\delta}_p^{i1,i2}(\theta_1,\theta_2)$, $\hat{\delta}_p^{i1}(\theta_1)$, $\hat{\delta}_p^{i2}(\theta_2)$ are the induced biases due to undetected gross errors. From (2.41), we see that we can express induced bias as a linear function of undetected gross errors: $\hat{\delta}_p^{i1,i2}(\theta_1,\theta_2) = \alpha_1\theta_1+\alpha_2\theta_2$, $\hat{\delta}_p^{i1}(\theta_1)=\alpha_1\theta_1$, $\hat{\delta}_p^{i2}(\theta_2)=\alpha_2\theta_2$. Then, the above expression can be rewritten as:

$P\left(\hat{m}_p \geq m_p^* \middle| i1, i2\right) = \frac{\Phi_{i1,i2}^2}{2}\frac{1}{2}P$, where:

$$P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\begin{array}{l} \dfrac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}(=F_{12}) \\ \text{(when both gross errors are detected)} \\ +\left\{\begin{array}{ll} [1-erf\left\{-\dfrac{\alpha_1\theta_1+\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}\right\}] & \alpha_1\theta_1+\alpha_2\theta_2 \leq 0 \\ [1+erf\left\{\dfrac{\alpha_1\theta_1+\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}\right\}] & \alpha_1\theta_1+\alpha_2\theta_2 > 0 \end{array}\right\} \dfrac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}(=F_0) \\ \text{(when no gross error is detected)} \\ +\left\{\begin{array}{ll} [1-erf\left\{-\dfrac{\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}\right\}] & \alpha_2\theta_2 \leq 0 \\ [1+erf\left\{\dfrac{\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}\right\}] & \alpha_2\theta_2 > 0 \end{array}\right\} \dfrac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}(=F_1) \\ \text{(when only gross error 1 is detected)} \\ +\left\{\begin{array}{ll} [1-erf\left\{-\dfrac{\alpha_1\theta_1}{\sqrt{2}\hat{\sigma}_p}\right\}] & \alpha_1\theta_1 \leq 0 \\ [1+erf\left\{\dfrac{\alpha_1\theta_1}{\sqrt{2}\hat{\sigma}_p}\right\}] & \alpha_1\theta_1 > 0 \end{array}\right\} \dfrac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}(=F_2) \\ \text{(when only gross error 2 is detected)} \end{array}\right\} d\theta_1 d\theta_2 \qquad (3.8)$$

$$P = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\theta_1, \theta_2) d\theta_1 d\theta_2$$

The financial loss DEFL is given by:

$DEFL^2 \,|\, i1, i2$

$$= \frac{1}{2} K_i T \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left\{ \begin{array}{l} \dfrac{\hat{\sigma}_p}{\sqrt{2\pi}} \\ \text{(when both gross errors are detected)} \\ + \left[\dfrac{\hat{\sigma}_p e^{-(A_1)^2}}{\sqrt{2\pi}} - \dfrac{\sqrt{2}\hat{\sigma}_p A_1}{2} \begin{cases} [1-erf(A_1)] & if\ A_1 > 0 \\ [1+erf(-A_1)] & if\ A_1 \le 0 \end{cases} \right] \\ \text{(when no gross error is detected)} \\ + \left[\dfrac{\hat{\sigma}_p e^{-(A_2)^2}}{\sqrt{2\pi}} - \dfrac{\sqrt{2}\hat{\sigma}_p A_2}{2} \begin{cases} [1-erf(A_2)] & if\ A_2 > 0 \\ [1+erf(-A_2)] & if\ A_2 \le 0 \end{cases} \right] \\ \text{(when only gross error 1 is detected)} \\ + \left[\dfrac{\hat{\sigma}_p e^{-(A_3)^2}}{\sqrt{2\pi}} - \dfrac{\sqrt{2}\hat{\sigma}_p A_3}{2} \begin{cases} [1-erf(A_3)] & if\ A_3 > 0 \\ [1+erf(-A_3)] & if\ A_3 \le 0 \end{cases} \right] \\ \text{(when only gross error 2 is detected)} \end{array} \right\} \frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}} \frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}} d\theta_1 d\vartheta_2 \qquad (3.9)$$

where $A_1 = \dfrac{\alpha_1\theta_1 + \alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$, $A_2 = \dfrac{\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$ and $A_3 = \dfrac{\alpha_1\theta_1}{\sqrt{2}\hat{\sigma}_p}$.

Clearly, we see that these integral expressions require integrating a discontinuous function that changes form in different regions.

3.2.1 Approximation Method

To calculate P and DEFL, one needs to evaluate integrals at the form

$$J = \int_a^b e^{-z^2}\, erf\{z\}\, dz \qquad (3.10)$$

The essence of the approximation method is to replace one of the integrand functions by a piece wise linear function. Thus if the error function is replaced by a linear expression, J becomes

$$J^* = \sum \int_{a_i}^{b_i} (y_{1i} + y_{2i} z) e^{-z^2} dz \qquad (3.11)$$

which can be calculated analytically. Certain choices of piece wise linear functions guarantee that the answer is an underestimate J^*_L or an overestimate J^*_U of J (Figure 3.2). Thus, the integral J can be approximated as the average of underestimate J^*_L and overestimate J^*_U, that is:

$$J = (J_L^* + J_U^*)/2 \qquad (3.12)$$

Because J_U^* and J_L^* are the overestimator and the underestimator of J, then, the maximum error is given by $(J_U^* - J_L^*)$.

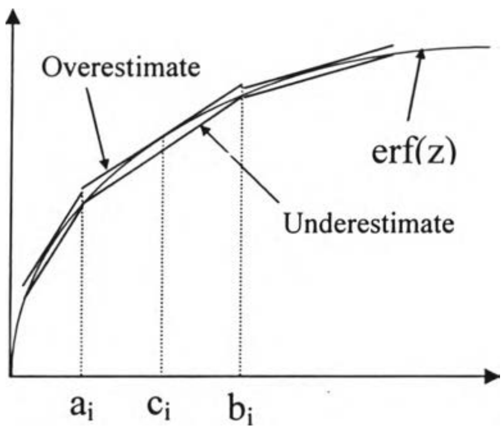

Figure 3.2 Approximation method.

To deal with discontinuity of the function, we partition sample space of variables into regions as follows:

$$
\begin{aligned}
P &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(\theta_1,\theta_2)d\theta_1 d\theta_2 \\
&= \left[\left\{\int_{-\infty}^{-K_1} + \int_{K_1}^{\infty}\right\} + \int_{-K_1}^{K_1}\right]\left[\left\{\int_{-\infty}^{-K_2} + \int_{K_2}^{\infty}\right\} + \int_{-K_2}^{K_2}\right] F(\theta_1,\theta_2)d\theta_1 d\theta_2 \\
&= \left\{\int_{-\infty}^{-K_1} + \int_{K_1}^{\infty}\right\}\left\{\int_{-\infty}^{-K_2} + \int_{K_2}^{\infty}\right\} F(\theta_1,\theta_2)d\theta_1 d\theta_2 \quad (P_1) \\
&+ \left\{\int_{-\infty}^{-K_1} + \int_{K_1}^{\infty}\right\}\int_{-K_2}^{K_2} F(\theta_1,\theta_2)d\theta_1 d\theta_2 \quad (P_2) \\
&+ \int_{-K_1}^{K_1}\left\{\int_{-\infty}^{-K_2} + \int_{K_2}^{\infty}\right\} F(\theta_1,\theta_2)d\theta_1 d\theta_2 \quad (P_3) \\
&+ \int_{-K_1}^{K_1}\int_{-K_2}^{K_2} F(\theta_1,\theta_2)d\theta_1 d\theta_2 \quad (P_4)
\end{aligned}
\tag{3.13}
$$

The regions are depicted as follows:

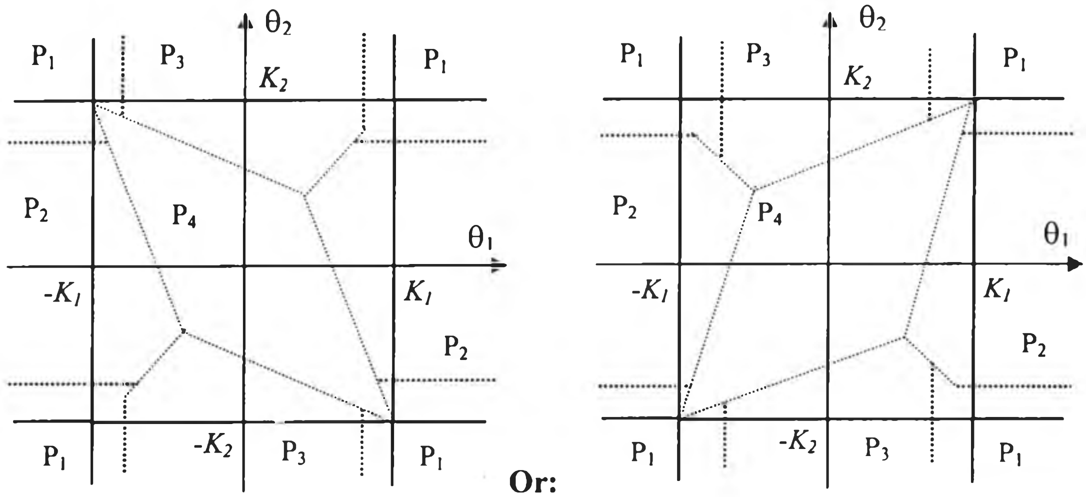

Figure 3.3 Regions used in calculation.

In the first region (P_1) both gross errors are detected ($F(\theta_1,\theta_2) = F_{12}$) and the first term ($P_1$) has analytical form. In the second region (P_2) gross error θ_1 is surely detected and there are two possibilities: one is that gross error θ_2 is also detected ($F(\theta_1,\theta_2) = F_{12}$) when $Z_2' = \frac{|W'_{i2i2}\theta_2|}{\sqrt{W'_{i2i2}}} \geq \xi \Rightarrow |\theta_2| \geq \delta_2$. The other is that gross error θ_2 is undetected ($F(\theta_1,\theta_2) = F_1$) when $Z_2' = \frac{|W'_{i2i2}\theta_2|}{\sqrt{W'_{i2i2}}} \leq \xi \Rightarrow |\theta_2| \leq \delta_2$

P_2 is partitioned into two terms as follows:

$$\begin{aligned}
P_2 &= \left\{ \int_{-\infty}^{-K_1} + \int_{K_1}^{\infty} \right\} \int_{-K_2}^{K_2} F(\theta_1,\theta_2) d\theta_1 d\theta_2 \\
&= \left\{ \int_{-\infty}^{-K_1} + \int_{K_1}^{\infty} \right\} \left[\left\{ \int_{-K_2}^{-\delta_2} + \int_{\delta_2}^{K_2} \right\} + \int_{-\delta_2}^{\delta_2} \right] F(\theta_1,\theta_2) d\theta_1 d\theta_2 \\
&= \left\{ \int_{-\infty}^{-K_1} + \int_{K_1}^{\infty} \right\} \left\{ \int_{-K_2}^{-\delta_2} + \int_{\delta_2}^{K_2} \right\} F_{12} d\theta_1 d\theta_2 (P_{2a}) \\
&+ \left\{ \int_{-\infty}^{-K_1} + \int_{K_1}^{\infty} \right\} \int_{-\delta_2}^{\delta_2} F_1 d\theta_1 d\theta_2 (P_{2b})
\end{aligned} \tag{3.14}$$

P_{2a} can be calculated analytically and P_{2b} needs to be calculated approximately.

In the third region (P_3), gross error θ_2 is surely detected and there are two possibilities: one is that gross error θ_1 is also detected ($F(\theta_1,\theta_2) = F_{12}$) when

$Z_1' = \frac{|W_{i1i1}'\theta_1|}{\sqrt{W_{i1i1}'}} \geq \xi \Rightarrow |\theta_1| \geq \delta_1$. The other is that gross error θ_1 is undetected ($F(\theta_1,\theta_2) = F_2$) when $Z_1' = \frac{|W_{i1i1}'\theta_1|}{\sqrt{W_{i1i1}'}} \leq \xi \Rightarrow |\theta_1| \leq \delta_1$

Similarly to P_2, P_3 is partitioned into two terms as follows:

$$\begin{aligned} P_3 &= \int_{K_1}^{K_1} \left\{ \int_{-\infty}^{-K_2} + \int_{K_2}^{\infty} \right\} F(\theta_1,\theta_2) d\theta_1 d\theta_2 \\ &= \left[\left\{ \int_{K_1}^{-\delta_1} + \int_{\delta_1}^{K_1} \right\} + \int_{-\delta_1}^{\delta_1} \right] \left\{ \int_{-\infty}^{-K_2} + \int_{K_2}^{\infty} \right\} F(\theta_1,\theta_2) d\theta_1 d\theta_2 \\ &= \left\{ \int_{K_1}^{-\delta_1} + \int_{\delta_1}^{K_1} \right\} \left\{ \int_{-\infty}^{-K_2} + \int_{K_2}^{\infty} \right\} F_{12} d\theta_1 d\theta_2 (P_{3a}) \\ &+ \int_{\delta_1}^{\delta_1} \left\{ \int_{-\infty}^{-K_2} + \int_{K_2}^{\infty} \right\} F_2 d\theta_1 d\theta_2 (P_{3b}) \end{aligned} \tag{3.15}$$

The term P_{3a} can be calculated analytically; the term P_{3b} needs to be calculated approximately.

The remaining term:

$$P_4 = \int_{K_1}^{K_1} \int_{K_2}^{K_2} F(\theta_1,\theta_2) d\theta_1 d\theta_2$$

All four possibilities can happen in this region. This term needs to be calculated approximately.

In general, to calculate the terms (e.g. P_4) approximately, the following steps are performed:

- Divide sample space of variables into subregions.
- In each subregion, check what possibility is (i.e., check for the presence & the number of undetected gross errors), then use the corresponding integrand function in that subregion.
- Determine the sign of the induced bias $\hat{\delta}_p^{i1,i2}(\theta_1,\theta_2) = \alpha_1\theta_1 + \alpha_2\theta_2, \hat{\delta}_p^{i1}(\theta_1) = \alpha_1\theta_1, \ldots$, to determine which form of the error function is in that subregion.
- Calculate the corresponding term in that subregion using approximation method.
- Calculate, for example P_4, as summation of corresponding terms in the subregions.

$$P_4 = \int_{K_1}^{K_1} \int_{K_2}^{K_2} F(\theta_1,\theta_2)d\theta_1 d\theta_2$$
$$= \sum_{a_i,b_i} \sum_{a_{i2},b_{i2}} \int_{a_i}^{b_i} \int_{a_{i2}}^{b_{i2}} F(\theta_1,\theta_2)d\theta_1 d\theta_2 \quad (3.16)$$

The financial loss DEFL can be calculated at the same way, the difference is the integrand function.

The procedure to calculate integral expression as described above can be extended to calculate P & DEFL when more than two gross errors are present in the system. The principle is to partition sample space of variables into regions at much the same way. There are regions that we know for sure information about the presence of undetected gross errors (like P_1, P_2, P_3) and the corresponding integral terms in these regions can be calculated analytically or approximately; but there are also regions where many possibilities can happen (like P_4) and the calculation requires us to perform the five steps as described above.

3.2.2 Monte Carlo Method

The principle of Monte Carlo method is briefly described as follows: Suppose we need to evaluate the following integral:

$$\mu = \int \dots \int_{\Omega_X} g(\boldsymbol{x}) f_x(\boldsymbol{x}) d\boldsymbol{x} \quad (3.17)$$

in which $g(x)$ is a function of $\boldsymbol{x}$, a vector of random variables described by a joint probability density function (pdf) $f_x(x)$; Ω_X is the space of $\boldsymbol{x}$.

If $g(x)\, f_x(x)$ is bounded on Ω_X and Ω_X is a bounded subset of R^k, the above integral can be replaced by the following integral:

$$\mu = \int \dots \int_{R^k} I_{\Omega_X}(\boldsymbol{x}) g(\boldsymbol{x}) f_x(\boldsymbol{x}) d\boldsymbol{x} \quad (3.18)$$

where:

$$I_{\Omega_X}(\boldsymbol{x}) = \begin{cases} 1 & if \quad \boldsymbol{x} \in \Omega_X \\ 0 & if \quad \boldsymbol{x} \notin \Omega_X \end{cases} \quad (3.19)$$

$I_{\Omega_X}(\boldsymbol{x})$ is the indicator function for the set Ω_X (Evans and Swartz, 2000).

To estimate the integral μ given by (3.17), we sample randomly a sequence of $\boldsymbol{x}_i$, i=1,2,...,N, from the density function $f_x(\boldsymbol{x})$ and the space R^k and compute the sample mean:

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} I_{\Omega_X}(x)_i\, g(x)_i \tag{3.20}$$

for computationally reasonable values of N. Since x_i , i=1,2,...,k are independent identically distributed random variables, it can be shown that E[$\hat{\mu}$] = μ, i.e., $\hat{\mu}$ is an unbiased estimator of μ (Lu & Zhang, 2003).

Monte Carlo method is applied to calculate the probability P& the financial loss DEFL. In this approach, P & DEFL are considered as summation of separate integral terms. The Monte Carlo method is used to calculate these integral terms and then P & DEFL can be calculated as summation of them. Therefore, it can be written as shown below:

(with R^2 = $[-\infty, \infty]x[-\infty, \infty]$)

$$P\left(\hat{m}_p \geq m_p^* \middle| i1, i2\right) = \frac{\Phi^2_{i1,i2}}{2}\frac{1}{2}P, \text{ where:}$$

$$P = \begin{cases} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_1(\theta_1,\theta_2)\dfrac{e^{-\frac{(\theta_1-\hat{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\hat{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 \quad (= PL_1) \\ \text{(when both gross errors are detected)} \\ +\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_2(\theta_1,\theta_2)\begin{Bmatrix}[1-erf\{-A_1\}] & \text{if } A_1 \leq 0 \\ [1+erf\{A_1\}] & \text{if } A_1 > 0\end{Bmatrix}\dfrac{e^{-\frac{(\theta_1-\hat{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\hat{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 (= PL_2) \\ \text{(when no gross error is detected)} \\ +\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_3(\theta_1,\theta_2)\begin{Bmatrix}[1-erf\{-A_2\}] & \text{if } A_2 \leq 0 \\ [1+erf\{A_2\}] & \text{if } A_2 > 0\end{Bmatrix}\dfrac{e^{-\frac{(\theta_1-\hat{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\hat{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 (= PL_3) \\ \text{(when only gross error 1 is detected)} \\ +\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I_4(\theta_1,\theta_2)\begin{Bmatrix}[1-erf\{-A_3\}] & \text{if } A_3 \leq 0 \\ [1+erf\{A_3\}] & \text{if } A_3 > 0\end{Bmatrix}\dfrac{e^{-\frac{(\theta_1-\hat{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\dfrac{e^{-\frac{(\theta_2-\hat{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 (= PL_4) \\ \text{(when only gross error 2 is detected)} \end{cases} \tag{3.21}$$

$$P = PL_1 + PL_2 + PL_3 + PL_4$$

where I_1, I_2, I_3 and I_4 are indicator functions for the corresponding integrals;

$A_1 = \dfrac{\alpha_1\theta_1 + \alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$, $A_2 = \dfrac{\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$ and $A_3 = \dfrac{\alpha_1\theta_1}{\sqrt{2}\hat{\sigma}_p}$.

$$(\mathrm{DEFL}^2|\mathrm{i1,i2})\frac{2}{K_sT} =$$

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{\hat{\sigma}_p}{\sqrt{2\pi}}I_1(\theta_1,\theta_2)\frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 \text{ (when both gross errors are detected)} \qquad (= FL_1)$$

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I_2(\theta_1,\theta_2)\left[\frac{\hat{\sigma}_p e^{-(A_1)^2}}{\sqrt{2\pi}}-\frac{\sqrt{2}\hat{\sigma}_p A_1}{2}\begin{cases}[1-erf(A_1)] & if\ A_1>0\\ [1+erf(-A_1)] & if\ A_1\le 0\end{cases}\right]\frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 \quad (= FL_2)$$

(when no gross error is detected) (3.22)

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I_3(\theta_1,\theta_2)\left[\frac{\hat{\sigma}_p e^{-(A_2)^2}}{\sqrt{2\pi}}-\frac{\sqrt{2}\hat{\sigma}_p A_2}{2}\begin{cases}[1-erf(A_2)] & if\ A_2>0\\ [1+erf(-A_2)] & if\ A_2\le 0\end{cases}\right]\frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 \quad (= FL_3)$$

(when only gross error 1 is detected)

$$+\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}I_4(\theta_1,\theta_2)\left[\frac{\hat{\sigma}_p e^{-(A_3)^2}}{\sqrt{2\pi}}-\frac{\sqrt{2}\hat{\sigma}_p A_3}{2}\begin{cases}[1-erf(A_3)] & if\ A_3>0\\ [1+erf(-A_3)] & if\ A_3\le 0\end{cases}\right]\frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}d\theta_1 d\theta_2 (= FL_4)$$

(when only gross error 2 is detected)

$$(DEFL^2|i1,i2)\frac{2}{K_sT} = FL_1 + FL_2 + FL_3 + FL_4$$

where I_1, I_2, I_3 and I_4 are indicator functions for the corresponding integrals;

$A_1=\frac{\alpha_1\theta_1+\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$, $A_2=\frac{\alpha_2\theta_2}{\sqrt{2}\hat{\sigma}_p}$ and $A_3=\frac{\alpha_1\theta_1}{\sqrt{2}\hat{\sigma}_p}$.

In these specific cases, we have:

$$f_{\underline{x}}(\underline{x}) = f(\theta_1,\theta_2) = \frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}} \qquad (3.23)$$

The procedure to compute the P & DEFL by Monte Carlo method comprises of N trials, each trial comprises of 2 steps:

(i): generate random numbers θ_1 and θ_2 according to the probability distribution function $f(\theta_1,\theta_2)$ and the space $R^2 = [-\infty,\infty]x[-\infty,\infty]$:

$$f_{\underline{x}}(\underline{x}) \equiv f(\theta_1,\theta_2) = \frac{e^{-\frac{(\theta_1-\bar{\delta}_1)^2}{2\rho_1^2}}}{\rho_1\sqrt{2\pi}}\frac{e^{-\frac{(\theta_2-\bar{\delta}_2)^2}{2\rho_2^2}}}{\rho_2\sqrt{2\pi}}$$

Because distributions of gross errors are uncorrelated, this task is equivalent to generating θ_i according to the normal distribution $N(\bar{\delta}_i,\rho_i)$.

(ii): with these values of θ_1 and θ_2, check for the presence and the number of undetected gross errors (by using the MIMT test) and calculate the values $I_{\Omega_X}(\underline{x})g(\underline{x})_i$ corresponding to the integral terms PL_i and FL_i as follows:

$$I_{\Omega_X}(\underline{x})g(\underline{x})_1 = \begin{cases} 1 \text{ (for } PL_1\text{) or } \dfrac{\hat{\sigma}_p}{\sqrt{2\pi}} \text{ (for } FL_1\text{) if both gross errors are detected} \\ 0 \text{ other cases} \end{cases} \tag{3.24}$$

$$I_{\Omega_X}(\underline{x})g(\underline{x})_2 = \begin{cases} \left.\begin{cases} [1 - erf\{-A_1\}] & if\ A_1 \le 0 \\ [1 + erf\{A_1\}] & if\ A_1 > 0 \end{cases}\right\} \text{ if no bias is detected} \\ 0 \text{ other cases} \end{cases} \quad \text{(for } PL_2\text{)} \tag{3.25}$$

$$\text{Or } = \begin{cases} \dfrac{\hat{\sigma}_p e^{-(A_1)^2}}{\sqrt{2\pi}} - \dfrac{\sqrt{2}\hat{\sigma}_p A_1}{2}\begin{cases} [1 - erf(A_1)] & if\ A_1 > 0 \\ [1 + erf(-A_1)] & if\ A_1 \le 0 \end{cases} \text{ if no bias is detected} \\ 0 \text{ other cases} \end{cases} \quad \text{(for } FL_2\text{)}$$

$$I_{\Omega_X}(\underline{x})g(\underline{x})_3 = \begin{cases} \left.\begin{cases} [1 - erf\{-A_2\}] & if\ A_2 \le 0 \\ [1 + erf\{A_2\}] & if\ A_2 > 0 \end{cases}\right\} \text{ if only bias 1 is detected} \\ 0 \text{ other cases} \end{cases} \quad \text{(for } PL_3\text{)} \tag{3.26}$$

$$\text{Or } = \begin{cases} \dfrac{\hat{\sigma}_p e^{-(A_2)^2}}{\sqrt{2\pi}} - \dfrac{\sqrt{2}\hat{\sigma}_p A_2}{2}\begin{cases} [1 - erf(A_2)] & if\ A_2 > 0 \\ [1 + erf(-A_2)] & if\ A_2 \le 0 \end{cases} \text{ if only bias 1 is detected} \\ 0 \text{ other cases} \end{cases} \quad \text{(for } FL_3\text{)}$$

$$I_{\Omega_X}(\underline{x})g(\underline{x})_4 = \begin{cases} \left.\begin{cases} [1 - erf\{-A_3\}] & if\ A_3 \le 0 \\ [1 + erf\{A_3\}] & if\ A_3 > 0 \end{cases}\right\} \text{ if only bias 2 is detected} \\ 0 \text{ other cases} \end{cases} \quad \text{(for } PL_4\text{)} \tag{3.27}$$

$$\text{Or } = \begin{cases} \dfrac{\hat{\sigma}_p e^{-(A_3)^2}}{\sqrt{2\pi}} - \dfrac{\sqrt{2}\hat{\sigma}_p A_3}{2}\begin{cases} [1 - erf(A_3)] & if\ A_3 > 0 \\ [1 + erf(-A_3)] & if\ A_3 \le 0 \end{cases} \text{ if only bias 2 is detected} \\ 0 \text{ other cases} \end{cases} \quad \text{(for } FL_4\text{)}$$

where A_1, A_2, A_3 are shown above.

The estimators for PL_i or FL_i can be calculated from equation (3.20). Then P & DEFL can be calculated as sum of integral terms PL_i & FL_i (Eq. 3.21 & 3.22).