

Chapter 6. The Treatment of Distance Dependent Scattering

6.1 Distance Dependent Scattering

In the prototype x-ray scanning system, scattering images are obtained to provide information on material density. As described in the Section 3.2, the forward and backward scattering is related to solid angle. Scattering is therefore dependent on the distance of the scanned object from the x-ray source. Figure 6.1-1 shows an example of backward scattering, where a polyethylene plate was scanned with different distances from x-ray source. The object is far away from x-ray source in Figure 6.1-1a, and backward scattering is low. In Figure 6.1-1b, the object was closer to x-ray source, and backward scattering therefore is stronger than the previous case. A similar phenomenon occurs on forward scattering.

In the real world, an object may be placed anywhere on the conveyer belt, so the measured intensity will contain errors relative to the ideal intensity. This makes classification results less reliable. Typical results for a 1.25 cm polyethylene plate is given in Figure 6.1-2. It can be seen that the forward scattering $F_L(d)$, where d is the distance from x-ray source with a reference to tunnel on the x-ray source side, becomes stronger and back scattering, $B_L(d)$, becomes weaker when the object is farther away from the x-ray source.

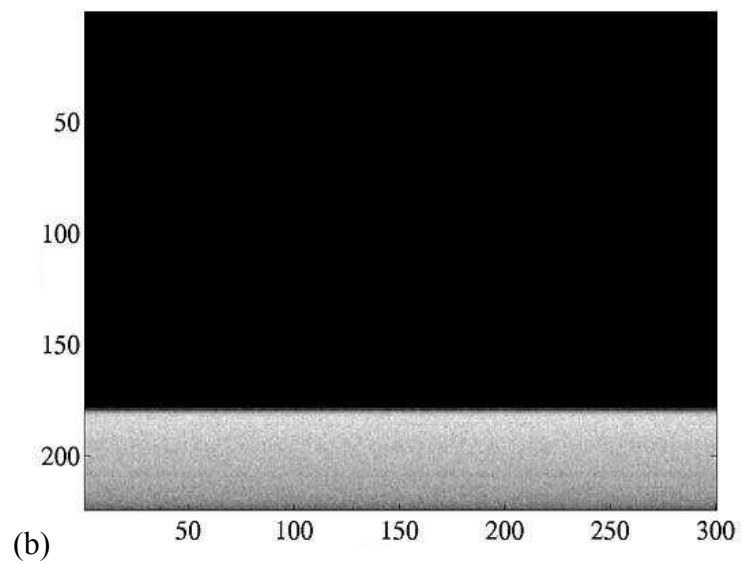
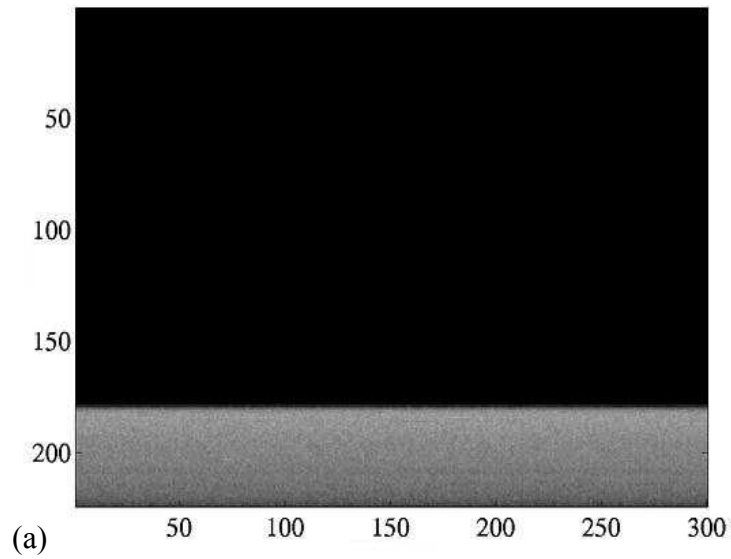


Figure 6.1-1 Backward scattering of a polyethylene plate: (a) far away from x-ray source, (b) close to x-ray source.

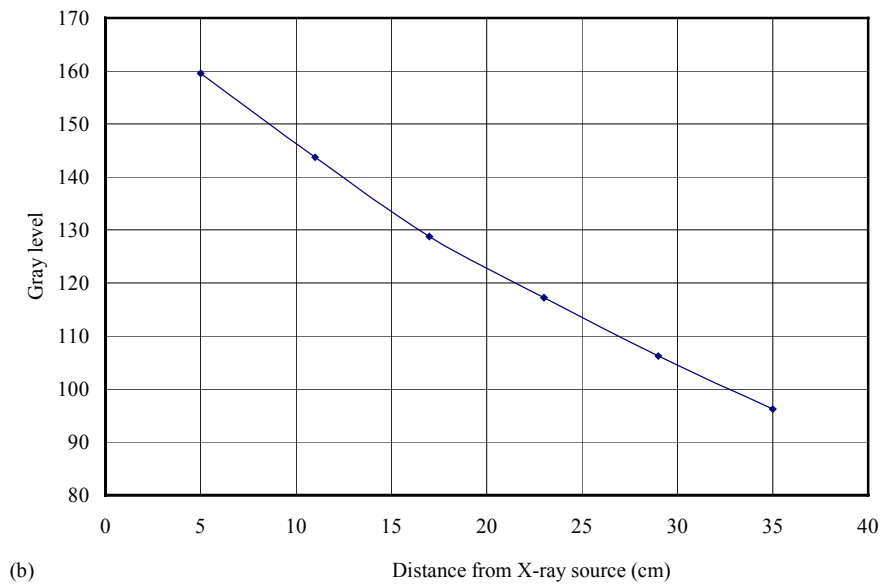
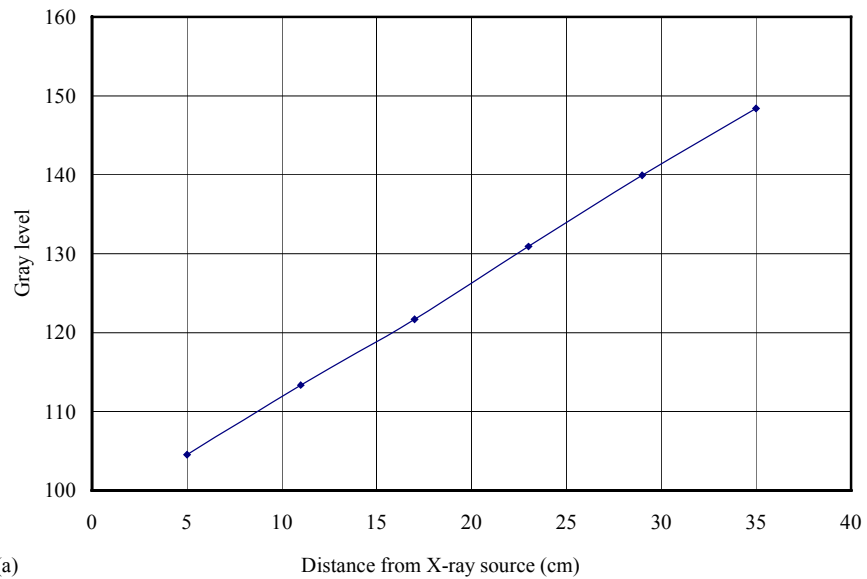


Figure 6.1-2 Variation of scattering gray levels with distance: (a) forward scattering, (b) backward scattering.

Note that one possibility is to create a constant signal value by combining the two scatter signals, such as the average value $(F_L(d) + B_L(d))/2$. The variation after such a processing is narrowed to about 10 percent of its absolute value (see Figure 6.1-3), but it is still too big for us to accept. A new approach has to be developed to address this variation.

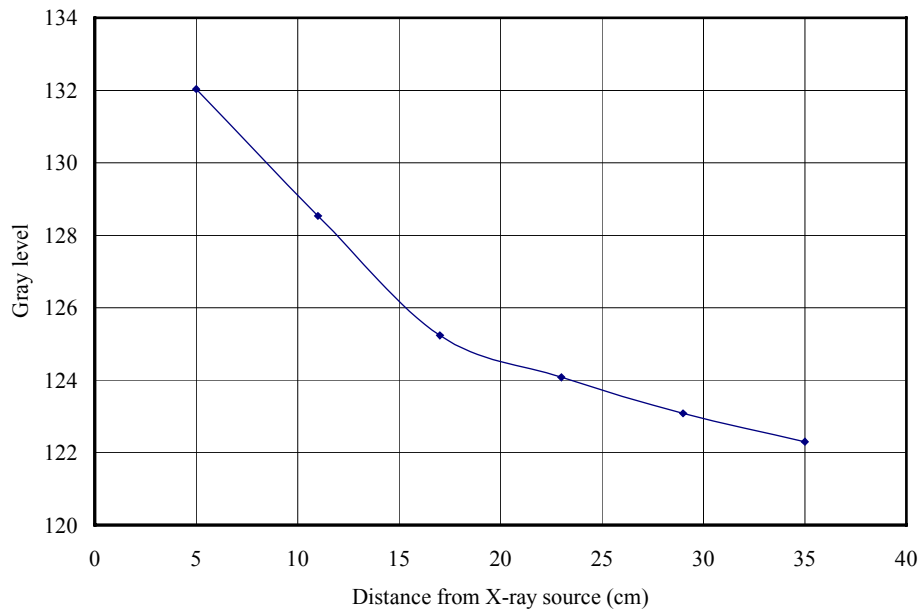


Figure 6.1-3 Variation of the averaged scattering gray levels with distance.

6.2 Treatment of Variation

6.2.1 Model description

As discussed above, the scattering signals vary as a function of distance from the x-ray source. Further study has shown that backward scattering is approximately proportional to the distance, and forward scattering is approximately inversely proportional to the distance. A possible method to treat these variations, i.e., to make them become less dependent on distance, is function approximation. By using higher orders of scattering signals rather than the simple average above, a fairly good result can be obtained. Its mathematical description is as follows:

$$\{a_0 F_L(d) + a_1 F_L^2(d) + \dots + a_{n-1} F_L^n(d)\} + \{b_0 B_L(d) + b_1 B_L^2(d) + \dots + b_{m-1} B_L^m(d)\} = K \quad (6.1)$$

In this equation, K is an assumed constant, a_{i-1} is the corresponding coefficient for the i th order of forward scattering, and b_{j-1} is the corresponding coefficient for the j th order of backward scattering. Empirically, it is also known that the increase of $F_L(d)$ and decrease of $B_L(d)$ are nearly linear functions of the distance. The problem thus can be simplified by selecting a second-order approximation of (6.1), yielding

$$a_0 F_L(d) + a_1 F_L^2(d) + b_0 B_L(d) + b_1 B_L^2(d) = K \quad (6.2)$$

By using vector form, the above equation can be rewritten as

$$w^T x = K \quad (6.3)$$

where $w = [a_0 \ a_1 \ b_0 \ b_1]^T$, and $x = [F_L(d) \ F_L^2(d) \ B_L(d) \ B_L^2(d)]^T$. We need to find the coefficients w so that for any input vector x , Equation (6.3) is tenable.

6.2.2 Solving for the coefficients with adaptive training

Adaptive modeling can be used to estimate w . The diagram is shown in Figure 6.2-1. It is assumed that if a system has the inputs and a desired output for each input, then the system could calculate the error between the actual and the desired output. The weights are adjusted, using a gradient decent method, so as to minimize the mean squared error [WID60].

For an input vector x , the output and mean squared error are expressed by (6.4) and (6.5) respectively,

$$y = w^T x \quad (6.4)$$

$$E[e^2] = E[(K - w^T x)^2] = K^2 - 2Kw^T E[x] + w^T E[xx^T]w \quad (6.5)$$

where $E[\]$ denotes the expected value, and the expectation is taken over all sets of input/desired output pairs. The above equation can be written in the following convenient form,

$$E[e^2] = K^2 - 2Kw^T h + w^T R w \quad (6.6)$$

where $h = E[x]$ gives the expected value for input vectors, and $R = E[xx^T]$ is the input correlation matrix.

Equation (6.6) is a quadratic function, and its gradient is

$$\nabla E[e^2] = \nabla(K^2 - 2Kw^T h + w^T R w) = -2Kh + 2Rw \quad (6.7)$$

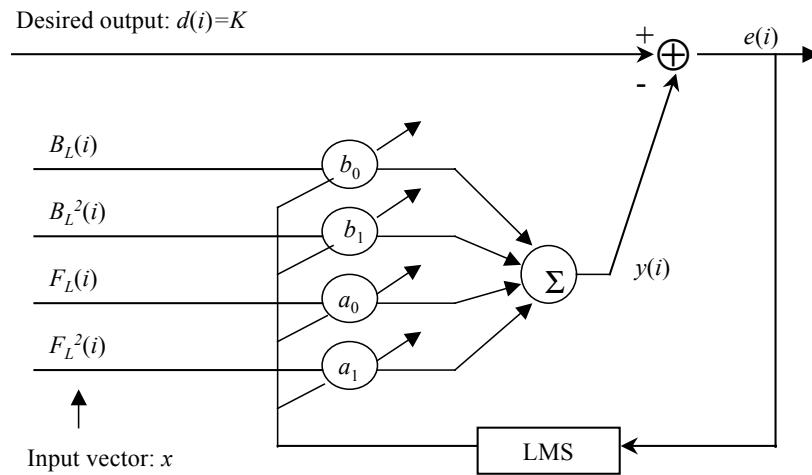


Figure 6.2-1 Block diagram of adaptive modeling.

The stationary point of $E[e^2]$ can be found by setting the gradient equal to zero,

$$-2Kh + 2Rw = 0 \tag{6.8}$$

Therefore, if the correlation matrix is positive definite there will be a unique stationary point, which will be a strong minimum,

$$w^* = KR^{-1}h \tag{6.9}$$

It is worth noting here that the existence of a unique solution depends only on the correlation matrix R . The characteristics of the input vectors therefore determine whether or not a unique solution exists [HAG95].

In general, however, it is not desirable and convenient to calculate h and R . For this reason an approximate steepest descent algorithm is used to estimate the gradient, which is called the Least Mean Square (LMS) algorithm. In LMS algorithm, the expectation of the squared error has been replaced by the squared error at iteration i ,

$$\hat{\nabla} E[e^2] = \nabla e^2(i) = 2e(i)\nabla e(i) = -2e(i)\nabla y(i) = -2e(i)x(i) \quad (6.10)$$

where $x(i)$ is the input vector at the i th iteration. The weights are adjusted according to,

$$w(i+1) = w(i) - \mu \nabla e^2(i) \quad (6.11)$$

where μ is step size of descent. By combining Equations (6.4), (6.10) and (6.11), the final algorithm is written as,

$$\left. \begin{aligned} y(i) &= w(i)^T x(i) \\ e(i) &= K - y(i) \\ w(i+1) &= w(i) + \mu \cdot e(i) \cdot x(i) \end{aligned} \right\} \quad (6.12)$$

The measured back and forward scatters are repeatedly used to update the four weights. When they are converged (see Figure 6.2-2), the network will give an output (Figure 6.2-3) with less than 4% errors, regardless of the distance variation that the object is away from the x-ray source.

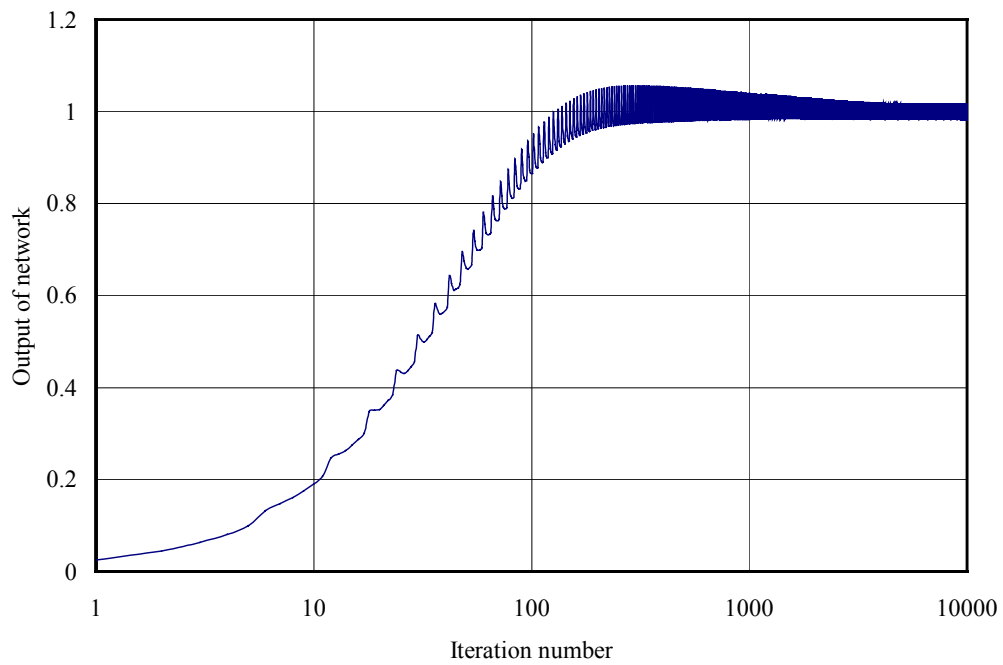


Figure 6.2-2 Learning curve of LMS algorithm

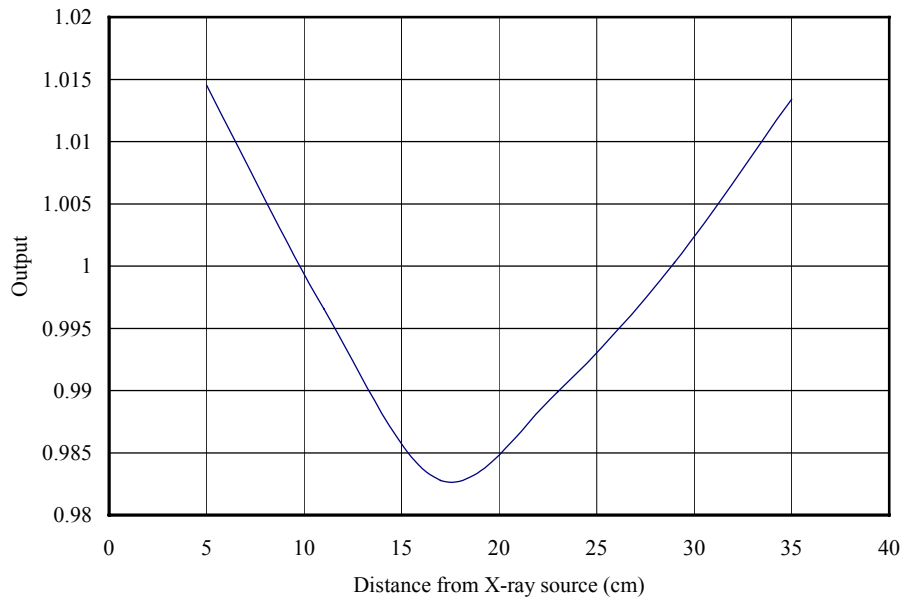


Figure 6.2-3 Output of the combination network processed with adaptive modeling. The desired output is set to 1.

6.2.3 Solving for the coefficients with Least Squares (LS)

Equation (6.3) can also be solved with linear least square's methods. The mathematical description of the problem is as follows: given $X \in \mathfrak{R}^{n \times 4}$ with $n > 4$, $K \in \mathfrak{R}^n$, select w to satisfy

$$\min_{w \in \mathfrak{R}^4} \|Xw - K\|_2 \quad (6.13)$$

where $X = [x_1, x_2, \dots, x_n]^T$, and n is the number of data points. The problem may be further interpreted as “Finding the linear combination of the four columns of X that is closest to K in

the L_2 norm.” We know that the j th column is composed of the j th elements of all input vectors, such as $v_j = [x_1(j), x_2(j), \dots, x_n(j)]^T$, thus yielding,

$$\left. \begin{aligned} v_1 &= [F_L(1), F_L(2), \dots, F_L(n)]^T \\ v_2 &= [F_L^2(1), F_L^2(2), \dots, F_L^2(n)]^T \\ v_3 &= [B_L(1), B_L(2), \dots, B_L(n)]^T \\ v_4 &= [B_L^2(1), B_L^2(2), \dots, B_L^2(n)]^T \end{aligned} \right\} \quad (6.14)$$

A theorem in numerical linear algebra [DEN83] states that the solution of Equation (6.13) is the set of the points $\{w^* : X^T(Xw^* - K) = 0\}$. If the columns of X are linearly independent, then w^* is unique, $X^T X$ is nonsingular, and

$$w^* = (X^T X)^{-1} X^T K \quad (6.15)$$

To apply the above equation, first we need to show that all columns of X are independent. For this purpose we give some theorems and one definition which are useful to our proof [LAN68] [HOF71].

Definition 6.1: Let A and B be $m \times n$ real matrices. We say that A is equivalent to B , and write $A \sim B$, if and only if, we can go from A to B by a finite number of elementary row and column operations.

Theorem 6.1: For a fixed m and n , the relation “ \sim ” defined in above is an equivalence relation on the set of all real $m \times n$ matrices.

Theorem 6.2: Equivalence matrices have the same row rank and the same column rank.

Theorem 6.3: The column rank of matrix A is equal to the maximum number of linearly independent column vectors of A .

The proof, for a representative set of data, is as follows:

- (i) Write up $X = [v_1, v_2, v_3, v_4]$,

$$X = \begin{bmatrix} 0.4392 & 0.1929 & 0.6703 & 0.4493 \\ 0.4763 & 0.2268 & 0.6038 & 0.3646 \\ 0.5112 & 0.2613 & 0.5412 & 0.2928 \\ 0.5500 & 0.3025 & 0.4927 & 0.2428 \\ 0.5879 & 0.3456 & 0.4464 & 0.1993 \\ 0.6236 & 0.3889 & 0.4041 & 0.1633 \end{bmatrix} \quad (6.16)$$

- (ii) Orthogonal-triangular decomposition, $X = QR$,

$$Q = \begin{bmatrix} -0.3351 & -0.5797 & 0.6196 & -0.4038 & -0.0665 & -0.0189 \\ -0.3634 & -0.4109 & -0.0518 & 0.7504 & 0.3282 & 0.1601 \\ -0.3900 & -0.2206 & -0.4928 & -0.0079 & -0.4960 & -0.5570 \\ -0.4196 & 0.0256 & -0.4071 & -0.3318 & -0.0958 & 0.7337 \\ -0.4485 & 0.3023 & -0.0836 & -0.2891 & 0.7018 & -0.3526 \\ -0.4758 & 0.5953 & 0.4449 & 0.2829 & -0.3744 & 0.0331 \end{bmatrix} \quad (6.17)$$

$$R = \begin{bmatrix} -1.3108 & -0.7160 & -1.2543 & -0.6662 \\ 0 & 0.0811 & -0.3679 & -0.3112 \\ 0 & 0 & 0.0593 & 0.0724 \\ 0 & 0 & 0 & -0.0021 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (6.18)$$

(iii) $\text{rank}(X) = \text{rank}(R) = 4$;

(iv) All four columns of X are linearly independent.

Substituting the measurement data into Equation (6.15) will yield w^* . A verification of the coefficients is shown in Figure 6.2-4. The variation after processing is narrowed into a range of less than 1%. By comparing Figure 6.2-4 with Figure 6.2-3, it was found the least squares method gives a better result, and the coefficient obtained with this method will be used in the prototype x-ray scanning system.

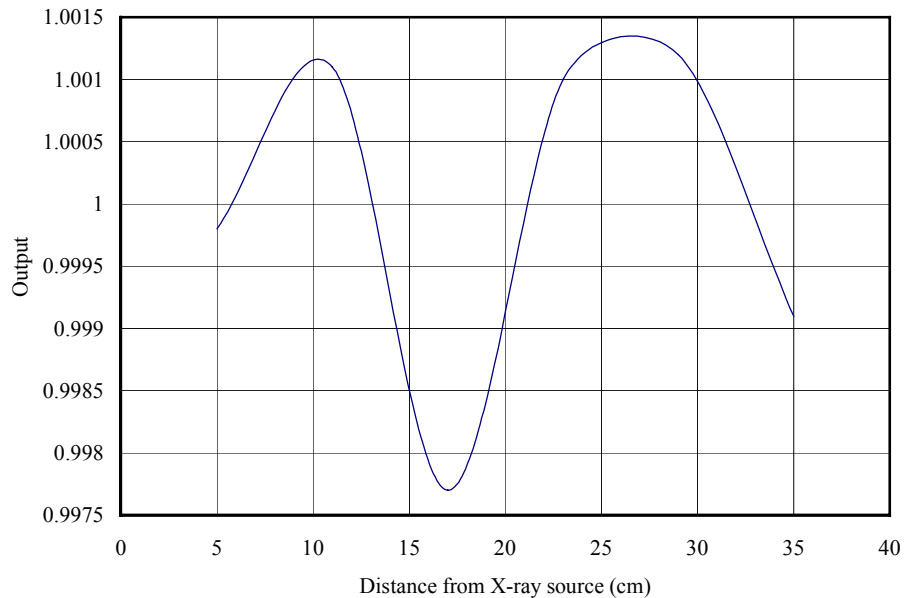


Figure 6.2-4 Output of the combination network processed with the linear least squares.