

A PERTURBATION THEORY APPROACH TO NON-LINEAR PROGRAMMING

James Thurber
Andrew Whinston
Purdue University
West Lafayette, IN 47907/USA

I. Perturbation Theory and Its Applications

The majority of problems which occur in mathematical physics are stated in the form of differential equations with some prescribed boundary and/or initial conditions. Generally, it is impossible to obtain closed form analytical solutions to such problems. A long standing method for obtaining approximate solutions is perturbation theory. One or more parameters enter the problem and for some fixed values of the parameters (usually taken to be zero), the problem is exactly solvable analytically. The equations and their solutions are expanded in terms of these parameters and one thereby obtains approximate solutions of the original problem.

We apply this method to non-linear mathematical programming, in order to obtain a sequence of approximate solutions which converges to solutions of the original problem. As a corollary to this approach, we are able to derive sufficient and/or necessary conditions for local optimum solutions in a very general setting. In particular, when the normal first order conditions fail due to the presence of cusps, we derive a new set of sufficient conditions.

The perturbation series approach leads to a new algorithmic development which is discussed in this paper and will be further amplified in a sequel.

II. Direct Expansion

The problem to be considered in this paper is:

$$\text{Minimize } F = F(\mathbb{Z}) \quad (1)$$

Subject to the constraints

$$G_\ell(\mathbb{Z}) = 0, \quad 1 \leq \ell \leq L \quad (2)$$

where $\mathbb{Z} = (\mathbb{Z}_1, \dots, \mathbb{Z}_R)$ is a vector with real components, and $F(\mathbb{Z})$ and the $G_\ell(\mathbb{Z})$ are real holomorphic functions in the components of \mathbb{Z} .

We study the problem in a neighborhood of a particular point

$$\mathbb{Z}_0 = (\mathbb{Z}_1^0, \mathbb{Z}_2^0, \dots, \mathbb{Z}_R^0) \quad (3)$$

and on feasible paths in this neighborhood passing through this point. These paths are constructed by means of a perturbation series to be described. In the neighborhood of \mathbb{Z}_0 , F and G_ℓ , $1 \leq \ell \leq L$ have power series expansions with non-zero radii of

of convergence. For a justification of the form of the perturbation expansion which we assume in this section, see reference 6.

In order to give a geometric interpretation of the terms in the perturbation series, it will be convenient to express the Taylor expansions of the constraint functions in terms of tensor notation. For our purposes, we will regard the tensor notation as an extension of matrix notation. In matrix notation, we have two indices i and j to denote the element m_{ij} in the i -th row and the j -th column of the matrix M . A matrix is a special case of a tensor. For tensors, we allow an arbitrary number of indices. A tensor of valence V will have V indices. Thus, a component of a tensor T of valence V will be denoted by t_{j_1, j_2, \dots, j_V} . Tensors of the same valence where the corresponding indices have the same range of values form a linear vector space, and they can be linearly combined to give tensors of the same type. We will consider such tensor spaces for complex as well as real valued components. Temporarily, we will only consider tensors where the indices j all have the same range. Below, it is assumed that this is always the case.

A tensor with exactly one index is a vector. A tensor with no indices is referred to as a scalar or simply a number. As is the case with matrices and vectors, tensors can also be multiplied. Two types of multiplication will concern us, namely tensor multiplication and contraction. The product P of two tensors S and T with components s_{j_1, \dots, j_r} and t_{i_1, \dots, i_u} respectively is the tensor with components

$$P_{j_1, \dots, j_r i_1, \dots, i_u} = s_{j_1, \dots, j_r} t_{i_1, \dots, i_u} \quad (4)$$

Multiplication of tensors is generally non-commutative; for example,

$$P_{12} = S_1 T_2 \neq S_2 T_1 = P_{21} \quad (5)$$

Tensor multiplication of tensors is, however, associative

$$R(ST) = (RS)T \quad (6)$$

and distributive

$$R(S + T) = RS + RT \quad (7)$$

$$(S + T)R = SR + TR \quad (8)$$

Contraction consists of tensor multiplication followed by a summation over pairs of indices which are equal; one member of each pair is taken from the two terms in the product. For example, suppose that the indices have the range $1 \leq j \leq 3$. T is of the form t_{ij} , and S is of the form s_{kl} ; then the tensor product ST of the form $t_{ij} s_{kl}$ can be contracted on the pair of indices j and k to give a tensor with the components

$$\sum_{j=1}^3 t_{ij} s_{j\ell}$$

where here, k is equated to j , and then j is summed. This is of course just equivalent to ordinary matrix multiplication of S and T , where tensors of valence two are treated as ordinary matrices. On the other hand, we could consider S to be of the form

$$S_{k\ell m}$$

and T to be of the form

$$t_{ijnq} ;$$

then ST is of the form $S_{k\ell m} t_{ijnq}$, and we could contract the indices ℓ and q to give a tensor of the form

$$U_{kmijn} = \sum_{\ell=1}^3 S_{k\ell m} t_{ijn\ell} . \quad (9)$$

We could also have contracted S and T on the two pairs of indices ℓ and q as well as k and n to give

$$V_{mij} = \sum_{\ell=1}^3 \sum_{k=1}^3 S_{k\ell m} t_{ijk\ell} . \quad (10)$$

Tensor multiplication of S and T followed by contraction on n pairs of indices will be denoted by

$$S \cdot^n T .$$

In the above examples, we could write

$$U = S \cdot^1 T \quad \text{and} \quad (11)$$

$$V = S \cdot^2 T . \quad (12)$$

In tensor analysis, one usually has two kinds of indices - superscripts as well as subscripts, and contraction occurs between indices of different type (i.e., super and sub-scripted indices); however, we will deal only with the case of so-called Cartesian tensors and the distinction will not be necessary for our purposes. We will find it useful also to regard contraction as represented by ordinary matrix multiplication. For example, in the tensor $S_{k\ell m}$, we can regard the pair of indices $k\ell$ as corresponding to the 3^2 components of a nine dimensional vector space and similarly for the indices nq of t_{ijnq} . From this point of view, we can regard S as a 9×3 matrix and T as a 9×9 matrix, and the 3×9 matrix V is formed by ordinary matrix multiplication from S and T . Thus, $V = S \cdot^t T$ where the superscript t denotes the transpose and the dot, ordinary matrix multiplication.

In the following, we will find both the tensor and matrix interpretations of contraction useful. We will now relax the restriction that the indices of tensors all have the same range; however, we can only contract on pairs of indices which have the same range of values.

One can extend the idea of tensors to tensor operators; thus the gradient operator $\nabla = \left(\frac{\partial}{\partial z_1}, \frac{\partial}{\partial z_2}, \dots, \frac{\partial}{\partial z_R} \right)$ can be regarded as a vector (tensor of valence one) operator with components $\frac{\partial}{\partial z_j}$ for $1 \leq j \leq R$; and ∇^2 , the tensor product of ∇ with itself is a partial differential operator with the components

$$\frac{\partial^2}{\partial z_i \partial z_j} \text{ for } 1 \leq i, j \leq R; \quad (13)$$

and similarly for ∇^3 , and so on.

If in ∇^2 , we equate the indices i and j and contract, we have

$$\nabla \cdot \nabla = \sum_{i=1}^R \frac{\partial^2}{\partial z_i^2} \quad (14)$$

which is the well known Laplace operator in R variables.

In our use of tensors, mostly symmetric tensors will occur. A tensor T of valence u is said to be symmetric if for every component t_{i_1, \dots, i_u} of T , we have

$$t_{i_1, \dots, i_u} = t_{i'_1, \dots, i'_u} \quad (15)$$

where the indices i'_1, \dots, i'_u are an arbitrary permutation of the indices i_1, \dots, i_u .

Basically the reason that mainly symmetric tensors are of interest to us is that in the Taylor Expansions of analytic functions to which the tensors will be applied, we know the order in which partial derivatives or more generally, directional derivatives are applied, is immaterial. We rearrange the order, and the result is the same.

Clearly sums and linear combinations of symmetric tensors are symmetric; however, the product of two symmetric tensors is in general, not symmetric. Thus we need the concept of a permutation operator acting on a tensor. A permutation operator Ω acts on a tensor T by permuting the indices of T to give another tensor \tilde{T} .

For example, let T be the tensor of valence four having components t_{ijkl} , and let Ω be the permutation operator

$$\Omega = \begin{pmatrix} 2143 \\ 1234 \end{pmatrix}. \quad (16)$$

That is, Ω takes the first index i and puts it in the place of the second index j . The second index j is put in the place of the first index i . The third index k is put in the place of the fourth index l , and the fourth index l is put in the place of the third index k .

Clearly there are $24 = 4!$ different such permutation operators Ω for tensors of valence four.

The symmetrized tensor \hat{T} corresponding to a tensor T of valence n is defined to be

$$\hat{T} = \frac{1}{n!} \left(\sum_{k=1}^{n!} \Omega_k T \right), \quad (17)$$

where the Ω_k are the $n!$ different permutation operators defined on tensors of valence n .

If T is a symmetric tensor, then clearly

$$\hat{T} = T. \quad (18)$$

In the application to the Taylor expansions of the constraint functions, the tensors which are tensor products of the gradient operator are clearly symmetric tensor operators. These tensors V^n are applied to the vector of constraint functions G at a particular point $Z = Z_0$. The resultant tensor $V_0^n G$ will be symmetric in all the indices arising from V^n but not from the index labeling the components of G . We will be interested in the left null spaces of the operators $V_0^n G$ and their orthogonal complementary spaces, since it is from these considerations that we will be able to scale the variables properly in terms of the perturbation parameter ϵ . The left indices are those corresponding to V^n , and the right indices are those belonging to G . From the symmetry in the left indices, we see that the projection operators π_n defining the left null spaces and the operators C_n giving the projection operators for the corresponding orthogonal complementary spaces must be symmetric in these same indices which are in fact the only indices occurring in the definition of the operators.

π_n and C_n which satisfy the relations

$$\pi_n \pi_n = \pi_n \quad (19)$$

$$C_n C_n = C_n \quad (20)$$

$$\pi_n C_n = C_n \pi_n = 0 \text{ (the 0-operator)}. \quad (21)$$

From the symmetry in the indices of π_n and C_n , we have the following useful property, namely, if $T_1 T_2 \dots T_k$ is a product of tensors lying in the domains of π_n and C_n , then a total contraction of all the indices of the product $T_1 \dots T_k$ $\pi_n T_1 T_2 \dots T_k$ with the corresponding indices of π_n gives a symmetric tensor, and similarly for $C_n T_1 T_2 \dots T_k$, and consequently

$$\pi_n T_1 T_2 \dots T_k = \pi_n T_{1'} T_{2'} \dots T_{k'} \quad (22)$$

where $1', 2', \dots, k'$ is an arbitrary permutation of $1, 2, \dots, k$, and similarly,

$$C_{n_1} T_1, \dots, T_k = C_{n_1} T_1, T_2, \dots, T_k, \dots \quad (23)$$

The functions defining the constraints are written as a row vector

$$G(\mathbb{Z}) = (G_1(\mathbb{Z}), G_2(\mathbb{Z}), \dots, G_L(\mathbb{Z})), \quad (24)$$

and the Taylor Expansions take the form

$$G(\mathbb{Z}_0 + \Delta\mathbb{Z}) = \sum_{n=0}^{\infty} \frac{(\Delta\mathbb{Z})^n n \cdot \nabla_0^n G}{n!} \quad (25)$$

where the following notational conventions are used:

$$a) \Delta\mathbb{Z} = (\Delta\mathbb{Z}_1, \Delta\mathbb{Z}_2, \dots, \Delta\mathbb{Z}_R) \quad (26)$$

and by $(\Delta\mathbb{Z})^n$, one means the n-fold tensor product; thus $(\Delta\mathbb{Z})^2$ is an $R \times R$ matrix with i, j entry equal to $\Delta\mathbb{Z}_i \Delta\mathbb{Z}_j$. We set $(\Delta\mathbb{Z})^0 = 1$ by definition. The term $(\Delta\mathbb{Z})^3$ would be a triply indexed array with i, j, k th entry equal to $\Delta\mathbb{Z}_i \Delta\mathbb{Z}_j \Delta\mathbb{Z}_k$, etc.

b) ∇ is the gradient operator written as a column vector. By ∇^n , again one means the tensor product, and by $\nabla_0^n G$, one means the n-fold gradient operator applied to G at the value $\mathbb{Z} = \mathbb{Z}_0$.

c) $n \cdot$ refers to an n-fold contraction of the indices of $(\Delta\mathbb{Z})^n$ and ∇^n . Thus

$$\frac{(\Delta\mathbb{Z})^n n \cdot \nabla^n}{n!} \quad (27)$$

is the usual scalar operator giving the nth order terms in the Taylor Expansions.

This representation of the Taylor Expansions enables us to describe in a simple fashion the perturbation expansion.

In the perturbation expansion, a perturbation parameter ϵ occurs, and the form of the expansion is a Puiseux series in ϵ . That is, the series consists of terms which are fractional powers of ϵ , namely terms of the form ϵ^α where α is a non-negative rational number. In fact, for our consideration, α will always be of the form $\frac{P}{N}$, where N is some fixed positive integer, and P is some non-negative integer. For practical calculations where in general, only a finite number of terms of the perturbation expansion are considered, ϵ is effectively a scaling parameter which determines the percent error we are willing to tolerate in satisfying the constraints. It will be convenient to regard $\nabla_0^n G$ as an R^n by L matrix, and of particular interest are the left null spaces of $\nabla_0^n G$ and their orthogonal complements. Our problem now is to determine the form of the solution to the constraints (2) locally. By assumption, the expansion converges for

$$|\Delta\mathbb{Z}_i| \leq r > 0 \quad (28)$$

for some radius of convergence r . We are interested in the $\Delta\mathbb{Z}$ such that $G(\mathbb{Z}_0) = 0$, and $G(\mathbb{Z}_0 + \Delta\mathbb{Z}) = 0$. We will justify later the expansion

$$\Delta\mathbb{Z} = \sum_{j=1}^{\infty} \epsilon^{\frac{P_j}{N}} \tilde{\mathbb{Z}}^j \quad (\text{Note: the } j \text{ in } \tilde{\mathbb{Z}}^j \text{ is a superscript or index and not an exponent}) \quad (29)$$

which we now assume to be valid.

The first operator or matrix of interest is the $V_O^1 G$ and the corresponding term in the Taylor Expansion is $\Delta Z \cdot 1 \cdot V_O^1 G$. For ΔZ having a direction belonging to the left null space of $V_O^1 G$, there is no contribution to the linear term in the Taylor Expansion. Thus, for an arbitrary ΔZ , the only contribution to the linear term is the orthogonal projection of ΔZ onto the orthogonal complement of the left null space of $V_O^1 G$.

Let π_1 denote the projection operator for the left null space of $V_O^1 G$. Then, $\pi_1 \Delta Z$ does not occur in the linear approximations to the constraints. Consider next the operator

$$\frac{V_O^2 G}{2!} \quad (30)$$

and the tensor product $(\Delta Z)^2$ written in the form:

$$(\Delta Z)^2 = (C_1 \Delta Z)^2 + (C_1 \Delta Z)(\pi_1 \Delta Z) + (\pi_1 \Delta Z)(C_1 \Delta Z) + (\pi_1 \Delta Z)^2 \quad (31)$$

where $C_1 = 1 - \pi_1$. (32)

If $\pi_2 (\pi_1 \Delta Z)^2 = (\pi_1 \Delta Z)^2$, then $\pi_1 \Delta Z$ as well as belonging to the left null space of $V_O^1 G$ also is such that its tensor product with itself belongs to the left null space of $\frac{V_O^2 G}{2!}$ and clearly does not contribute to leading order to the quadratic term in the Taylor Expansion.

For $\Delta Z \neq 0$, more generally, let us assume for $k \geq 2$, that

$$\begin{aligned} C_1 \Delta Z &= 0, \text{ hence } \pi_1 \Delta Z \neq 0 \\ C_2 (\Delta Z)^2 &= 0 \\ &\vdots \\ C_{k-1} (\Delta Z)^{k-1} &= 0; \end{aligned} \quad (33)$$

but $C_k (\Delta Z)^k \neq 0$;

then we will regard ΔZ as a term of order $\epsilon^{1/k}$ where ϵ is a small perturbation parameter in terms of which we scale our expansion which will give us an expansion for solving the constraints at regular points or in the neighborhood of cusps.

To see how the expansion works, we proceed as follows. We begin with the linear term in the Taylor expansion, namely

$$(\Delta Z) \cdot 1 \cdot V_O^1 G = (C_1 \Delta Z + \pi_1 \Delta Z) \cdot 1 \cdot V_O^1 G \quad (34)$$

$$= C_1 \Delta^{\mathbb{Z}} 1 \cdot V_0^1 G$$

$C_1 \Delta^{\mathbb{Z}}$ we take to be of order ϵ .

Since $\pi_1 \Delta^{\mathbb{Z}} 1 \cdot V_0^1 G = 0$, $\pi_1 \Delta^{\mathbb{Z}}$ will occur in the Taylor expansion no earlier than the tensor product of $\Delta^{\mathbb{Z}}$ with itself, we have

$$(\Delta^{\mathbb{Z}})^2 = (C_1 \Delta^{\mathbb{Z}})^2 + (C_1 \Delta^{\mathbb{Z}})(\pi_1 \Delta^{\mathbb{Z}}) + (\pi_1 \Delta^{\mathbb{Z}})(C_1 \Delta^{\mathbb{Z}}) + (\pi_1 \Delta^{\mathbb{Z}})^2. \quad (35)$$

The only part of $(\Delta^{\mathbb{Z}})^2$ which contributes to the second degree terms in the Taylor expansion, namely

$$\frac{(\Delta^{\mathbb{Z}})^2 2 \cdot V_0^2 G}{2!}$$

is $C_2 (\Delta^{\mathbb{Z}})^2$. Since C_2 is symmetric in its indices,

$$C_2 (C_1 \Delta^{\mathbb{Z}})(\pi_1 \Delta^{\mathbb{Z}}) = C_2 (\pi_1 \Delta^{\mathbb{Z}})(C_1 \Delta^{\mathbb{Z}}); \quad (36)$$

thus

$$C_2 (\Delta^{\mathbb{Z}})^2 = C_2 (C_1 \Delta^{\mathbb{Z}})^2 + 2C_2 (C_1 \Delta^{\mathbb{Z}})(\pi_1 \Delta^{\mathbb{Z}}) + C_2 (\pi_1 \Delta^{\mathbb{Z}})^2. \quad (37)$$

We wish to determine those terms of $C_2 (\Delta^{\mathbb{Z}})^2$ which are of order ϵ . Clearly $C_2 (C_1 \Delta^{\mathbb{Z}})^2$ is of order ϵ^2 , and to a first approximation, we can for small ϵ , neglect this contribution to $C_2 (\Delta^{\mathbb{Z}})^2$.

It is reasonable to choose a vector of the form $\pi_1 \Delta^{\mathbb{Z}}$ to be of order $\epsilon^{\frac{1}{2}}$ for then $C_2 (\pi_1 \Delta^{\mathbb{Z}})^2$ will be of order ϵ . If $C_2 (\pi_1 \Delta^{\mathbb{Z}})^2 \neq 0$, we make this choice. The following difficulty arises if two vectors $\Delta^{\mathbb{Z}}_1$ and $\Delta^{\mathbb{Z}}_2$ are such that

$$C_2 (\pi_1 \Delta^{\mathbb{Z}}_1)^2 \neq 0 \text{ and } C_2 (\pi_1 \Delta^{\mathbb{Z}}_2)^2 \neq 0; \quad (38)$$

we cannot in general assert that $C_2 \{ \pi_1 (\alpha \Delta^{\mathbb{Z}}_1 + \beta \Delta^{\mathbb{Z}}_2) \}^2 \neq 0$ for arbitrary α and β , so we proceed as follows. For $\Delta^{\mathbb{Z}}_1$ such that

$$\pi_1 \Delta^{\mathbb{Z}}_1 = \Delta^{\mathbb{Z}}_1 \quad (39)$$

and

$$C_2 (\pi_1 \Delta^{\mathbb{Z}}_1)^2 \neq 0, \quad (40)$$

choose $\Delta^{\mathbb{Z}}_1$ to be of order $\epsilon^{\frac{1}{2}}$.

Next choose $\Delta^{\mathbb{Z}}_2$ to be linearly independent of $\Delta^{\mathbb{Z}}_1$ and such that

$$C_2(\pi_1 \Delta Z_2)^2 \neq 0 \quad (41)$$

if possible. Then choose ΔZ_3 to be linearly independent of ΔZ_1 and ΔZ_2 such that

$$\pi_1 \Delta Z_3 = \Delta Z_3 \quad (42)$$

and such that

$$C_2(\pi_1 \Delta Z_3)^2 \neq 0 \quad (43)$$

if possible and continue as far as possible obtaining linearly independent

$$\Delta Z_1, \Delta Z_2, \dots, \Delta Z_n$$

such that

$$\pi_1 \Delta Z_k = \Delta Z_k \quad (44)$$

and

$$C_2(\pi_1 \Delta Z_k)^2 \neq 0 \text{ for } 1 \leq k \leq n. \quad (45)$$

Then ΔZ_k for $1 \leq k \leq n$ are taken to be of order $\epsilon^{\frac{1}{2}}$.

The vectors in the subspace determined by C_1 and $\Delta Z_1, \dots, \Delta Z_n$ may not generate the full vector space of possible ΔZ . If they do, then we are in a position to assign a unique order ϵ^α to each vector; if not, then there must exist vectors $\Delta Z \neq 0$ such that

$$\pi_1 \Delta Z = \Delta Z, \quad (46)$$

$$C_2(\pi_1 \Delta Z)^2 = 0, \quad (47)$$

or equivalently

$$\pi_2(\pi_1 \Delta Z)^2 = (\pi_1 \Delta Z)^2. \quad (48)$$

We must determine the order of such a ΔZ . Such a ΔZ might satisfy the relation

$$C_2(\Delta Z)W \neq 0 \quad (49)$$

where W is in the subspace generated by $\Delta Z_1, \Delta Z_2, \dots, \Delta Z_n$. If this is so, let us consider vectors H of the form

$$H = \lambda \Delta Z + W \quad (50)$$

for arbitrary W and scalars λ .

By assumption,

$$\pi_1 H = \lambda \pi_1 \Delta Z + \pi_1 W = \lambda \Delta Z + W = H, \quad (51)$$

and therefore,

$$C_2(\pi_1 H)^2 = \lambda^2 C_2(\Delta Z)^2 + 2\lambda C_2(\Delta Z)W + C_2 W^2 \quad (52)$$

using the symmetry in the indices of C_2 .

Since $C_2(\Delta Z)^2 = 0$, and $C_2 \Delta Z W \neq 0$, we have

$$C_2(\pi_1 H)^2 = 2\lambda C_2(\Delta Z)W + C_2 W^2 \neq 0 \quad (53)$$

for λ sufficiently large.

But $\pi_1 H = H$ is linearly independent of $\Delta Z_1, \Delta Z_2, \dots, \Delta Z_n$, and thus we could have chosen $H = \Delta Z_{n+1}$, a contradiction. Another possibility is that ΔZ might satisfy a relation of the form

$$C_2(\Delta Z)u \neq 0 \quad (54)$$

where

$$\pi_1 u = u. \quad (55)$$

Such a u is of order ϵ , and unless the order of ΔZ is chosen to be $\epsilon^0 = 1$, we cannot have $C_2(\Delta Z)u$ of order ϵ . Thus if ΔZ is not of order $\epsilon^0 = 1$, we have

$$C_2(\Delta Z)u = 0. \quad (56)$$

Thus we must look to cubic or higher degree terms to find the order of ΔZ . Therefore, we look for ΔZ such that ΔZ is linearly independent of $\Delta Z_1, \dots, \Delta Z_n$, and the vectors V such that $C_1 V = C$ and such that

$$\pi_1 \Delta Z = \Delta Z, \quad (57)$$

$$C_2(\pi_1 \Delta Z)^2 = 0, \quad (58)$$

and

$$C_3(\pi_1 \Delta Z)^3 \neq 0 \quad (59)$$

If we find such a ΔZ , we take it to be of order $\epsilon^{1/3}$ as before we look for another such ΔZ linearly independent of the first, choose it to be of order $\epsilon^{1/3}$, and so on until we obtain a maximal such linearly independent set of vectors, all of order $\epsilon^{1/3}$.

As before, we can show that no vector satisfying

$$\pi_1 \Delta Z = \Delta Z \quad (60)$$

$$C_2(\pi_1 \Delta Z)^2 = 0 \quad (61)$$

linearly independent of the above vectors of order $\epsilon^{1/3}$ and occurring as a factor of T

such that

$$C_3 T \neq 0 \quad (62)$$

can occur. Similarly we continue and obtain a maximal linearly independent set of vectors ΔZ of order $\epsilon^{1/k}$ satisfying

$$\begin{aligned} \pi_1 \Delta Z &= \Delta Z \\ C_2 (\pi_1 \Delta Z)^2 &= 0 \\ &\vdots \\ C_{k-1} (\pi_1 \Delta Z)^{k-1} &= 0 \\ C_k (\pi_1 \Delta Z)^k &\neq 0. \end{aligned} \quad (63)$$

Since our original vector space is of finite dimension, there must be a maximum finite such value of k , and all other vector ΔZ must satisfy the conditions

$$\pi_1 \Delta Z = \Delta Z \quad (64)$$

and

$$C_n (\pi_1 \Delta Z)^n = 0 \quad (65)$$

for all positive integers n . In addition, for all T in the domain of C_n for which ΔZ occurs as a factor and all other factors are of order ϵ or $\epsilon^{\frac{1}{2}}$ or ... $\epsilon^{1/k}$, we have

$$C_n T = 0$$

By an easy extension of the argument used in the construction of H (see (50) with as many λ as ΔZ , we can similarly show that a T in which more than one such different ΔZ occur as factors, we have also

$$C_n T = 0 \quad (67)$$

for T in the domain of C_n .

Let ΔZ_i denote the above factors of T , and now define

$$H_n = \sum_{i=1}^j \lambda_i \Delta Z_i + R \quad (68)$$

where R consists of terms corresponding to C_{n-1} , C_{n-2} , ..., C_1 . By assumption,

$$C_n H_n^n = 0. \quad (69)$$

On the other hand, expanding H_n^n , we have a term having some of the ΔZ_i as factors which is different from 0. There is a non zero such term with maximum index i , $1 \leq i \leq j$. We can then choose this λ_j so large in ratio to the remaining λ_i that if the sum of the terms with λ_j is $\neq 0$, then $C_n H_n^n \neq 0$. If there is only one such term

by assumption, it is $\neq 0$, and we have the desired contradiction. If not, then among these terms there must be a non zero term with maximum index $i < j$. Denote this index by k . We then choose λ_j and λ_k sufficiently large but with λ_k small in ratio to λ_j and repeat this process until after a finite number of steps, we arrive at the desired contradiction. If there are any $\Delta Z \neq 0$ for which

$$C_n T = 0 \quad (70)$$

whenever ΔZ occurs as a factor T for all $n \geq 1$, then ΔZ corresponds to a variable a linear combination of the original variables, which does not occur in the Taylor expansions of the constraint function. That is, the constraint functions of G are locally independent of ΔZ . By the analyticity of the G , this means that the constraint functions are globally independent of ΔZ . By holding the remaining variables fixed, we optimize with respect to ΔZ . This gives a new criterion function which is independent of such ΔZ . We can then proceed to optimize with respect to the remaining variables ignoring such ΔZ , providing the criterion function is such that the interchange of limits is valid. For rational criterion functions with denominator bounded away from zero, this is certainly the case. If there is any problem of order of limits, we simply keep such ΔZ , treating them as parameters in the criterion function. We are now prepared to give the proper scaling of each component ΔZ in terms of a perturbation parameter ϵ where

$$\epsilon \geq 0, \quad (71)$$

and ρ_k is the projection operator for terms of index k . Define

$$\begin{aligned} \rho_0(\Delta Z) &= C_1 \Delta Z = \epsilon \tilde{W}_0 \\ \rho_{k^*} \Delta Z &= \epsilon \frac{1}{k^*} \tilde{W}_1 \\ &\vdots \\ \rho_{k^*} \Delta Z &= \epsilon \frac{1}{k^*} \tilde{W}_n. \end{aligned} \quad (72)$$

The vectors $\tilde{W}_0, \tilde{W}_1, \dots, \text{ and } \tilde{W}_n$ are in general, functions of ϵ but are of order

$$\tilde{W}_j = O(1) = O(\epsilon^0) \text{ for } 0 \leq j \leq n. \quad (73)$$

and

$$\Delta Z = \epsilon \tilde{W}_0 + \epsilon \frac{1}{k^*} \tilde{W}_1 + \dots + \epsilon \frac{1}{k^*} \tilde{W}_n. \quad (74)$$

The leading or dominant terms of the Taylor expansion are

$$C_1 \Delta Z \cdot 1 \cdot V_0^1 G + \sum_{j=1}^n \frac{\left(\rho_{k^*} \Delta Z \right)_j^{k^*} \cdot V_0^{k^*} G}{(k_j^*)!}$$

$$= \epsilon \left(\tilde{W}_0^* \cdot \mathbb{1} \cdot \mathbb{V}_0^1 G + \sum_{j=1}^n \frac{(\tilde{W}_j^*)^{k_j^*} \cdot \mathbb{V}_0^1 G}{(k_j^*)!} \right) \quad (75)$$

With the above scaling in terms of ϵ , the dominant terms are all of the same order in ϵ . The first approximation to the non-linear constraints (1) becomes

$$\tilde{W}_0^* \cdot \mathbb{1} \cdot \mathbb{V}_0^1 G + \sum_{j=1}^n \frac{(\tilde{W}_j^*)^{k_j^*} \cdot \mathbb{V}_0^1 G}{(k_j^*)!} = 0 \quad (76)$$

and if \tilde{W}_0^* can be chosen as the dependent variable

$$\tilde{W}_0^* = - \left\{ \sum_{j=1}^n \frac{(\tilde{W}_j^*)^{k_j^*} \cdot \mathbb{V}_0^1 G}{(k_j^*)!} \right\} (\mathbb{V}_0^1 G)^\dagger \quad (77)$$

where $(\mathbb{V}_0^1 G)^\dagger$ denotes the Penrose pseudo inverse of $\mathbb{V}_0^1 G$ where \tilde{W}_j^* denotes the zeroth order terms in the Puiseux Expansions of

$$\tilde{W}_j, \quad 0 \leq j \leq n. \quad (78)$$

To obtain (77), simply expand (2) in a Taylor series as given by (74), then substitute relations (72) into (2), expand each of the \tilde{W}_j as Puiseux series and collect together all terms having the same power of ϵ . Solve for \tilde{W}_0^* using the Penrose pseudo inverse $(\mathbb{V}_0^1 G)^\dagger$, and since by definition \tilde{W}_0^* is independent of ϵ , we have (77).

To obtain higher order approximations to the constraints, if we can take the

$$\tilde{W}_j \text{ for } 1 \leq j \leq n \quad (79)$$

as the independent variables and write

$$\tilde{W}_0 = - \left(\sum_{m=1}^{\infty} \frac{(\Delta \mathbb{Z})^m \cdot \mathbb{V}_0^m G}{m!} \right) \frac{(\mathbb{V}_0^1 G)^\dagger}{\epsilon}. \quad (80)$$

Higher approximations to \tilde{W}_0 than given by (77) can be given by iteration as follows

$$\tilde{W}_0^{(\ell+1)} = - \left(\sum_{m=1}^{\infty} \frac{[(\Delta \mathbb{Z})^{(\ell)}]^m \cdot \mathbb{V}_0^m G}{m!} \right) \frac{(\mathbb{V}_0^1 G)^\dagger}{\epsilon}. \quad (81)$$

By $(\Delta \mathbb{Z})^{(\ell)}$, we mean the ℓ th approximation to $\Delta \mathbb{Z}$, the 0-th approximation to \tilde{W}_0 is given by (77).

We take

$$\tilde{W}_j^{(\ell+1)} = \tilde{W}_j^{(\ell)} = \tilde{W}_j \text{ for } 1 \leq j \leq n \quad (82)$$

since the \tilde{W}_j ($1 \leq j \leq n$) are the independent variables, and the limit as $\ell \rightarrow \infty$ of

$$\tilde{W}_0^{(k)} \text{ is } \tilde{W}_0.$$

When

$$\rho_{k_1^*} = 0 \text{ but } C_1 = 0, \quad (83)$$

we have a situation where in general the Kuhn-Tucker constraint qualifications are violated and there is no linear approximation which represents the non-linear constraints in a reasonable manner. The example of a cusp that we treat in the example section is just such a situation. Nonetheless our prescription for scaling the variables in terms of powers of ϵ remains valid. The terms corresponding to the lowest powers of ϵ could be treated similarly to the way we treated the independent variables \tilde{W}_0 above. However, in general coupled non-linear algebraic equations of degree k_1^* would have to be solved.

The situation where $\rho_{k_1^*} \neq 0$ is more complicated. A vector ΔZ such that $\Delta Z = \rho_{k_1^*} \Delta Z$ corresponds to a direction for which there is no effective constraint. In this case, all derivatives mixed or otherwise in the direction ΔZ are equal to 0, and the direction is unconstrained. The value of ΔZ places no restriction on variables in directions orthogonal to ΔZ .

The directions corresponding to $\rho_{k_1^*}, \rho_{k_2^*}, \dots, \rho_{k_n^*}$ are scaled as before. For each ΔZ such that $\Delta Z = \rho_{k_1^*} \Delta Z$ it may happen that it occurs in a term of the Taylor expansion of the criterion function with non zero coefficient where only derivatives in the directions ΔZ and/or directions corresponding to the projection operators $\rho_{k_1^*}, \dots, \rho_{k_n^*}$ occur. In this case, $\epsilon^0 = 1$ gives the correct scaling in such a way that ΔZ occurs in as low an order term as possible. Note that all exponents α of ϵ^α must be of the form $\alpha = \frac{P}{N}$ for some fixed integer N and some integer $P \geq 0$ (see reference 6). It may or may not happen that the term where ΔZ gives the dominant contribution is of order ϵ . Examples can be given to show that ΔZ may have its dominant contribution in a term of order higher than ϵ . We can now give a general sufficient condition for local optimality. At

$$Z = Z_0, \quad (84)$$

write

$$\Delta Z = C_1 \Delta Z + \rho_{k_1^*} \Delta Z + \dots + \rho_{k_n^*} \Delta Z + \rho_{k_\infty^*} \Delta Z \quad (85)$$

where C_1 may or may not equal zero.

The dependent variables of ΔZ are now replaced by their expansion in terms of the independent variables and substituted into the expression

$$F(Z_0 + \Delta Z) = F(Z_0) + \sum_{m=1}^{\infty} \frac{(\Delta Z)^m \cdot V_0^m}{m!} \quad (86)$$

Ignoring $F(\bar{Z}_0)$ we keep the dominant term in ϵ . If this term is of a definite sign and never zero unless all the independent variables are equal to zero, then we must be at a local optimum. We are at a maximum if the sign is negative and a minimum if it is positive. A necessary condition for local optimality is that the leading non-constant term be semi-definite. That is, it should be of a definite sign or possibly equal zero for non-zero values of the independent variables.

III. Example

As an example, we consider one that is often quoted in the literature to illustrate what can happen when the Kuhn-Tucker constraint qualification is violated.

The criterion function to be maximized is

$$F = X, \quad (87)$$

and the constraints are

$$\begin{aligned} Y - (1-X)^3 &\leq 0 \\ X, Y &\geq 0. \end{aligned} \quad (88)$$

At the point $Y = 0, X = 1$, we have on writing $Y = \Delta Y, X = 1 + \Delta X$

$$\begin{aligned} \Delta Y - (\Delta X)^3 &\leq 0 \\ 1 + \Delta X &\geq 0, \Delta Y \geq 0. \end{aligned} \quad (89)$$

Keeping only the linear terms in ΔX and ΔY and regarding ΔX as small compared to one, we have

$$\Delta Y \leq 0 \quad (90)$$

and

$$\Delta Y \geq 0 \quad (91)$$

or simply

$$\Delta Y = 0 \quad (92)$$

as our linear approximation to relation (88).

Since ΔX is taken to be small compared to one, there is effectively no sign constraint on ΔX in the linear approximations. Thus, according to the linear approximations, we seek to maximize

$$F = 1 + \Delta X \quad (93)$$

subject to $\Delta X = 0$. To the linear approximations, we ought to be able to increase ΔX by some small positive amount, however clearly $X > 1$ is feasible, and it is obvious that $\Delta X = 0$ or $X = 1$ is in fact already the optimum value. Our method is designed, in fact, to deal with just such situations, where in addition, the result may no longer be obvious.

In order to prepare this problem for our technique, we rewrite it in the equivalent form

$$\begin{aligned} F &= U^2 = X \\ Y &= V^2 \end{aligned} \tag{94}$$

and

$$V^2 - (1 - U^2)^3 + W^2 = 0, \tag{95}$$

where U , V , and W are real valued variables. Setting $U_0 = 1$, $V_0 = 0$, and $W_0 = 0$, we write

$$\begin{aligned} U &= 1 + \Delta U \\ V &= \Delta V \\ W &= \Delta W. \end{aligned} \tag{96}$$

The Taylor expansion for $g = V^2 - (1 - U^2)^3 + W^2$ is simply

$$G = (\Delta V)^2 + (\Delta W)^2 + 8(\Delta U)^3 + 12(\Delta U)^4 + 6(\Delta U)^5 + (\Delta U)^6. \tag{97}$$

According to our prescription for scaling, we write

$$\begin{aligned} \Delta V &= \epsilon^{1/2} \tilde{V} \\ \Delta W &= \epsilon^{1/2} \tilde{W} \end{aligned} \tag{98}$$

and

$$\Delta U = \epsilon^{1/3} \tilde{U}.$$

Then

$$G = \epsilon(\tilde{V}^2 + \tilde{W}^2 + 8\tilde{U}^3) + 12\epsilon^{4/3}\tilde{U}^4 + 6\epsilon^{5/3}\tilde{U}^5 + \epsilon^2\tilde{U}^6. \tag{99}$$

To leading order, the constraints are solved by the relation

$$\tilde{U}^3 = -\frac{1}{8}(\tilde{V}^2 + \tilde{W}^2). \tag{100}$$

Clearly $\tilde{U} \leq 0$ and

$$F = 1 + 2\Delta U + (\Delta U)^2 = 1 + 2\epsilon^{1/3}\tilde{U} + \epsilon^{2/3}\tilde{U}^2. \tag{101}$$

Neglecting the trivial constant term 1 to leading order F is described by

$$2\epsilon^{1/3}\tilde{U} = \frac{-2\epsilon^{1/3}}{8^{1/3}}(\tilde{U}^2 + \tilde{W}^2)^{1/3}. \tag{102}$$

But $\epsilon > 0$, and the increment in F is negative definite. It is 0 only for $\tilde{U} = \tilde{W} = 0$. Thus, we know we must be at a local maximum.

REFERENCES

1. R. Bellman, Perturbation Techniques in Mathematics, Physics, and Engineering, Holt, Rinehart, and Winston, 1964.
2. T. L. Boullion and P. L. Odell, Generalized Inverse Matrices, Wiley-Interscience, 1971.
3. M. Van Dyke, Perturbation Methods in Fluid Mechanics, Academic Press, 1964.
4. R. C. Gunning and H. Bossi, Analytic Functions of Several Complex Variables, Prentice-Hall, 1965.
5. O. L. Magasarian, Nonlinear Programming, McGraw-Hill, 1969.
6. Bliss, Algebraic Functions, Dover Publishing Company, 1966.