Higher Order Sensitivity Theory and Nonlinear Optimization in Fusion Neutronics Studies

T. Wu and C.W. Maynard

January 1981

UWFDM-402

FUSION TECHNOLOGY INSTITUTE

UNIVERSITY OF WISCONSIN

MADISON WISCONSIN

Higher Order Sensitivity Theory and Nonlinear Optimization in Fusion Neutronics Studies

T. Wu and C.W. Maynard

Fusion Technology Institute
University of Wisconsin
1500 Engineering Drive
Madison, WI 53706

http://fti.neep.wisc.edu

January 1981

UWFDM-402

HIGHER ORDER SENSITIVITY THEORY AND NONLINEAR OPTIMIZATION IN FUSION NEUTRONICS STUDIES[†]

Tang Wu[*]
and
Charles W. Maynard

Fusion Engineering Program
Nuclear Engineering Department
University of Wisconsin
Madison, Wisconsin 53706

January 1981

UWFDM-402

[†]Research supported in part by the Department of Energy and Wisconsin Electric Utilities Research Foundation

[*]Present address: General Electric Company, P.O. Box 508, Sunnyvale, California 94086

ABSTRACT

Higher order sensitivity terms can be derived by the utilization of functional analysis and the Green's function and are explicitly expressed as a combination of the flux, adjoint flux, and the Green's function, all from the unperturbed system, along with the perturbation in the transport operator. A simple neutron attenuation problem is examined to check the applicability of the first order sensitivity approximation for different magnitudes of variations. The quadratic approach in a nonlinear optimization problem is formulated using the results of the second order sensitivity approximation. It yields satisfactory results when applied to the NUWMAK inner shield design for the purpose of minimizing atomic displacements.

A. Higher Order Sensitivity Theory

The method of first-order perturbation theory or variational principles has been utilized historically in obtaining the sensitivity function for a response R with respect to the data field $\{\Sigma\}$.[1-3] The variation in the response, δR, can be expressed as

$$\delta R \simeq (\delta\Sigma_r, \phi) - (\phi^*, \delta L\phi), \qquad (1)$$

where, ϕ and ϕ^* are forward and adjoint fluxes, respectively, δL is the variation in transport operator L, (,) denotes integration over the phase space, and Σ_r is the response function such that $R = (\Sigma_r, \phi)$. Eq. (1) is said to have an error of order 2. Usually Eq. (1) would be adequate for estimating the variation in the response provided the perturbation in the data field is small. However, in neutronics design we often encounter problems that are not limited to a small variation in the data field and/or problems in which the response cannot be linearized with respect to the data field, where the first order approximation would provide a poor estimate of δR. In that case, a higher-order approach would be needed. Unfortunately, the simple perturbation theory and variational approach fail to give analytic and systematic solutions for the higher order sensitivity terms.

In another approach Oblow[4] suggested a differential method which features the implementation of a direct derivative and Taylor series expansion of the response. The response R is expressed as an integral functional of the form

$$R = \int F[\Sigma(x), \phi(x)]\, dx, \tag{2}$$

where F is an explicit function of Σ and ϕ, and x is a point in phase space. A small perturbation $\delta\Sigma(x)$ is applied to the system and the variation in response will be

$$\delta R = R[\Sigma(x) + \delta\Sigma(x)] - R[\Sigma(x)]$$

$$= \int \frac{dR}{d\Sigma(x)}\, \delta\Sigma(x)dx + \frac{1}{2!}\int \frac{d^2R}{d\Sigma(x)^2}[\delta\Sigma(x)]^2 dx + \ldots \quad . \tag{3}$$

Then the rest of the procedure will be the calculation of derivatives $dR/d\Sigma(x)$, $d^2R/d\Sigma(x)^2$, ..., etc. Thus, this approach enables us to derive an expression for the variation in response as a sum of a Taylor expansion. However, there are three questions concerning the validity of this approach which need to be examined: (1) The dimensions are not consistent in Eq. (2). (2) The definition of the derivative is not clear since R is a functional of Σ, not a function. (3) It is not proved that the Taylor expansion of a functional will have the form of Eq. (3) . Without answering these three questions, one cannot take for granted that this approach is valid.

In this section we will discuss a more rigorous approach for obtaining the higher order sensitivity terms. This approach features the employment of functional derivatives as well as the functional Taylor expansions. The Green's function plays a significant role here. Some mathematical background for the functional analysis will be briefly discussed. The derivation of the higher order sensitivity terms will be presented in a consistent manner. Also included is an illustrative example of calculating the higher order terms for a simple neutron attenuation problem in pure absorbers, with comparison to the theoretical predictions.

a. General Discussions of Functional Analysis

The definition of a linear vector space, or simply, a vector space, is an extension of the familiar concept of vectors. Here we will simply say that a vector space X over a field F is a set of vectors, without citing its formal mathematical definition.[5,6] X is a real vector space if the field is the familiar real number field R. A real vector space will usually be denoted by X or Y. However, Euclidean n-space is denoted by R^n and the real line by R. When X is an inner product space we will denote the inner product by $<,>$. A norm on X is defined as a mapping $N: X \to R$, mapping vectors in X, a vector space over a field F, into real numbers if and only if

(i) $N(x) \geq 0$ for every $x \in X$, and $N(x) = 0$ if and only if $x = 0$

(ii) $N(ax) = |a|\, N(x)$ for every $x \in X$ and for every $a \in F$

(iii) $N(x+y) \leq N(x) + N(y)$ for every $x, y \in X$.

$N(x)$ is also called the norm of x and the standard notation

$$N(x) = \| x \|$$

is used. A vector space X is called a normed space or a normed linear space if X is normable, i.e., some norm can be defined on the vectors $x \in X$. A mapping J defined from a subset D of a vector space X into a vector space Y is called an operator (or a transformation) and is denoted by $J : X \to Y$ or $J : D \to Y$. The set D is called the domain of J. The vector space of all operators from X into Y is denoted by $[X,Y]$. A real-valued operator, i.e., a member of $[X,R]$ where R is the real number field, is called a functional. The operator $J : X \to Y$ is said to be

(i) homogeneous if $J(ax) = aJ(x)$ and

(ii) additive if $J(x+y) = J(x) + J(y)$

for all $a \in R$ and all x and y in the domain of J. An additive and homogeneous operator is said to be linear. By a nonlinear operator we mean any operator which is not necessarily linear. If X and Y are topological vector spaces, - where some form of topology can be defined on the vector space, e.g., measure, norm -, then the vector space of all continuous linear operators from X into Y is

denoted by $L_1[X,Y]$. Sometimes $L_1[X,Y]$ is denoted by X^* and called the dual space of X. If X and Y are normed linear spaces and J is a linear operator from X into Y, then J said to be <u>bounded</u> if there exists $M>0$ such that $||J(x)|| \leq M||x||$ for all $x \in X$. It is well known that J is continuous if and only if it is bounded. If we let $||J|| = \inf\{M : ||J(x)|| \leq M||x||, x \in X\}$, the $L_1[X,Y]$ becomes a normed linear space. In particular X^* is a normed linear space whenever X is a normed linear space. Consequently, $L_n[X,Y]$ is a normed linear space whenever X is a normed linear space, where $L_n[X,Y]$ denotes the space $L_1[X,L_{n-1}[X,Y]]$. This property will play an important role in generating higher order sensitivity terms.

In general the concept of the derivative of an operator in a vector space can be viewed as an extension of the definition of the gradient for a real-valued function in R^n space. There are two kinds of derivatives commonly used by mathematicians, namely, the Gâteaux derivative and the Fréchet derivative[7]. The fundamental idea of the differential calculus is considered to be the local approximation of operators by linear operators. Recall that for a function $J : R \to R$, the derivative of J at f for f in R is defined as

$$J'(f) = \lim_{t \to 0} \left[\frac{J(f+t) - J(f)}{t} \right] \qquad (4)$$

and the line

$$J(f) = J(\hat{f}) + J'(\hat{f})\ (f-\hat{f}) \tag{5}$$

is a good approximation to $J(f)$ for f near $\hat{f}$. Let h denote a small increment in f at $\hat{f}$. Then the second term on the right hand side in Eq. (5) is $J'(\hat{f})\cdot h$, which can be interpreted as an operator $J'(\hat{f})$ mapping h to $J'(\hat{f})\ (h)$. If instead we consider $J : R^2 \to R$, then Eq. (4) no longer makes sense since we cannot add scalars and vectors. However, Eq. (4) can be generalized for an operator $J : X \to Y$ where X need only be a vector space and Y a topological vector space such that

$$J'(f)\ (h) = \lim_{t\to 0} \left[\frac{J(f+th) - J(f)}{th}\right] \cdot h$$

$$= \lim_{t\to 0} \left[\frac{J(f+th) - J(f)}{t}\right] \tag{6}$$

for f and h in X and t in R. Eq. (6) leads to the formal definition of the Gâteaux derivatives:

Consider an operator $J : X \to Y$ where X is a vector space and Y is a topological vector space. Given f and h in X suppose

$$DJ(f)\ (h) = \lim_{t\to 0}\left[\frac{J(f + th) - J(f)}{t}\right] \tag{7}$$

exists. Then $DJ(f)\ (h) \in Y$ is called the Gâteaux derivative (some

mathematicians prefer the terminology of Gâteaux variation or differential) of J at f in the direction of h and we say J is Gâteaux differentiable at f in the direction h. Also, J is Gâteaux differentiable at f when J is Gâteaux differentiable at f in every direction. In this case the operator

$$DJ(f) : X \to Y$$

which assigns to each $h \in X$ the vector $DJ(f)(h) \in Y$ is called the Gâteaux derivative of J at f. And the operator

$$DJ: X \to [X,Y]$$

which assigns to $f \in X$ the operator $DJ(f) \in [X,Y]$ is called the Gâteaux derivative of J.

The usefulness of the Gâteaux derivative is limited by the fact that the vector space [X,Y] is not a topological space since there is no way to define a topology on it. Consequently, we cannot consider the Gâteaux derivative of the Gâteaux derivative, which is an essential necessity in higher order variations. This deficiency is caused by the weak condition that X need only be a vector space. For a normed linear space which we are really interested in, another form of derivatives can be defined:

Consider $J: X \to Y$ where both X and Y are normed linear spaces. Given $f \in X$, if a linear operator $J'(f) \in L_1[X,Y]$ exists such that

$$\lim_{\|h\|\to 0} \frac{\|J(f+h) - J(f) - J'(f)(h)\|}{\|h\|} = 0, \quad h\in X, \tag{8}$$

then $J'(f)$ is called the Fréchet derivative of J at f.

The operator

$$J' : X \to L_1[X,Y]$$

which assigns $J'(f)$ to f is called the Fréchet derivative of J. As we mentioned before, $L_1[X,Y]$ is a normed linear space when X and Y are both normed linear spaces. Therefore, we are able to define J'', the Fréchet derivative of the Fréchet derivative of J. Clearly

$$J'' : X \to L_2[X,Y]$$

where $L_2[X,Y]$ denotes $L_1[X, L_1[X,Y]]$. Consequently, the n-th Fréchet derivative of J is by definition the Fréchet derivative of $J^{(n-1)}$, the (n-1)-th Fréchet derivative of J, and

$$J^{(n)} : X \to L_n[X,Y].$$

The Fréchet derivatives have some familiar names in R^n space. Consider a functional $J: R^n \to R$, the following expression holds:

$$J'(f)(h) = \langle \nabla(f), h \rangle, \quad f, h \in R^n, \tag{9}$$

where the vector $\nabla J(f)$ is the gradient of J at f.

The Fréchet derivative of the gradient of J, for historical reasons, is called the Hessian of J and denoted by H. If $J : R^n \to R$, then $\nabla J = \left(\frac{\partial J}{\partial f_1}, \ldots, \frac{\partial J}{\partial f_n} \right) : R^n \to R^n$; hence,

$$H(f)(h) = \begin{pmatrix} \frac{\partial^2 J}{\partial f_1 \partial f_1} & \cdots & \frac{\partial^2 J}{\partial f_1 \partial f_n} \\ \vdots & & \\ \frac{\partial^2 J}{\partial f_n \partial f_1} & \cdots & \frac{\partial^2 J}{\partial f_n \partial f_n} \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} . \tag{10}$$

Also, for $f,h,k \in R^n$, we have

$$J''(f)(h,k) = \langle H(f)(h), k \rangle .$$

The Taylor's series expansion for an operator in a complete normed linear space can have several forms. The following expression, Young's form of Taylor's theorem, will be used in connection with the Fréchet derivatives. Let an operator $J : X \to Y$ be n times Fréchet differentiable at a point $f \in X$. Then

$$J(f+h) = J(f) + J'(f)(h) + \frac{1}{2} J''(f)(h,h) + \ldots$$

$$+ \frac{1}{n!} J^{(n)}(f)(h,\ldots,h) + r(f;h), \tag{11}$$

where

$$\lim_{h \to 0} \frac{\| r(f;h) \|}{\| h \|^n} = 0.$$

The materials discussed in this section will be utilized in the next section for deriving higher order sensitivity terms. The vector space will be a set of cross sections as a function of phase space variables, reaction types and nuclides. The main functional is the response R. Fréchet differentiability is assumed without proof. The following notations are used throughout the next section:

$$\begin{aligned}
\delta J(f;h) &= J'(f)\ (h) \\
\delta^2 J(f;h^2) &= J''(f)\ (h,h) \\
\delta^2 J(f;h_1,h_2) &= J''(f)\ (h_1,h_2) \\
&\vdots \\
\delta^n J(f;h^n) &= J^{(n)}(f)\ (h,\ldots,h),
\end{aligned}$$

where J is an operator or functional, f is a vector, h is an increment in f, and $J^{(n)}(f)$ is the n-th Fréchet derivative of J at f.

b. Derivation of Higher Order Sensitivity Terms via Green's Function

Let R be the response of interest and R can be expressed as a functional of the form

$$R = \int_{X_o} F[\Sigma,\phi]\, dX_o, \tag{12}$$

where Σ is the data field and a point in vector space, X_o is a

point in phase space, and ϕ is the flux which is the solution of the transport equation:

$$L(X)\,\phi(X) = S(X). \tag{13}$$

By assuming that R is n-th Fréchet differentiable at Σ with variation h in Σ (i.e., $h = \delta\Sigma$), the Taylor series expansion of R in h can be expressed as

$$\delta R = R(\Sigma+h) - R(\Sigma)$$

$$= \sum_{n=1}^{\infty} \frac{1}{n!}\,\delta^n\,R(\Sigma;h^n)\ , \tag{14}$$

where $\delta^n R(\Sigma;h^n)$ is the n-th Fréchet derivative of R at Σ in the direction of h.

In fusion reactor neutronics we are usually interested in a linear functional response of the flux, which has the form $F = \Sigma_r\phi$ with Σ_r being the response function defined in a detector domain D. That is,

$$R = \int_D \Sigma_r(X_o)\phi(X_o)\,dX_o. \tag{15}$$

The Fréchet derivative of R, $\delta R(\Sigma;h)$, can therefore be derived by differentiating Eq. (15):

$$\delta R(\Sigma;h) = \int_D [\phi(X_o)\ \delta\Sigma_r(X_o;h) + \Sigma_r(X_o)\ \delta\phi(X_o;h)]dX_o, \tag{16}$$

where $\delta\phi(X_o;h)$ and $\delta\Sigma_r(X_o;h)$ are the Frechet derivatives of $\phi(X_o)$ and $\Sigma_r(X_o)$, respectively, at Σ with the increment h. Here we have used the important fact that $\phi(X_o)$ is also a functional of Σ via the transport equation.

Since Σ_r is usually linear to Σ, i.e., $\delta^n\Sigma_r(X_o;h^n) = 0$ for $n \geq 2$, the n-th Frechet derivative of R can be easily obtained by differentiating Eq. (16) n-1 times:

$$\delta^n R(\Sigma;h^n) = \int_D n\, \delta\Sigma_r(X_o;h)\, \delta^{n-1}\phi(X_o;h^{n-1})\, dX_o + \int_D \Sigma_r(X_o)\, \delta^n\phi(X_o;h^n)\, dX_o \,. \tag{17}$$

Hence, the remaining job is to find an expression for the n-th Frechet derivative of $\phi(X_o)$.

Here we will calculate $\delta^n(X_o;h^n)$ by following a procedure suggested by Dubi and Dudziak[8] which utilizes the Green's function and the adjoint Green's function. The role of the Green's functions in calculating higher-order sensitivity terms was first pointed out by Albert.[9] Let us define the Green's function $G(X,X_1)$ as a solution of the equation

$$L(X)\, G(X,X_1) = \delta(X-X_1), \tag{18}$$

where X, X_1 are points in phase space and $\delta(X-X_1)$ is the phase

space delta function. Clearly $G(X,X_1)$ gives the flux at X due to a unit point source at X_1. Since the transport operator is linear, the flux at X_o due to an external source distribution S(X) can be presented as

$$\phi(X_o) = \int_X G(X_o,X)\, S(X)\, dX \ . \tag{19}$$

Similarly the adjoint Green's function $G^*(X,X_2)$ is defined as the solution of the adjoint transport equation

$$L^+(X)\, G^*(X,X_2) = \delta(X-X_2) \ . \tag{20}$$

Therefore, $G^*(X,X_2)$ has the meaning of the adjoint flux at X due to a unit source at X_2. The adjoint flux at X_o due to an adjoint source distribution $S^*(X)$ is then

$$\phi^*(X_o) = \int_X G^*(X_o,X)S^*(X)\, dX \ . \tag{21}$$

The relation between the Green's function and the adjoint Green's function could easily be obtained from the well-known inner-product relationship:

$$\int_X G^*(X,X_2)L(X)G(X,X_1)dX = \int_X G(X,X_1)L^+(X)G^*(X,X_2)dX,$$

$$\text{or} \quad \int_X G^*(X,X_2)\, \delta(X-X_1)dX = \int_X G(X,X_1)\, \delta(X-X_2)\, dX \ . \tag{22}$$

That is,

$$G^*(X_1,X_2) = G(X_2,X_1) \ . \tag{23}$$

Substituting G^* for G in Eq. (19), we get

$$\phi(X_o) = \int_X G^*(X,X_o)\, S(X)\, dX \ . \tag{24}$$

The n-th Frechet derivative of $\phi(X_o)$ is therefore

$$\delta^n\phi(X_o;h^n) = \int_X \delta^n G^*(X,X_o;h^n)\, S(X)\, dX \ . \tag{25}$$

From Eq. (25) it is obvious that our problem now is to find an expression for $\delta^n G^*(X,X_o;h^n)$.

The definition of the adjoint Green's function,

$$L^+(X)G^*(X,X_o) = \delta(X - X_o), \tag{26}$$

yields the following expression by differentiating Eq. (26):

$$L^+(X)\ \delta G^*(X,X_o;h) = -\ \delta L^+(X;h)\ G^*(X,X_o) \ . \tag{27}$$

Now using the fact that L^+ is linear in Σ, we get

$$\delta^n L^+(X;h^n) = 0 \quad \text{for} \quad n \geq 2. \tag{28}$$

Differentiating Eq. (27) once more and using Eq. (28), we have

$$L^+(X)\ \delta^2 G^*(X,X_o;h^2) = -2\ \delta L^+(X;h)\ \delta G^*(X,X_o;h) \ . \tag{29}$$

Therefore, a recursion relation can be found in the following

form:

$$L^{+}(X)\ \delta^{n}G^{*}(X,X_{o};h^{n}) = -n\ \delta L^{+}(X;h)\delta^{n-1}\ G^{*}(X,X_{o};h^{n-1})\ . \qquad (30)$$

Bearing Eq. (30) in mind, we now proceed to evaluate the first-order Frechet derivative of G^{*} by employing the following expression:

$$G^{*}(X_{2},X_{o}) = \int_{X_{1}} G^{*}(X_{1},X_{o})\,\delta(X_{1} - X_{2})\,dX_{1}\ .$$

$$\begin{aligned}
\delta G^{*}(X_{2},X_{o};h) &= \int_{X_{1}} \delta G^{*}(X_{1},X_{o};h)\ L(X_{1})\ G(X_{1},X_{2})\ dX_{1} \\
&= \int_{X_{1}} G(X_{1},X_{2})\ L^{+}(X_{1})\ \delta G^{*}(X_{1},X_{o};h)\ dX_{1} \\
&= -\int_{X_{1}} G(X_{1},X_{2})\ \delta L^{+}(X_{1};h)G^{*}(X_{1},X_{o})dX_{1} \\
&= -\int_{X_{1}} G^{*}(X_{1},X_{o})\ \delta L\ (X_{1};h)\ G(X_{1},X_{2})\ dX_{1} \\
&= -\int_{X_{1}} G(X_{o},X_{1})\ \delta L\ (X_{1};h)\ G(X_{1},X_{2})\ dX_{1}\ . \qquad (31)
\end{aligned}$$

Similarly for the second-order derivative,

$$\begin{aligned}
&\delta^{2}G^{*}(X_{3},X_{o};h^{2}) \\
&= \int_{X_{2}} \delta^{2}G^{*}(X_{2},X_{o};h^{2})\,\delta\,(X_{2} - X_{3})\,dX_{2} \\
&= \int_{X_{2}} \delta^{2}G^{*}(X_{2},X_{o};h^{2})L(X_{2})G(X_{2},X_{3})\,dX_{2}
\end{aligned}$$

$$= \int_{X_2} G(X_2,X_3)L^+(X_2)\ \delta^2G^*(X_2,X_o;h^2)\ dX_2$$

$$= -2\int_{X_2} G(X_2,X_3)\ \delta L^+(X_2;h)\delta G^*(X_2,X_o;h)\ dX_2$$

$$= -2\int_{X_2} \delta G^*(X_2,X_o;h)\ \delta L(X_2;h)\ G(X_2,X_3)\ dX_2$$

$$= 2\int_{X_1}\int_{X_2} G(X_o,X_1)\ \delta L(X_1;h)\ G(X_1,X_2)\ \delta L(X_2;h)G(X_2,X_3)$$

$$dX_1dX_2\ . \qquad (32)$$

By looking at the similarity between Eq. (31) and Eq. (32), our intuition suggests that a general form for the n-th Frechet derivative of G^* would be

$$\delta^nG^*(X_{n+1},X_o;h^n) = (-1)^n\ n!\int_{X_1}\ldots\int_{X_n} P_n(X_o,\ldots,X_n)G(X_n,X_n+1)$$

$$dX_1\ldots dX_n\ , \qquad (33)$$

where $P_n(X_o,\ldots,X_n)$ is the flux perturbation operator and is defined as

$$P_n(X_o,\ldots,X_n) \equiv \prod_{i=1}^{n} G(X_{i-1},X_i)\ \delta L(X_i;h)\ . \qquad (34)$$

To prove the validity of Eq. (33), we follow the procedure of mathematical induction. We already know that Eq. (33) holds for n=1 (Eq. (31)). Suppose that Eq. (33) is true for n=k, then for n=k+1,

$$\delta^{k+1}G^*(X_{k+2},X_o;h^{k+1})$$

$$= \int_{X_{k+1}} \delta^{k+1} G^*(X_{k+1}, X_o; h^{k+1})\, \delta(X_{k+1} - X_{k+2})\, dX_{k+1}$$

$$= \int_{X_{k+1}} \delta^{k+1} G^*(X_{k+1}, X_o; h^{k+1})\, L(X_{k+1})\, G(X_{k+1}, X_{k+2})\, dX_{k+1}$$

$$= \int_{X_{k+1}} G(X_{k+1}, X_{k+2})\, L^+(X_{k+1})\, \delta^{k+1} G^*(X_{k+1}, X_o; h^{k+1})\, dX_{k+1}$$

$$= -(k+1) \int_{X_{k+1}} G(X_{k+1}, X_{k+2})\, \delta L^+(X_{k+1}; h)\, \delta^k G^*(X_{k+1}, X_o; h^k)\, dX_{k+1}$$

$$= -(k+1) \int_{X_{k+1}} \delta^k G^*(X_{k+1}, X_o; h^k)\, \delta L(X_{k+1}; h)\, G(X_{k+1}, X_{k+2})\, dX_{k+1}$$

$$= (-1)^{k+1} (k+1)! \int_{X_1} \cdots \int_{X_{k+1}} P_k(X_o, \ldots, X_k) G(X_k, X_{k+1}) \delta L(X_{k+1}; h) G(X_{k+1}, X_{k+2})\, dX_1 \cdots dX_{k+1}$$

$$= (-1)^{k+1} (k+1)! \int_{X_1} \cdots \int_{X_{k+1}} P_{k+1}(X_o, \ldots, X_{k+1}) G(X_{k+1}, X_{k+2})\, dX_1 \cdots dX_{k+1},$$

i.e., Eq. (33) holds for n=k+1. Thus, we have proved that, by induction, Eq. (33) is true for all positive integer n.

The rest of the derivation is straightforward. Substituting Eq. (33) into Eq. (25) and using the characteristic of the Green's function,

$$\int_{X_{n+1}} G(X_n,X_{n+1})\, S(X_{n+1})\, dX_{n+1} = \phi(X_n) ,$$

yields:

$$\delta^n\phi(X_o;h^n) = (-1)^n\, n!\int_{X_1}\cdots\int_{X_n} P_n(X_o,\ldots,X_n)\,\phi(X_n)\, dX_1\ldots dX_n . \qquad (35)$$

The n-th Frechet derivative of the response is then represented by

$$\delta^n R(\Sigma;h^n) = (-1)^{n-1}\, n!\int_{X_o}\cdots\int_{X_{n-1}} \delta\Sigma_r(X_o;h)P_{n-1}(X_o,\ldots,X_{n-1})\,\phi(X_{n-1})dX_o\ldots dX_{n-1} + (-1)^n\, n!\int_{X_o}\cdots\int_{X_n} \Sigma_r(X_o)\, P_n(X_o,\ldots,X_n)\phi(X_n)\, dX_o\ldots dX_n . \qquad (36)$$

The first term is called the direct effect which is non-zero only if the response function in the detector domain is perturbed. The second term is called the indirect effect or the flux perturbation term which arises from the perturbation of the flux in the detector region through the particle transport when the system is perturbed.

Looking at Eq. (36) in the case of n=1, i.e., the first-order term, we get

$$\delta R(\Sigma;h) = \int_{X_o} \delta\Sigma_r(X_o;h)\phi(X_o)\, dX_o$$

$$-\int_{X_o}\int_{X} \Sigma_r(X_o)G(X_o,X)\delta L(X;h)\phi(X)\, dX_o\, dX$$

$$= \int_{X_o} \delta\Sigma_r(X_o;h)\ \phi\ (X_o)\, dX\ .$$

$$-\int_{X} \phi^*(X)\ \delta L(X;h)\ \phi\ (X)\, dX\ , \qquad (37)$$

which is identical with the result of first-order perturbation theory or variational principles. Again the physical meaning of Eq. (37) is clear. The integration over X in the indirect effect term yields the perturbation of the flux in the detector region due to a change in the transport operator over all the phase space, while the integration over X_o in the direct effect term merely accounts for contributions from any specific variation of the response function itself in the detector domain.

There are some intriguing aspects associated with the higher-order sensitivity analysis. First, the n-th order term is explicitly expressed as a combination of the flux, adjoint flux, and a detailed Green's function, all from the unperturbed system, along with the perturbation in the transport operator. Once the Green's function $G(X_1,X_2)$ is defined for every pair (X_1,X_2) within the perturbed region, the perturbation can be calculated to any

degree of accuracy in principle. Also, the additional information required to calculate the second-order term, i.e., the Green's function, will suffice to calculate any higher-order terms. However, the applicability of this approach may be severely restricted by the fact that the calculation of the Green's function might require a tremendous amount of effort, unless we can obtain an analytic solution or approximation for the Green's function. One important observation is that we are not limited to small perturbations since the Taylor's series expansion is infinite, provided the expansion converges.

In cross section sensitivity analysis, the first-order term is linear in δL. Therefore, the total cross section sensitivity is naturally the sum of the partial sensitivities in first-order analysis. This property often serves as an indication of whether the sensitivity calculation is self-consistent. However, the second and higher order terms are no longer linear in δL which can be seen from Eq. (37) and the definition of P_n. The total cross section sensitivity of higher-order terms will involve cross terms from the partial sensitivities and is not merely the sum of the partial sensitivities.

c. An Illustrative Example

Consider a monoenergetic, monodirectional neutron beam of intensity S $n/(cm^2\text{-sec})$, incident normally on the surface of an

infinite planar slab of thickness a. The slab consists of a pure absorber with the macroscopic absorption cross section Σ (Fig. 1.a). The response of interest is taken as the neutron flux emerging from the slab. Hence, the response function is a delta function $\delta(X-a)$ and the response R is $\phi(a)$. The flux distribution across the slab can be obtained analytically as

$$\phi(X) = S\, e^{-\Sigma X} , \quad 0 \leq X \leq a . \tag{38}$$

For any two points X_1 and X_2 inside the slab, the Green's function is

$$G(X_1,X_2) = e^{-\Sigma (X_1 - X_2)} \text{ for } X_1 > X_2,$$

$$= 0 \quad \text{otherwise.} \tag{39}$$

Now consider a perturbation applied to the system with a uniform increment of Σ such that $\Sigma \rightarrow \Sigma+h$. Thus, we have

$$\delta L\,(X) = h \quad \text{for } 0 \leq X \leq a. \tag{40}$$

From the previous section, the change in the response, δR, can be expressed as

$$\delta R = \sum_n \frac{1}{n!}\, \delta^n R(\Sigma;h^n)$$

$$= \sum_n (-1)^n \int_{X_n} \cdots \int_{X_o} \Sigma_r(X_o) P_n(X_o,\ldots,X_n) \phi(X_n)\, dX_o \ldots dX_n , \tag{41}$$

where P_n is the flux perturbation operator given by

$$P_n(X_o,\ldots X_n) = \prod_{i=1}^{n} G(X_{i-1},X_i)\delta L(X_i) \quad . \tag{42}$$

Let $(\delta R)_n$ be the n-th order contribution to δR such that $\delta R = \sum_n (\delta R)_n$.

Upon utilizing Equations (38) to (42) and the following expression

$$\int_{X_o} \Sigma_r(X_o)\, G(X_o,X_1)\, dX_o = G(a,X_1) = e^{-\Sigma(a-X_1)} \quad , \tag{43}$$

we get

$$\begin{aligned}(\delta R)_n &= (-1)^n S e^{-\Sigma a} h^n \int_o^a \int_{X_n}^a \cdots \int_{X_2}^a e^{\Sigma\{X_1-(X_1-X_2)-\cdots \\ &\qquad -(X_{n-1}-X_n)-X_n\}} dX_1 \ldots dX_n \\ &= R(-h)^n \int_o^a \int_{X_n}^a \cdots \int_{X_2}^a dX_1 \ldots dX_{n-1} dX_n \quad . \end{aligned} \tag{44}$$

The integration can be carried out by using the following identity

$$\int_{X_n}^a \cdots \int_{X_2}^a dX_1 \ldots dX_{n-1} = \frac{1}{(n-1)!}(a - X_n)^{n-1} \quad , \tag{45}$$

which can be proved by mathematical induction. Thus, we have

$$\begin{aligned}&\int_o^a \int_{X_n}^a \cdots \int_{X_2}^a dX_1 \ldots dX_{n-1}\, dX_n \\ &= \int_o^a \frac{1}{(n-1)!}(a - X_n)^{n-1}\, dX_n\end{aligned}$$

$$= \int_0^a \frac{1}{(n-1)!} \, Y^{n-1} dY$$

$$= \frac{a^n}{n!} \tag{46}$$

That is,

$$(\delta R)_n = R(-ha)^n/n! \ , \tag{47}$$

and

$$\delta R = R \sum_{n=1}^{\infty} (-ha)^n/n! \ . \tag{48}$$

Let δR_N denote the prediction by N-th order sensitivity theory with the definition

$$\delta R_N = \sum_{n=1}^{N} (\delta R)_n = R \sum_{n=1}^{N} (-ha)^n/n! \ . \tag{49}$$

Theoretically the perturbed response can be found as

$$R' = e^{-\Sigma' a}$$

$$= e^{-(\Sigma+h)a}$$

$$= R \, e^{-ha} \tag{50}$$

Therefore, $\delta R = R' - R$

$$= R \, (e^{-ha} - 1)$$

$$= R \sum_{n=1}^{\infty} (-ha)^n/n! \ , \tag{51}$$

which is identical to Eq. (48). It is obvious that the change in response with system perturbation can be accurately predicted by the sensitivity theory via Green's function approach for this simple system. The relative difference between the perturbed response predicted by the N-th order sensitivity approximation, R'_N, and the true perturbed response, R'_{true}, is therefore a function of N and ah:

$$d = (R'_N - R'_{true})/R'_{true}$$

$$= \frac{\sum_{n=0}^{N} (-ah)^n/n!}{\sum_{n=0}^{\infty} (-ah)^n/n!} - 1 . \qquad (52)$$

Table 1 tabulates the values of d in units of percentage as a function of N and ah with the blanks representing those d less than 1%. A graphical presentation is shown in Fig. 1.b where a range between 1% and 100% for d is used.

From Table 1 one can observe the range of applicability of the N-th order sensitivity approximation for a simple particle attenuation problem as a function of ah. For instance, the first-order sensitivity theory would be accurate enough to predict the perturbed response if the perturbation ah is 0.2 and the error tolerance in predicting the response is 2%. On the other

hand, one would have to use the fourth-order approximation to satisfy the same error tolerance if ah were 1.0 since the first-order approximation will yield an error of 100%. Recall that $ah = (\Sigma a)\ (\Delta\Sigma/\Sigma)$, where Σa is the thickness in mean free path (mfp) and $\Delta\Sigma/\Sigma$ is the relative change in the cross section. In the cross section sensitivity analysis for this simple system, the applicable range of the fractional cross section change can therefore be expressed as a function of thickness, in units of mfp, and the order of sensitivity approximation.

Discussion of the simple flux attenuation problem in a system of pure absorbers can be easily extended to shielding analysis in fusion reactor designs. As reported by Abdou[10], it is found that the total energy leakage to the superconducting toroidal field coils (TFC) is exponentially attenuated as a function of the thickness of the bulk shield in a tokamak fusion device. Also, Wu and Maynard[11] find that both the neutron energy leakage to the magnet and the atomic displacement in the aluminum stabilizer of the magnet, in the NUWMAK design, are most sensitive to the tungsten 14 MeV (n,2n) cross section which is the most significant neutron removal cross section at that energy. Let σ_{ij} represent the transfer cross section from energy group i to j and let group 1 be the 14 MeV group. Then the removal cross section for group 1, σ_1^r, can be expressed as

$$\sigma_1^r = \sigma_1^t - \sigma_{11} = \sigma_1^a + \sum_{j>1} \sigma_{1j} ,$$

where σ^t and σ^a represent total and absorption cross sections, respectively. For a shielding problem where the 14 MeV group has the dominant effect, as in the case of radiation damage to a TFC in tokamak fusion reactors, the sensitivity analysis would have similar characteristics as the pure absorber if Σ is replaced by σ_1^r, provided the in-group scattering (σ_{11}) is negligible. Thus, the result from the simple flux attenuation problem, Table 1 and Fig. 1.b could provide a basis of checking the range of applicability for the N-th order sensitivity approximation when the system is perturbed in fusion reactor shielding design.

B. Second Order Approximation in Nonlinear Optimization

The theory of higher order sensitivity approximation discussed in the previous section can be applied to an interesting problem of finding an optimal set of variables that will maximize or minimize a given detector response in fusion neutronics design. An optimization procedure is reported by Greenspan et al.,[12] through the use of the second order sensitivity approximation, where the sensitivity terms are derived from a differential approach suggested by Oblow.[4] In this section we will utilize the results formulated in the previous section for the optimization problem in

source-driven systems.

Consider a nonlinear real-valued function $y = f(x)$. To find a point $x = x^*$ extremizing y the necessary condition is that $y'(x^*) = 0$ for a constraint-free problem. For the case that an analytic expression of $y'(x)$ for all x is available, the problem reduces to zero-finding for $y'(x)$, which is considered to be an easier task. Unfortunately, that is not the usual case. An alternative way is to use Taylor's expansion and retain the first two terms:

$$\delta y = y - y_o = f'(x_o)\delta x + \frac{1}{2} f''(x_o) (\delta x)^2 , \tag{53}$$

i.e., a quadratic approximation of y around x_o. The condition of optimum is then

$$\partial(\delta y)/\partial(\delta x) = 0 \quad \text{at } x = x^* , \tag{54}$$

which enables us to find the optimal step size

$$(\delta x)^* = -f'/f'' . \tag{55}$$

and the change in y is

$$\delta y = -(f')^2/2f'' . \tag{56}$$

The iteration procedure continues for a new point $x_1 = x_o + (\delta x)^*$ until a pre-determined stopping rule is satisfied and the optimal is found. Thus, the quadratic approximation has the capability of

optimally determining the step size for the next point in the iteration process while the first-order approximation fails to do so. Consequently, one has to take some educated guess regarding the step size in the linear approximation. A small step length obviously would require more iterations to reach the optimal point and a large step length may overshoot the solution. Thus, the main impact of the quadratic approximation on optimization problems is its ability to reduce the number of iterations although some additional effort for calculating the second-order derivative is essential.

There is a very interesting result regarding the difference in the change of y. Let the optimal step length in Eq. (55) be applied to the linear approximation:

$$y - y_o = f'(x_o)\,(x - x_o)\ . \tag{57}$$

By Eq. (55),

$$(\delta y)_{1st\ order} = -(f')^2/f''\ . \tag{58}$$

Therefore, for the same optimal step length, the prediction of δy in the linear approximation is exactly twice that of the quadratic approximation, i.e.,

$$(\delta y)_{1st\ order} = 2(\delta y)_{2nd\ order}\ . \tag{59}$$

This identity could serve as a basis for checking the consistency in our calculations of f, f', and f'' as well as the question of whether the quadratic approximation is appropriate.

For a function of many variables, the above iteration procedure of one variable should still be adequate by replacing f' and f'' by the gradient vector and the Hessian matrix, respectively. As stated before, the formulation of functional analysis could be viewed as a natural extension of analysis for functions of many variables. It is therefore to be expected that, for the optimization problem of a functional, some forms of gradient-like and Hessian-like operators would exist in the mathematical formulations. The next section presents a rigorous derivation of the second order optimization procedure.

a. Mathematical Formulation of Quadratic Approximation

In the previous section, we derived the formulation of the n-th Frechet derivative for a response R as a functional of the data field Σ in the direction h. The mathematical expression is given in Eq. (36). Consider the cases n=1 and n=2:

$$\delta R(\Sigma;h) = \int_{X_o} \delta\Sigma_r(X_o;h)\phi(X_o)\, dX_o$$

$$-\int_X \int_{X_o} \Sigma_r(X_o)G(X_o,X)\delta L(X;h)\phi(X)\, dX_o\, dX, \qquad (60)$$

and

$$\delta^2 R(\Sigma;h,h) = -2\int_{X_1}\int_{X_o} \delta\Sigma_r(X_o;h)G(X_o,X_1)\delta L(X_1;h)\phi(X_1)\, dX_o dX_1$$

$$+2\int_{X_2}\int_{X_1} \phi^*(X_1)\delta L(X_1;h)G(X_1,X_2)\delta L(X_2;h)\phi(X_2)\, dX_1\, dX_2 \,. \tag{61}$$

Suppose that we are interested in the second-order derivative in the directions of h_1 and h_2, where h_1 and h_2 are not necessarily the same, then the simple expression of Eq. (61) will not be suitable for $\delta^2 R(\Sigma;h_1,h_2)$. However, $\delta^2 R(\Sigma;h_1,h_2)$ is still attainable upon following the same steps taken in deriving Eq. (36). Here we will only present the final result without showing the detailed steps:

$$\delta^2 R(\Sigma;h_1,h_2)$$

$$= -\int_{X_o}\int_{X_1} \{\delta\Sigma_r(X_o;h_1)G(X_o,X_1)\, L(X_1;h_2)$$

$$+ \delta\Sigma_r(X_o;h_2)G(X_o,X_1)\, L(X_1;h_1)\}\phi(X_1)dX_1 dX_o$$

$$+\int_{X_1}\int_{X_2} \phi^*(X_1)\{\delta L(X_1;h_1)G(X_1,X_2)\delta L(X_2,h_2)$$

$$+ \delta L(X_1;h_2)G(X_1,X_2)\delta L(X_2;h_1)\}\phi(X_2)dX_2 dX_1$$

$$= \int_{X_1}\int_{X_2} G(X_1,X_2)\{[\delta L^+(X_1;h_1)\phi^*(X_1) - \delta\Sigma_r(X_1;h_1)]\delta L(X_2;h_2)$$

$$+[\delta L^{+}(X_1;h_2)\phi^{*}(X_1) - \delta\Sigma_r(X_1;h_2)]\delta L(X_2;h_1)\}\phi(X_2)$$

$$dX_2 dX_1 \ . \tag{52}$$

Let $S(y)$ be the sensitivity vector and $H_{12}(y_1,y_2)$ be the Hessian of R at Σ in the directions of $h_1(y_1)$ and $h_2(y_2)$ with the following definitions:

$$\delta R(\Sigma;h) = \int_y S(y)\ h(y)\ dy \tag{63}$$

and

$$\delta^2 R(\Sigma;h_1,h_2) = \int_{y_1}\int_{y_2} H_{12}\ (y_1,y_2)h_1(y_1)h_2(y_2)dy_1dy_2, \tag{64}$$

where y, y_1, and y_2 are points in the phase space.

Clearly S and H are the analogy of the gradient vector and Hessian matrix, respectively, of the Euclidean space.

Comparing Eq. (60) and Eq. (63), we get

$$S(y) = \int_X \left[\frac{\partial\Sigma_r(X)}{\partial\Sigma(y)} - \phi^{*}(X)\ \frac{\partial L(X)}{\partial\Sigma(y)}\right]\phi(X)\ dX \ . \tag{65}$$

Similarly, from Eq. (62) and Eq. (64), we have

$$H_{12}(y_1,y_2) = \int_X \left[\Psi_1^{*}(X,y_1)\ \frac{\partial L(X)}{\partial\Sigma(y_2)} + \Psi_2^{*}(X,y_2)\ \frac{\partial L(X)}{\partial\Sigma(y_1)}\right]$$

$$\phi(X)dX, \tag{66}$$

where $\Psi_i^*(X,y_i)$ is the solution of the following adjoint transport equation:

$$L^+(X)\Psi_i^*(X,y_i) = \frac{\partial L^+(X)}{\partial \Sigma(y_i)}\,\phi^*(X) - \frac{\partial \Sigma_r(X)}{\partial \Sigma(y_i)} . \qquad (67)$$

b. Methodology of Nonlinear Optimization

Let us concentrate on the practical problem of finding the optimal distribution of nuclide densities that will maximize or minimize the response of a given detector. The system variables are the nuclide density distributions, $N = (N_1(r), N_2(r), \ldots, N_I(r))$ and the response is denoted by R. The problem is then

$$\text{Minimize } R(N) ,$$

$$\text{or Maximize } R(N) . \qquad (68)$$

Let h denote the variation of N. In general, R is a nonlinear functional of N. The quadratic approximation of R yields, by utilizing Eqs. (63) to Eq. (67):

$$\delta R(N) = R(N+h) - R(N)$$

$$= \sum_{i=1}^{I} \int_r S_i(r)h_i(r)\,dr$$

$$+ \frac{1}{2}\sum_{i,j=1}^{I} \int_r \int_{r'} H_{ij}(r,r')\,h_i(r)h_j(r')\,drdr' , \qquad (69)$$

where

$$S_i(r) = \left\langle \phi, \frac{\partial \Sigma_r}{\partial N_i(r)} \right\rangle - \left\langle \phi^*, \frac{\partial L}{\partial N_i(r)} \phi \right\rangle, \quad i=1,2,\ldots,I \tag{70}$$

$$H_{ij}(r,r') = \left\langle \Psi_i^*(X,r), \frac{\partial L}{\partial N_j(r')} \phi(X) \right\rangle$$

$$+ \left\langle \Psi_j^*(X,r'), \frac{\partial L}{\partial N_i(r)} \phi(X) \right\rangle, \quad i,j=1,\ldots,I \tag{71}$$

$\langle , \rangle$ denotes integration over the velocity space $(E,\underline{\Omega})$ and $\Psi_i^*(X,r)$ is the solution of the equation

$$L^+(X)\Psi_i^*(X,r) = \frac{\partial L^+(X)}{\partial N_i(r)} \phi^*(X) - \frac{\partial \Sigma_r(X)}{\partial N_i(r)} \quad . \tag{72}$$

If the condition is constraint-free, Eq. (69) is the basic equation for determining the optimal density variations $h_i(r)$, $i=1,\ldots,I$. The condition for the optimal density variations can be obtained by requiring

$$\partial(\delta R)/\partial h_i = 0,$$

Then

$$S_i(r) + \sum_{j=1}^{I} \int_{r'} H_{ij}(r,r')h_j(r')dr' = 0, i=1,\ldots,I \tag{73}$$

The solution of this set of equations gives the optimal density variations which eventually, by iterating the procedure, would bring the optimal density distribution that extremizes the response.

Let us pause for a moment and count the number of transport calculations required for each iteration step. From Eq. (73), we know that the availability of $S_i(r)$ and $H_{ij}(r,r')$ is a necessity for each iteration step. $S_i(r)$ is attainable from Eq. (70) which implies that two transport calculations for the forward and adjoint flux are inevitable in the linear approximation. It is a common practice to divide the spatial dependence into a certain number of spatial zones, say M, in the transport calculation. It is also clear that Eq. (71) and Eq. (72) account for the additional number of transport calculations which is IM. Hence, the advantage of the quadratic approach - being able to converge to the optimal solution more rapidly - could be jeopardized by the tremendous amount of additional effort required for obtaining the Hessian which, for most practical problems, is prohibitively expensive. Fortunately, some remedies are available to reduce the number of additional transport calculations and will definitely make the quadratic approach reasonably practical.

There is an inherent constraint on the variations of the nuclide densities, namely, the total volume occupied by all the nuclides of interest at any given location is kept constant despite

Using Eq. (75), the first-order term in Eq. (69) can be rearranged:

$$\sum_{i=1}^{I} \int_r S_i(r)h_i(r)\, dr$$

$$=\sum_{i=1}^{I-1} \int_r S_i(r)h_i(r)dr + \int_r S_I(r)h_I(r)dr$$

$$=\sum_{i=1}^{I-1} \int_r \left[S_i(r) - b_i(r)S_I(r)\right] h_i(r)dr$$

$$=\sum_{i=1}^{I-1} \int_r \hat{S}_i(r)h_i(r)dr, \qquad (77)$$

where, $\hat{S}_i(r)$ is called the substitution sensitivity vector and is defined as

$$\hat{S}_i(r) = S_i(r) - b_i(r)S_I(r)$$

$$= \left\langle \phi, \frac{\partial \hat{\Sigma}_r}{\partial N_i(r)} \right\rangle - \left\langle \phi^*, \frac{\partial \hat{L}}{\partial N_i(r)}\, \phi \right\rangle, \qquad (78)$$

where

$$\frac{\partial \hat{\Sigma}_r}{\partial N_i(r)} = \frac{\partial \Sigma_r}{\partial N_i(r)} - b_i(r)\, \frac{\partial \Sigma_r}{\partial N_I(r)} \qquad (79)$$

the density variations. Let $N_i^o(r)$ denote the maximum possible density for nuclide i at r, i.e., $N_i(r)=N_i^o(r)$ when $N_j(r)=0$ for all $j\neq i$. Also let $b_i(r)$ represent $N_I^o(r)/N_i^o(r)$. Then the constraint can be expressed as

$$\sum_{i=1}^{I} \frac{N_i(r)}{N_i^o(r)} = 1$$

or

$$N_I(r) + \sum_{i=1}^{I-1} b_i(r)N_i(r) = N_I^o(r), \tag{74}$$

which leads to

$$h_I(r) = -\sum_{i=1}^{I-1} b_i(r)h_i(r) \ . \tag{75}$$

Then our original unconstrained problem will have a new form:

$$\text{Min(Max)} \quad R(N)$$

$$\text{subject to} \quad \sum_{i=1}^{I} N_i(r)/N_i^o(r) = 1 \ . \tag{76}$$

It is a common practice in mathematical programming to reduce the number of independent variables in a problem with equality constraints by either direct substitution if possible or by the Lagrange multipliers technique[13]. Here we will use a direct substitution to eliminate the dependence on $N_I(r)$.

$$\frac{\partial \hat{L}}{\partial N_i(r)} = \frac{\partial L}{\partial N_i(r)} - b_i(r) \frac{\partial L}{\partial N_I(r)} \qquad (80)$$

$$i = 1, \ldots, I-1 .$$

Rearranging the second-order term is more difficult and tedious. The final result is shown as follows:

$$\sum_{i,j=1}^{I} \int_r \int_{r'} H_{ij}(r,r') h_i(r) h_j(r') dr\, dr'$$

$$= \sum_{i,j=1}^{I-1} \int_r \int_{r'} \hat{H}_{ij}(r,r') h_i(r) h_j(r') dr\, dr' \ , \qquad (81)$$

where $\hat{H}_{ij}$ is called the substitution Hessian and is defined as:

$$\hat{H}_{ij}(r,r') = \left\langle \hat{\Psi}_i^* \ , \frac{\partial \hat{L}}{\partial N_j} \phi \right\rangle + \left\langle \hat{\Psi}_j^* \ , \frac{\partial \hat{L}}{\partial N_i} \phi \right\rangle \ , \qquad (82)$$

where $\hat{\Psi}_i^* = \Psi_i^* - b_i \Psi_I^*$ and is the solution of the equation

$$L^+ \hat{\Psi}_i^*(X,r) = \frac{\partial \hat{L}^+(X)}{\partial N_i(r)} \phi^*(X) - \frac{\partial \hat{\Sigma}_r(X)}{\partial N_i(r)} \ , \quad i=1,\ldots,I-1. \qquad (83)$$

The quadratic approximation of R(N) with I-1 independent variables $N_1,\ldots,N_{I-1}$ is then expressed as

$$\delta R(N) = R(R + h) - R(N)$$

$$= \sum_{i=1}^{I-1} \int_r \hat{S}_i(r) h_i(r) dr$$

$$+ \frac{1}{2} \sum_{i,j=1}^{I-1} \int_r \int_{r'} \hat{H}_{ij}(r,r') h_i(r) h_j(r') dr\, dr' \quad . \qquad (84)$$

Then a similar optimum condition can be obtained in Eq. (73) with the replacement of S_i and H_{ij} by $\hat{S}_i$ and $\hat{H}_{ij}$, respectively. Notice that we have reduced the number of additional transport calculations from IM to (I-1) M in Eq. (83). For a system of many spatial zones, this is still not considered a practical approach.

A further improvement in reducing the number of transport calculations can be obtained by employing the concept of the optimal-gradient method in mathematical programming[13]. The gradient algorithm is by far the most widely applied method of solving optimization problems. The basic idea comes from a result of variational calculus which shows that the smallest distance between two points on a multidimensional surface results from following the gradient path. Consider a simple mathematical programming problem: $\min_X f(X)$ where f is a real-valued function and X is a n-dimensional vector. Let ∇f denote the gradient of f. The gradient algorithm suggests that, for p-th iteration step at X^P, the next point X^{P+1} can be chosen as

$$X^{P+1} = X^P + k \nabla f(X^P) \ . \qquad (85)$$

where $k > 0$ for maximization,

$k < 0$ for minimization.

If k is small, the discrete algorithm will closely approximate the gradient path, but convergence will be slow since the gradient must be calculated many times. On the other hand, if k is large, convergence will initially be very fast, but the algorithm will oscillate about the optimum.

The dilemma in selecting an adequate k for each iteration leads to the development of the so-called optimal-gradient method. For the p-th iteration step at X^P, the step size k is chosen by solving a one-dimensional minimization problem:

$$\min_k f(X^P + k \nabla f(X^P)),$$

that is, to find an optimal k* that will minimize f along the gradient direction at X^P. If $f(X^{P+1})$ can be known analytically as a function of k, then k* is the solution of $\partial f(X^{P+1})/\partial k = 0$. Otherwise, the one-dimensional search techniques, e.g., dichotomous search, equal-interval search, Fibonacci search, golden-section search, and random search, can be applied to search for the optimal step size k*. Unfortunately, every one-dimensional

search method needs to know the exact value of f(X) for all X and is simply not applicable to reactor systems since each f(X) evaluated at X requires exactly one transport calculation. However, the quadratic approximation for the response R(N) does provide a basis for analytically selecting an optimal step size along the gradient direction.

Let us turn our attention back to the original problem of extremizing R(N). Upon adopting the concept of the optimal-gradient method, we are able to define a step size a_i along the gradient direction $\hat{S}_i(r)$, i=1,...,I-1, such that for each step the variation in nuclide densities $h_i(r)$ is solely determined by

$$h_i(r) = a_i \hat{S}_i(r), \; i = 1,\ldots,I-1 \; . \tag{86}$$

The quadratic approximation in Eq. (84) therefore has the following expression

$$\delta R(a) = \sum_{i=1}^{I-1} a_i \int_r \hat{S}_i^2(r)\, dr + \frac{1}{2} \sum_{i,j=1}^{I-1} a_i a_j \int_r \int_{r'} \hat{H}_{ij}(r,r') \hat{S}_i(r) \hat{S}_j(r')\, dr dr' \tag{87}$$

The second order term in Eq. (87) can be further simplified by defining

$$\tilde{\Psi}_i^*(X) = \int_r \hat{\Psi}_i^*(X,r)\, \hat{S}_i(r)\, dr, \tag{88}$$

and therefore $\tilde{\Psi}_i^*(X)$ is the solution of the equation

$$L^+ \tilde{\Psi}_i^*(X) = \int_r \left[\frac{\partial \hat{L}^+(X)}{\partial N_i(r)} \phi^*(X) - \frac{\partial \hat{\Sigma}_r(X)}{N_i(r)} \right] \hat{S}_i(r) dr \tag{89}$$

Combining Eqs. (82), (88), and (89), we get

$$\frac{1}{2} \sum_{ij} a_i a_j \int_r \int_{r'} \hat{H}_{ij}(r,r')\, \hat{S}_i(r)\, \hat{S}_j(r')\, dr\, dr'$$

$$= \frac{1}{2} \sum_{ij} a_i a_j \int_r \int_{r'} \left\langle \hat{\Psi}_i^*(X,r), \frac{\partial \hat{L}}{\partial N_j(r')} \phi \right\rangle \hat{S}_i(r) \hat{S}_j(r') dr dr'$$

$$+ \frac{1}{2} \sum_{ij} a_i a_j \int_r \int_{r'} \left\langle \hat{\Psi}_j^*(X,r'), \frac{\partial \hat{L}}{\partial N_i(r)} \phi \right\rangle \hat{S}_i(r) \hat{S}_j(r') dr dr'$$

$$= \frac{1}{2} \sum_{ij} a_i a_j \int_{r'} \left\langle \tilde{\Psi}_i^*, \frac{\partial \hat{L}}{\partial N_j(r)} \phi \right\rangle \hat{S}_j(r') dr'$$

$$+ \frac{1}{2} \sum_{ij} a_i a_j \int_r \left\langle \tilde{\Psi}_j^*, \frac{\partial \hat{L}}{\partial N_i(r)} \phi \right\rangle \hat{S}_i(r)\, dr$$

$$= \sum a_i a_j \int_r \left\langle \tilde{\Psi}_i^* , \frac{\partial \hat{L}}{\partial N_j(r)} \phi \right\rangle \hat{S}_j(r) dr .$$

With the following definition of $\tilde{H}_{ij}$

$$\tilde{H}_{ij}(r) = 2 \left\langle \tilde{\Psi}_i^* , \frac{\partial \hat{L}}{\partial N_j(r)} \phi \right\rangle , \tag{90}$$

Eq. (87) can be formulated as

$$\delta R(a) = \sum_{i=1}^{I-1} a_i \int_r \hat{S}_i^2(r) dr$$

$$+ \frac{1}{2} \sum_{i,j=1}^{I-1} a_i a_j \int_r \tilde{H}_{ij}(r) \hat{S}_j(r) dr . \tag{91}$$

The condition for the optimal step size can be obtained by applying the equation $\partial(\delta R)/\partial a_i = 0$ to Eq. (91), then

$$\int \hat{S}_i^2(r) dr + \sum_{j=1}^{I-1} a_j^* \int_r \tilde{H}_{ij}(r) \hat{S}_j(r) dr = 0, i = 1, \ldots, I-1 \tag{92}$$

The solution of this system of I-1 equations gives a set of optimal step sizes $(a_1^*, \ldots, a_{I-1}^*)$ along the gradient direction $(\hat{S}_1, \ldots, \hat{S}_{I-1})$ for the optimization process which would most efficiently arrive at the optimal solution that extremizes the

detector response of interest. This approach has the most significant advantage in both efficient convergence to optimum and requiring a minimum number of transport calculations. The latter point can be verified by looking at Eq. (89) which shows that only I-1 additional transport calculations are needed instead of (I-1)M.

To illustrate the optimization procedure discussed above, we consider a two component problem, i.e., I=2. Thus, there is only one independent variable. From Eq. (92) we get

$$a_1^* = - \frac{\int \hat{S}_1^2(r)dr}{\int \tilde{H}_{11}(r)\hat{S}_1(r)dr} \quad . \tag{93}$$

Hence

$$h_1(r) = a_1\hat{S}_1(r) = - \frac{\hat{S}_1(r) \int \hat{S}_1^2(r)dr}{\int \tilde{H}_{11}(r)\hat{S}_1(r)dr} \quad , \tag{94}$$

and the variation of R in the quadratic approximation is

$$(\delta R)_{\text{quadratic}} = - \frac{\left[\int \hat{S}_1^2(r)dr\right]^2}{2\int \tilde{H}_{11}(r)\hat{S}_1(r)dr} \quad , \tag{95}$$

while the linear approximation yields

$$(\delta R)_{\text{linear}} = - \frac{\left[\int \hat{S}_1^2(r)dr\right]^2}{\int \tilde{H}_{11}(r)\hat{S}_1(r)dr} \quad , \tag{96}$$

i.e., $(\delta R)_{linear} = 2(\delta R)_{quadratic}$, which is merely a characteristic of the quadratic approximation as mentioned in the previous section.

The problem associated with the existence of inequality constraints in a nonlinear optimization is more difficult to handle and will not be addressed here. However, in the optimization problem we discussed, there are some inevitable inequality constraints. In particular the nuclide density distribution is expected to have some definite upper and lower limits such that $N_{min}(r) \leq N(r) \leq N_{max}(r)$. Theoretically, the inequality constraints can be converted into a set of necessary conditions by the Kuhn-Tucker theorem in an inequality-constrained problem. There are many gradient-related techniques which have been proposed to solve constrained optimization problems. In general they can be divided into two categories: (1) boundary-following methods, e.g., feasible-directions method and gradient-projection method, (2) penalty-function methods. A general discussion of these methods can be found in Reference 13.

C. Application of Optimization Method to NUWMAK Shield Design

In the previous section we outlined the procedure and mathematical formulation for a nonlinear optimization problem through the use of second-order sensitivity theory, i.e., the quadratic approximation. Historically, there is only one

attempt, made by Greenspan et al.[12], to determine the optimal density distribution for a mixture of iron and water as shielding materials in a typical radiation shielding problem. The shield is 100 cm thick and is optimized to provide minimum displacements per atom (dpa) at the back of the shield when a source of 14-MeV neutrons impinges on the front side. In this section, the optimization method associated with the quadratic approximation is applied to the inner shield of a conceptual tokamak fusion reactor design-NUWMAK.

The description of the overall system design, as well as the neutronics analysis of NUWMAK can be found elsewhere[14]. One peculiar characteristic of the NUWMAK design which directly influences the inner blanket and shield design is its compactness. The space from the plasma chamber first wall to the dewar of the superconducting magnet at the midplane is only 105 cm thick. In addition, there is 20 cm near the first wall to be reserved for the Li-Pb blanket so as to maintain minimum thermal cycling of the structure. Thus, a space of only 85 cm is actually available to provide adequate shielding for the protection of the superconducting magnet from the environment of intense radiation (~5 MW/m^2 neutron wall loading). In this regard, the use of metallic tungsten is essential to attenuate sufficiently the neutron and gamma radiations emerging from the blanket since

tungsten has a relatively high atomic density, 0.0629×10^{24} atoms/cm^3, considerably higher than most of the high Z shielding materials, (e.g., 0.0335×10^{24} atoms/cm^3 for lead). The schematic of a one-dimensional cylindrical model based on the plasma minor radius is shown in Fig. 2.a for the inner blanket/shield of NUWMAK. A lead zone of 5 cm near the magnet is retained for the purposes of reducing the gamma heating and dose rate in the magnet. In the actual NUWMAK design, a 40 cm-thick tungsten zone followed by a 40 cm-thick boron carbide zone is used in the $W+B_4C$ zone. The radiation damage results for this configuration are (1) the maximum atomic displacement rate in the aluminum stabilizer is $2x10^{-6}$ dpa/yr which necessitates periodic annealing for the resistivity change approximately every two years, and (2) the dose rate in the epoxy-based super-insulators is $3x10^{7}$ rad/yr. These should last the plant life.

The primary motive for an optimization study in the design of the NUWMAK inner shield is to find an optimal combination of W and B_4C throughout the 80 cm-thick $W+B_4C$ zone that will either minimize the dose rate or the dpa rate. The shield zone is divided into 8 intervals of 10 cm each which are numbered as intervals 1 to 8 with interval 8 being the one nearest to the Pb zone. Transport calculations, both forward and adjoint, are performed with a P_3S_4 approximation using the ANISN[15] code. A data

library with 25 neutron and 21 gamma groups is produced by group-collapsing of the DLC-41/VITAMIN-C library[16]. A second-order optimization computer code, OPTIMAL, is employed in the optimization process. The optimization code consists of three major parts: (1) A revised ANISN that will solve the usual forward and adjoint transport equations as well as the generalized adjoint transport equation (Eq. 89) where the adjoint distributed source is angularly dependent and may be negative, the latter could lead to negative fluxes and necessitates some modifications in flux convergency testing. (2) A sensitivity module which executes Equations (78) and (90). (3) A management module which handles the input and output and performs the actual iteration process.

From Eq. (89), it is realized that only one additional transport calculation is required for each iteration since there is only one independent variable, i.e., the spatial distribution nuclide density for either W or B_4C. The variable can also be converted into the volume percent occupied by either material at a particular interval. The starting point for the optimization process is the final design of the NUWMAK inner shield, i.e., 40 cm-thick tungsten followed by 40 cm-thick boron carbide. Table 2 presents the summary result of the detailed step-by-step process that will minimize either the dpa rate in the Al stabilizer or the dose rate in the epoxy-based superinsulator. A set of

inequality constraints is applied so that $0 \leq V_i \leq 100$, where V_i is the volume percent occupied by W in interval i(i=1,...,8). The R values in Table 2 are un-normalized. The conversion factors can be easily obtained by realizing that Step 1 represents the initial condition and therefore has the responses equivalent to a dpa rate of 2×10^{-6} dpa/yr and a dose rate of 3×10^{7} rad/yr. dR_1 and dR_2 represent the results calculated by Eqs. (96) and Eq. (95), respectively. dR_{true} is the difference in the response between two successive steps. Fig. 2.b shows how the response decreases as a function of iteration steps where dpa_{min} is the dpa rate for the last iteration step which gives the optimal combination of W and B_4C that minimizes the dpa rate. The case for dose rate is given in Fig. 2.c. The optimal distributions of the B_4C volume percentages for the conditions of minimum dpa and minimum dose are given in Fig.2.d, with the initial distribution shown for comparison.

Some observations can be made from the results of the optimization processes shown in Table 2 and Fig. 2. For the dpa as the response of interest, the dpa rate reduces to almost half of its original value after just one iteration step. It eventually shrinks to about one-fourth of the original dpa after completion of the optimization process. For the dose rate, the first iteration step also gives a reduction of one-half, although the subsequent iteration steps show little improvement on

minimizing the dose rate. The distribution of the W and B_4C combination that gives the smallest dpa rate clearly demonstrates the dominance of tungsten due to its effective shielding ability for high energy neutrons through the (n,2n) and inelastic scattering reactions. Nonetheless, the presence of B_4C near the magnet is essential for absorbing low energy neutrons which also have significant impact on the total dpa. As a matter of fact, the dpa rate, in relative units, for a shield consisting entirely of tungsten is 4.8 which is larger than that of the original design where the relative dpa is 3.7. As far as the dose rate is concerned, tungsten has less effect than in the case of the dpa rate and the importance of B_4C increases accordingly. Recall that the total dose rate is the combined effects of neutron and gamma radiation. For the initial material composition the total dose of 1.443 is the sum of 1.137 and 0.306 coming from neutrons and photons, respectively. After one iteration step, which raises the tungsten volume percentages from 0 to 30-40% for intervals 5 to 8, the gamma dose reduces to an almost negligible value of 0.05 while the neutron dose is 0.68. Thus, it is reasonable to suspect that the original shield combination, 40 cm W followed by 40 cm B_4C, does not have enough high Z materials in the back of the shield to absorb the gamma radiation, even though the 5 cm-thick Pb zone near the magnet has been retained.

An important conclusion can be drawn from Fig. 2.d. For a combination of W and B_4C as a shielding material, the optimal combination that will minimize the radiation damage parameters in the superconducting magnet may vary from one response to the other. In the practical design of reactor shielding, the optimization process should apply to the most crucial response which is the limiting factor for the reactor performance. In addition, the resources problem and economic considerations, such as the high cost of tungsten, should also be taken into account as one of the constraints in the shielding design process.

There is one significant and intriguing value of the quadratic optimization method that is worth mentioning. The optimal-gradient method enables us to find the optimal spatial distribution of the material densities (or, equivalently, the volume fractions) with only one additional transport calculation per iteration step for a 2-materials problem. On the other hand, most of the perturbation-related and variational methods can only deal with a certain number of variables (for instance, the zone thickness of W in the W+B_4C system) and obtaining a detailed spatial distribution is by no means an easy task.

There are some restrictions imposed on the application of the optimization procedure discussed above. First, the choice of the starting point is of prime importance in solving the nonlinear

optimization problem. An ill-chosen initial guess may lead to a local maxima or minima instead of the global one. Thus, a good understanding of the problem is essential so that the best educated guess can be taken before proceeding to the actual optimization process. Secondly, the response must be a slowly-varying functional of the variables and saddle points should not exist, in order to converge to the optimal solution efficiently. In principle a truly quadratic function only needs one step to arrive at the optimal point. Thus, for a functional that is approximately quadratic the convergency rate is expected to be very fast. As for a saddle point where the second-order derivative vanishes, the step size is infinite and some precautionary steps must be taken to avoid any confusion in computer programming. Finally, the constraints, equality or inequality, cannot be too complicated, otherwise the simple formulation discussed in Section B would not be applicable and the nonlinear optimization problem may become too difficult to deal with.

The optimization process for the W and B_4C shielding system presented in this section can be generalized to a system comprising other candidate shielding materials, for instance, a stainless steel and B_4C system, an iron and water system, a tungsten, B_4C, and lead system, etc. One potential application of the quadratic optimization methodology is in the design of the

fusion-fission hybrid systems where the spatial distributions of structural materials, fertile materials, fissile materials, reflectors, and neutron multipliers can be optimized to achieve the design goal for a specific hybrid system.

References

1. A. Gandini, J. Nucl. Energy, Parts A/B, 21, 755 (1967).

2. W.M. Stacey, Jr., Nucl. Sci. Eng., 48, 444 (1972).

3. T.J. Hoffman, Nucl. Sci. Eng., 50, 300 (1973).

4. E.M. Oblow, Nucl. Sci. Eng., 59, 187 (1976); see also, Nucl. Sci. Eng., 65, 428 (1978).

5. C. Goffman and G. Pedrick, First Course in Functional Analysis, Prentice-Hall, New Jersey (1969).

6. J.T. Oden, Applied Functional Analysis, Prentice-Hall, New Jersey (1979).

7. L.B. Rall, ed., Nonlinear Functional Analysis and Applications, Academic Press, New York (1971).

8. A. Dubi and D. J. Dudziak, "Higher-order Terms in Sensitivity Analysis via a Differential Approach," paper submitted to Nucl. Sci. Eng. (1979).

9. T.E. Albert, Trans. Am. Nucl. Soc., 17, 546 (1973).

10. M.A. Abdou, "Fusion Reactor Shielding," paper presented at the IAEA Advisory Group Meeting on Fusion Reactor Technology, Vienna, Dec. 11-15, 1978.

11. T. Wu and C.W. Maynard, "The Application of Uncertainty Analysis in Conceptual Fusion Reactor Design," Proc. Seminar/Workshop on Theory and Application of Sensitivity and Uncertainty Analysis, August 22-24, 1978. ORNL/RSIC-42, pp. 191-217, Oak Ridge National Laboratory (1978).

12. E. Greenspan, D. Gilai, and E.M. Oblow, Nucl. Sci. Eng., 68, 1 (1978).

13. D.A. Wismer and R. Chattergy, Introduction to Nonlinear Optimization, North-Holland Inc., New York (1978).

14. B. Badger et al., "NUWMAK: A Tokamak Reactor Design Study," UWFDM-330, Fusion Engineering Program, Nuclear Engineering Department, University of Wisconsin-Madison (1979).

15. W.W. Engle Jr., "A User's Manual for ANISN," K-1693, Oak Ridge Gaseous Diffusion Plant (1967).

16. RSIC Data Library Collection, DLC-41/VITAMIN-C Radiation Shielding Information Center, Oak Ridge National Laboratory (1977).

Table 1

Percent Difference in Predicted Response between N'th-order Sensitivity Result and True Result, $(R'_N - R'_{true})/R'_{true} \times 100\%$

	Order of Approximation									
ah	1	2	3	4	5	6	7	8	9	10
.2	-2									
.4	-10	1								
.6	-27	6								
.8	-56	16	-3							
1.0	-100	36	-9	2						
1.2	-166	73	-23	6	-1					
1.4	-262	135	-50	15	-4					
1.6	-397	237	-101	34	-9	2				
1.8	-584	396	-192	73	-23	6	-1			
2.0	-839	639	-346	146	-51	15	-4			
2.2	-1180	1000	-601	280	-107	35	-10	3		
2.4	-1640	1530	-1010	516	-216	77	-24	7	-2	
2.6	-2250	2300	-1650	916	-417	161	-54	16	-4	1
2.8	-3060	3390	-2630	1580	-777	323	-117	39	-11	3
3.0	-4120	4920	-4120	2660	-1410	628	-243	83	-26	7

Table 2

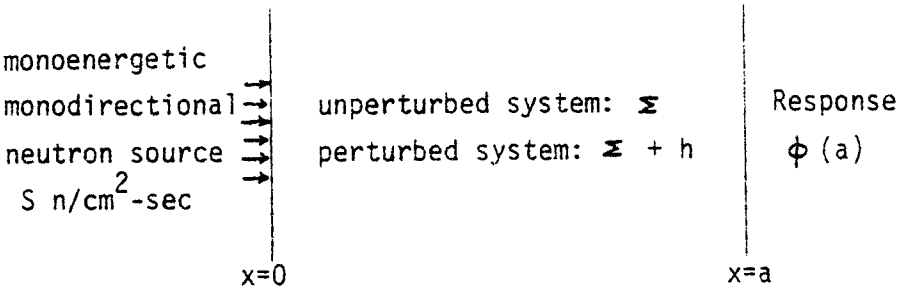
Summary of Optimization Process for (a) Dpa in Al Stabilizer and (b) Dose in Epoxy Superinsulator for NUWMAK Inner Shield.

(a) dpa

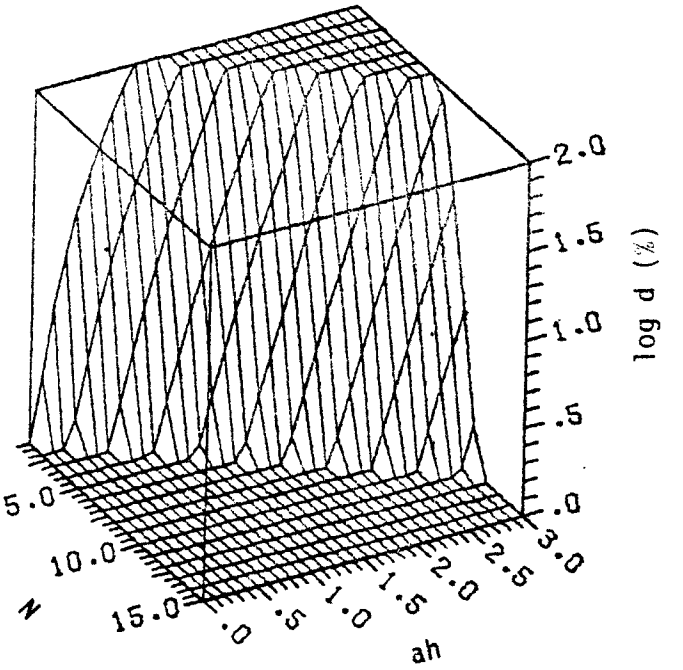
Iteration	V_1	V_2	V_3	V_4	V_5	V_6	V_7	V_8	R	dR_1	dR_2	dR_{true}
1	100	100	100	100	0	0	0	0	13.5	-5.09	-2.55	-6.09
2	100	100	100	100	17	19	19	14	7.41	-3.18	-1.59	-2.24
3	100	100	100	100	39	42	40	29	5.17	-1.83	-0.92	-1.07
4	100	100	100	100	68	70	63	37	4.10	-0.65	-0.32	-0.43
5	100	100	100	100	92	85	65	6	3.67	-0.14	-0.07	-0.05
6	100	100	100	100	100	92	68	0	3.62			
(b) dose												
1	100	100	100	100	0	0	0	0	1.443	-2.582	-1.291	- .714
2	100	98	97	99	31	29	38	32	.729	- .087	- .043	- .011
3	100	99	98	100	32	40	39	31	.718	- .085	- .043	- .056
4	100	100	100	100	50	56	48	11	.663	- .014	- .007	- .004
5	100	100	99	98	53	57	46	0	.658			

Note: (1) V_i denotes percent volume occupied by tungsten in iterval i.

(2) All values of R, dR_1, dR_2, and dR_{true} are un-normalized numbers.



(a) A System of Simple Neutron Attenuation in a Slab of Pure Absorber.

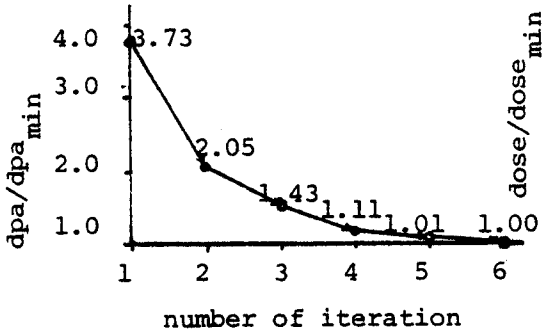


(b) Graphical Display of d, $(R'_N - R'_{true})/R'_{true}$, as a Function of N (Order of Sensitivity) and ah (Cross Section Perturbation) for a Simple Neutron Attenuation System.

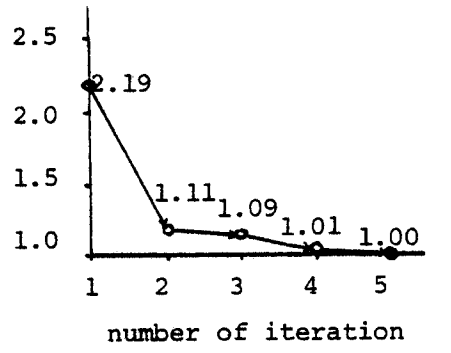
Fig. 1

material	plasma	vacuum	Li-Pb	W + B_4C	Pb	magnet
thickness (cm)	113	15	20	80	5	
radius (cm)	0	113	128	148	228	233

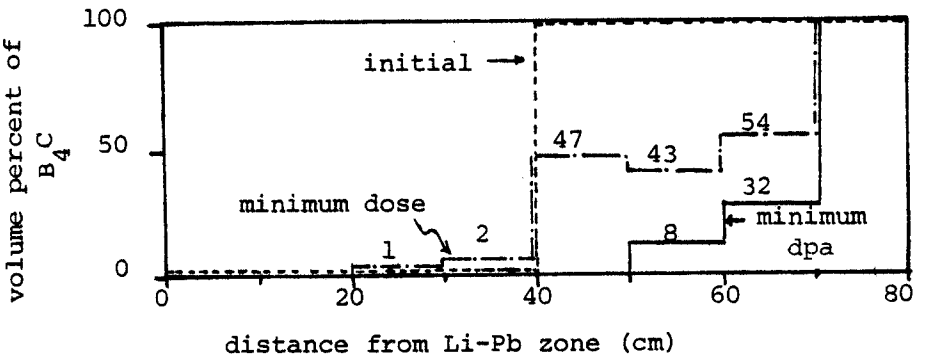
(a) Schematic of a one-dimensional cylindrical model for NUWMAK inner blanket/shield.



(b) dpa/dpa_{min} vs. iteration



(c) $dose/dose_{min}$ vs. iteration



(d) Spatial distribution of volume percent occupied by B_4C for initial, minimum dpa, and minimum dose conditions.

Fig. 2