1. Publication Nº	2. Version	3. Date	5. Distribution
INPE-4060-PRE/1019	*2nd-July 1987*	*Nov., 1986*	☐ Internal ☒ External ☐ Restricted

4. Origin	Program
DMC	*A81*

6. Key words - selected by the author(s)
MATHEMATICAL PROGRAMMING
GRADIENT PROJECTION METHODS
OPTIMIZATION
LINEAR ESTIMATION OF PARAMETERS

7. U.D.C.: *681.514*

8. Title *INPE-4060-PRE/1019*

A STOCHASTIC APPROACH TO GENERATE A PROJECTION OF THE GRADIENT TYPE METHOD

10. Nº of pages: *17*

11. Last page: *16*

12. Revised by

Jose Jaime da Cruz

9. Authorship *Atair Rios Neto*
Ricardo L.U.F. Pinto

Responsible author

13. Authorized by

Marco Antonio Raupp
Diretor Geral

14. Abstract/Notes

A stochastic approach is proposed to generate a direct search procedure where errors in the constraints are inherently treated. The objective is to have the possibility of directly considering the relative accuracy in the satisfaction of constraints, when one is numerically solving a nonlinear constrained optimization problem. This is done by adding to the numerical algorithm the feature of modelling as random variables the errors with which the constraints are supposed to be approximated. Use is made of linear estimation theory, reducing the search increment determination to a problem of parameter estimation with a priori information. This problem is solved with the Gauss-Markov estimator in the Kalman form. The first order, direct search method, that results can be viewed as a stochastic version of the projection of the gradient method. An analysis is made showing the relationships between the proposed stochastic version and the existing deterministic method.

15. Remarks
Submitted for publication in the Journal of Optimization: Theory and Applications

VIII CONGRESSO LATINO-AMERICANO E IBÉRICO
sobre métodos computacionais para engenharia
RIO DE JANEIRO, 3 a 5 de NOVEMBRO de 1987
PUC / RJ
TRABALHO Nº PP:

A STOCHASTIC APPROACH TO GENERATE A PROJECTION OF THE GRADIENT TYPE METHOD

ATAIR RIOS NETO
RICARDO L.U.F.PINTO
Instituto de Pesquisas Espaciais - INPE/MCT
Departamento de Mecânica Espacial e Controle - DMC
Caixa Postal 515
12201 - São José dos Campos - SP - Brazil

SUMÁRIO

Propõe-se uma aproximação para gerar um procedimento de busca direta no qual os erros nos vínculos são inerentemente tratados. Isto é feito adicionando-se ao algoritmo a característica de modelar como variáveis aleatórias os erros com os quais os vínculos são aproximados. Faz-se uso da teoria de estimação linear, resolvendo-se a determinação do incremento de busca com um estimador do tipo Gauss-Markov na forma de Kalman. É feita uma análise mostrando as relações entre a versão estocástica proposta e o método da projeção do gradiente determinístico.

SUMMARY

A stochastic approach is proposed to generate a direct search procedure where errors in the constraints are inherently treated. This is done by adding to the numerical algorithm the feature of modelling as random variables the errors with which the constraints are approximated. Use is made of linear estimation theory, solving the search increment determination problem with a Gauss-Markov estimator in the Kalman form. An analysis is made showing the relationships between the proposed stochastic version and the gradient projection method.

1. Introduction

In what follows a procedure is proposed for the numerical solution of constrained nonlinear programming problems. Adopting a first order, direct search scheme, the problem of determining the search increment in each iteration is treated as one of optimal linear estimation of parameters. This leads to a stochastic approach where the relative accuracy in the satisfaction of the constraints can be represented by random variables and thus directly treated in the resulting numerical algorithm.

The procedure was developed with the purpose of using linear estimation to have a direct search numerical scheme where the accepted errors in the constraints could be treated along and as part of the numerical procedure. What resulted, however, was soon seen to be related to the well known projection of the gradient method([8], [9] and [4]). The use of linear estimation theory (e.g. [2]) or, more specifically, of the Gauss-Markov parameter estimator in the Kalman form, resulted in a procedure([6],[9] and [1]) which can be viewed as a stochastic version of the projection of the gradient method. In this paper, an updated version is presented and an anlysis showing the relationships between the proposed stochastic version and the existing deterministic method is made.

The procedure is intended to be applicable to deterministic problems where one only wants to attain an approximate satisfaction of the constraints. Thus, it is not in the class of those procedures to solve stochastic nonlinear programming problems for which one only has access to random realizations of the objective function or of the constraints (e.g. [3],[5] and [10]).

In what follows, the paper is organized in six sections. In Section 2, the optimization problem to be numerically solved is formulated. The presentation and analysis of the proposed procedure are made in Section 3. In Section 4, the comparison with the deterministic projection of the gradient method is done. To illustrate aspects of implementation and basic characteristics, a simple numerical example is considered in Section 5. Final comments and conclusions are presented in Section 6.

2. Problem Formulation

The optimization problem for which a numerical solution is sought is, if a minimization is considered:

$$\text{Minimize} \quad : \quad f(\underline{x}) \tag{1}$$

$$\text{Subject to} : \quad \underline{g}(\underline{x}) \leq \underline{\delta} \tag{2}$$

$$\underline{h}(\underline{x}) = \underline{\varepsilon} \tag{3}$$

where $\underline{x}$ is the vector of parameters to be optimized; $f(\underline{x})$, $\underline{g}(\underline{x})$ and $\underline{h}(\underline{x})$ are fixed functions of $\underline{x}$ of dimensions 1, m_g and m_h, which are assumed to have the necessary regularity (e.g.[4]) to validate the mathematical developments and assumptions that follow; $\underline{\delta}$ and $\underline{\varepsilon}$ are vectors of errors to represent the accuracy of constraints satisfaction and defined as:

$$|\delta_i| \leq \delta_i^t, \; i = 1,2,\ldots,mg, \; |\varepsilon_i| \leq \varepsilon_i^t, \; i = 1,2,\ldots, m_h \;, \tag{4}$$

where the fixed and given values δ_i^t and ε_i^t characterize the region around zero within which errors are considered as tolerable. The inclusion of error bounds in the satisfaction of the constraints is a way of modelling the problem of what is to be the numerical zero for each constraint. That is, a way of characterizing as acceptable those errors which are within the error bounds.

3. Proposed Procedure

3.1 - General Scheme: Given a value $\bar{\underline{x}}$ from an initial guess or from an immediately previous iteration, a first order, direct search approach is adopted in a typical iteration to determine an approximate solution for the increment $\Delta\underline{x}$ in the problem:

$$\text{Minimize} \quad : f(\bar{\underline{x}} + \Delta\underline{x}) \tag{5}$$

$$\text{Subject to} : \underline{h}(\bar{\underline{x}} + \Delta\underline{x}) = \alpha\, \underline{h}(\bar{\underline{x}}) + \underline{\varepsilon} \tag{6}$$

$$g_i(\bar{\underline{x}} + \Delta\underline{x}) = \beta\, g_i(\bar{\underline{x}}) + \delta_i \tag{7}$$

where $g_i(\bar{\underline{x}}) \geq \delta_i^t$, $i = 1,2,\ldots, I_g$ represent the active constraints; and $0 \leq \alpha < 1$, $0 \leq \beta < 1$ are chosen close enough to one to lead to increments $\Delta\underline{x}$ of a first order of magnitude.

Linearized approximations are taken for the left hand sides of Equations (6) and (7) together with a stochastic interpretation for the errors $\underline{\varepsilon}$ and δ_i, resulting in:

$$(\alpha-1)\ \underline{h}(\bar{\underline{x}}) = (d\ \underline{h}(\bar{\underline{x}})/d\underline{x})\ \Delta\underline{x} + \underline{\varepsilon}^r \tag{8}$$

$$(\beta-1)\ g_i(\bar{\underline{x}}) = (dg_i(\bar{\underline{x}})/d\underline{x})\ \Delta\underline{x} + \delta_i^r \tag{9}$$

where $i = 1,2,\ldots, Ig$; and $\underline{\varepsilon}^r$ and δ_i^r are now assumed to be zero mean uniformly distributed random errors, modelled as:

$$E[\underline{\varepsilon}^r\ \underline{\varepsilon}^{rT}] = \text{diag}\ [e_i^2,\ i=1,2,\ldots,m_h],\ e_i^2 = (\varepsilon_i^t)^2/3\ ,$$

$$E[\underline{\delta}^r\ \underline{\delta}^{rT}] = \text{diag}\ [d_i^2,\ i=1,2,\ldots,I_g],\ d_i^2 = (\delta_i^t)^2/3\ ,$$

where $E[.]$ indicates the expected value of its argument; and ε_i^t and δ_i^t are as defined in Equation (4).

The condition of Equation (5) is approximated by the following <u>a priori information</u>

$$-\ p\ \underline{\nabla}\ f^T(\bar{\underline{x}}) = \Delta\underline{x} + \underline{\eta}\ , \tag{10}$$

where $p \geq 0$ is to be adjusted to guarantee a first order of magnitude for the increment, that is, such that $\Delta\underline{x}$ is small enough to permit the use of a linearized representation of $f(\bar{\underline{x}} + \Delta\underline{x})$; and $\underline{\eta}$ is taken as a zero mean uniformly distributed random vector, modelling the a priori searching error in the direction of the gradient $\underline{\nabla} f(\bar{\underline{x}})$, with:

$$E[\underline{\eta}\ \underline{\eta}^T] = \bar{\underline{P}}$$

as its diagonal covariance matrix. The values of the variances in $\bar{\underline{P}}$ are to be chosen such as to characterize an <u>adequate order of magnitude</u> for the dispersion of $\underline{\eta}$ as it will be made clear later.

The diagonal form adopted is to model the assumption that it is not imposed any a priori correlation between the errors in the gradient components.

The simultaneous consideration of conditions of Equations (8), (9) and (10) characterize a problem of parameter estimation, which in a compact notation can be put as follows:

$$\underline{\bar{X}} = \underline{X} + \underline{n} \quad , \tag{11}$$

$$\underline{Y} = \underline{M}\ \underline{X} + \underline{V} \ , \tag{12}$$

where $\underline{\bar{X}} \triangleq - p\ \underline{\nabla} f^T(\underline{\bar{x}})$ is the a priori information; $\underline{X} \triangleq \Delta x$; $\underline{Y}^T \triangleq [(\alpha-1)\ \underline{h}^T(\underline{\bar{x}}):(\beta-1)\ g_1(\underline{\bar{x}}):\ldots:(\beta-1)g_{I_g}(\underline{\bar{x}})]$ is the mx1 observation vector; $\underline{M}^T \triangleq [(d\underline{h}(\underline{\bar{x}})/d\underline{x})^T:(dg_1(\underline{\bar{x}})/d\underline{x})^T:\ldots:(dg_{I_g}(\underline{\bar{x}})/d\underline{x})^T]$; $\underline{V}^T \triangleq [(\varepsilon^r)^T:\ \delta_1^r:\ldots:\ \delta_{Ig}^r]$.

Adopting a criterion of linear, minimum varinace estimation, the optimal search increment can be determined using the classical Gauss-Markov estimator, which in Kalman form (e.g. [2]) gives:

$$\underline{\hat{X}} = \underline{\bar{X}} + \underline{K}(\underline{Y} - \underline{M}\ \underline{\bar{X}}) \ , \tag{13}$$

$$\underline{P} = \underline{\bar{P}} - \underline{KM\bar{P}} \quad , \tag{14}$$

$$\underline{K} = \underline{\bar{P}}\ \underline{M}^T(\underline{M\bar{P}M}^T + \underline{R})^{-1} \tag{15}$$

where $\underline{\bar{P}}$ is defined as before; $\underline{R} \triangleq E[\underline{VV}^T] \triangleq \mathrm{diag}\ [R_k, k=1,2,\ldots,m]$; and $\underline{P}$ has the meaning of being the covariance matrix of the errors in the components estimates of $\underline{X}$, i.e:

$$\underline{P} = E[(\underline{X} - \underline{\hat{X}})\ (\underline{X} - \underline{\hat{X}})^T]. \tag{16}$$

A formally equivalent way of obtaining the estimate of Equations (13) to (15) is to solve the deterministic optimization problem:

$$\underset{\{X\}}{\text{Minimize:}}\ \frac{1}{2}\ ((\underline{X}-\underline{\bar{X}})^T\underline{\bar{P}}^{-1}(\underline{X}-\underline{\bar{X}}) + (\underline{Y}-\underline{MX})^T R^{-1}(\underline{Y}-\underline{MX})) \tag{17}$$

This way of interpreting the solution shows that the numerical procedure gives in each iteration an increment $\hat{\underline{X}}$ which: (i) differs from $\bar{\underline{X}}$ according to the bounds imposed by the covariance matrix $\bar{\underline{P}}$; and (ii) controls the constraints components deviation from zero, according to the relative priorities imposed by the elements of the covariance matrix $\underline{R}$. It also shows that $\bar{\underline{P}}$ is to be related to $\underline{R}$ such as to guarantee priority to the constraints, in the sense that constraints satisfaction is to be looked as a necessary condition, in the search for the minimum.

3.2 - Analysis of Convergence: To build a numerical algorithm using the proposed procedure, the following types of iterations are considered:

(i) initial phase of acquisition of constraints, when starting from a feasible point that satisfies the inequality constraints, the search is done to capture the equality constraints, including those inequality constraints that eventually became active along this phase;

(ii) search of the minimum, when from a point that satisfies the constraints in the limits of the tolerable errors $\underline{V}$ in Equation (12), the search is done to take the objective function (Equation (5)) to get to the minimum; this search is conducted relaxing the order of magnitude of the error bounds around the constraints;

(iii) restoration of the constraints, when from a point that resulted from a type (ii) iteration, the search is done to restore constraints satisfaction, within the limits imposed by the error $\underline{V}$ in Equation (12).

In the limiting situation of perfect constraints satisfaction, a situation of convergence occurs when in a type (ii) iteration no feasible direction of search can be found and the search increment is zero. The geometrical interpretation of this situation is that the objective function gradient vector is orthogonal to the intersection of hyperplanes representing the active constraints. For the proposed procedure, the correspondent situation of convergence occurs when in a type (ii) iteration the estimate $\hat{\underline{X}}$ of the search increment results negligible, in spite of using a first order of magnitude p in Equation (11), together with a dispersion of $\underline{\eta}$ also of first order. What is to be considered negligible can be evaluated by analysing the following equation, in a situation of convergence ($\underline{X} \cong 0$):

$$\underline{e}_X \triangleq \hat{\underline{X}} - \underline{X} = (\underline{I} - \underline{KM})\ \bar{\underline{e}}_X + \underline{KV}\ , \tag{18a}$$

$$\cong \hat{\underline{X}} \qquad \cong (\underline{I} - \underline{KM})\ \bar{\underline{X}} + \underline{KV}\ , \tag{18b}$$

where Equation (12) and (13) were used, and $\bar{\underline{e}}_X \triangleq \bar{\underline{X}} - \underline{X}$. It is then seen that a negligible $\hat{\underline{X}}$ occurs when the first term in the right hand side of Equation (18b) is of the a order of magnitude less than or equal to that corresponding to the dispersion of the second error term, $\underline{KV}$. Notice that for a type (ii) iteration the result of Equation (18b) is also that given by Equation (13) when $\underline{Y}$ is approximated by $\underline{KV}$.

To verify convergence it is still necessary to check if, together with a negligible increment as given by Equation (18), all the inegualty constraints that are active have the tendency to remain active. In the limiting situation of perfect satisfaction of constraints, this corresponds to not having a feasible increment in the direction of the objective function gradient which could lead to a point inside the allowed constraint region. Geometrically this means that the vector opposite to the objective function gradient, which is orthogonal to the constraints, has no tendency to penetrate the allowable constraint region, for any of the active constraints.

In terms of the proposed procedure, this second condition is verified if:

$$\underline{M}^a\ \bar{\underline{X}} + \underline{V}^a \geq 0, \tag{19}$$

where $\underline{M}^a$ and $\underline{V}^a$ are respectively the partitions of $\underline{M}$ and $\underline{V}$ in Equation (12), in correspondence with the active inequality constraints; and the inequality sign is meant to be valid for each component of the left-hand side vector.

Another equivalent way of considering this condition results if in Equation (18) one notices that the true increment $\underline{X}$ is negligibly small and thus $\bar{\underline{X}}$ obeys the relationship:

$$(\underline{I} - \underline{K}^a\underline{M}^a)\ \bar{\underline{e}}_X + \underline{K}^a\underline{V}^a \cong \bar{\underline{X}} + \underline{K}^a(\underline{V}^a - \underline{M}^a\ \bar{\underline{X}}) \tag{20}$$

where, in Equation (13), $\underline{Y}^a = \underline{M}^a\ \bar{\underline{X}} + \underline{V}^a \cong \underline{V}^a$; and thus:

$$\bar{\underline{X}} = \underline{K}^a\underline{M}^a\bar{\underline{X}} + (\underline{I} - \underline{K}^a\underline{M}^a)\ \bar{\underline{e}}_X \tag{21}$$

resulting from Equation (19) that:

$$\underline{M}^a\underline{K}^a\underline{M}^a\bar{\underline{X}} + (\underline{M}^a(\underline{I} - \underline{K}^a\underline{M}^a)\ \bar{\underline{e}}_X + \underline{V}^a) \geq 0, \tag{22}$$

3.3 - Numerical Algorithm: The algorithm proposed is in analogy with that used in [4] for the projection of the gradient method. It can be summarized as follows:

1. Find the active constraints and form $\underline{M}$ in Equation (12). At most in each iteration of this type, adjust the following:(i) the parameters α and β in Equation (12), when in an initial phase of acquisition of constraints to guarantee the linear perturbation hypotesis; if the iteration is one where constraint restoration is being done take $\alpha = \beta = 1$; (ii) the parameters necessary to get a dispersion of $\underline{\eta}$ in Equation (11) of a first order of magnitude as compared to the dispersion as compared to the dispersion of $\underline{V}$ in Equation (12) (that is, such as to guarantee a higher priority to the constraints satisfaction or, equivalently, such as to give higher weight to the second term in the minimization of the cost in Equation (17)); and (iii) the parameter p, in Equation (11), with a value p_c compatible with the hypotesis that $(\bar{\underline{x}} + \bar{\underline{X}}(p_c))$ is inside the region of validity of linear perturbation of the constrainsts, that is,

p_c = Max {p: $\bar{\underline{x}} + \bar{\underline{X}}(p)$ is inside the region of validity of the linear perturbation of the constraints}.

Perform constraint restoration iterations using the algorithm of Equation (13) to (15), until:

$$|h_i(\underline{\bar{x}} + \underline{\hat{X}}(p_c))| \leq \varepsilon_i^t \ , \ |g_j(\underline{\bar{x}} + \underline{\hat{X}}(p_c))| \leq \delta_j^t \ , \tag{23}$$

where $i = 1,2,\ldots,m_h$ and $j = 1,2,\ldots, I_g$. Notice that in an iteration of this type it may be convenient not to reject a point which, in spite of not improving constraint satisfaction, resulted inside the region of validity of linear perturbation and led to a decrease in the objective function.

2. If necessary, recalculate the dispersion of $\underline{\eta}$ and the parameter p_c. With the algorithm of Equations (13) to (15) calculate the expression for $\underline{\hat{X}}(p)$ and the value of $\underline{K}$. Evaluate the estimate for $p = p_c$, i.e., calculate $\underline{\hat{X}}(p_c)$. Go to step 3, if it happens that:

$$|\underline{\hat{X}}(p_c) - \underline{K}\,\underline{Y}| \leq \gamma \ |\underline{K}\,\underline{E}_V| \ , \tag{24}$$

where $0 < \gamma \leq 1$, to be adjusted depending upon the problem; $\underline{E}_V^T \triangleq [\varepsilon_1^t, \varepsilon_2^t,\ldots, \varepsilon_{m_h}^t, \delta_1^t, \delta_2^t,\ldots, \delta_{I_g}^t]$, and the errors ε_i^t and δ_i^t are as defined in Equation (4). If not, determine p_d associated with:

$$\text{Min } \{f(\underline{\bar{x}} + \underline{\hat{X}}(p)): \ 0 \leq p \leq p_c\} \ , \tag{25}$$

set $\underline{\bar{x}}$ to $(\underline{\bar{x}} + \underline{\hat{X}}(p_d))$ and if constraint satisfation resulted destroyed return to step 1; otherwise repeat this step.

3. With $\underline{\bar{X}}(p_c)$ calculate the vector $\underline{M}^a\underline{\bar{X}}(p_c)$ in Equation (19). If it happens that:

$$M_i^a \ \underline{\bar{X}}(p_c) \geq - \delta_i^t, \ i = 1,2,\ldots,I_g \tag{26}$$

then stop. Otherwise, delete as active the inequality constraints corresponding to the most negative result in Equation (26) and return to step 2.

Notice that after restoration of constraints in the first step, following a search for the minimum in the second step, it may occur that the objective function increases as compared to its value at the beginning of this cycle of iterations. This can be remedied by restarting the cycle with a smaller p_d in the second step.

Among the many possibilities certainly existing for the determination of the parameters needed in the implementation of the proposed algorithm, the following are suggested:

(i) choose α and β of Equation (12), in each step, as suggested [6]:

$$(\alpha-1) = (\beta-1) = -s \tag{27}$$

$$s = \text{Min.}\{s_i, i=0,1,2\ldots,m: s_0=1,\ s_j^2 h_j^2(\bar{\underline{x}}) = q_j(3R_j),\ j=1,2\ldots,m_h;$$

$$s_{k+m_h}^2 g_k(\bar{\underline{x}}) = q_k(3R_{k+m_h}),\ k = 1,2,\ldots,\ I_g\}; \tag{28}$$

where $q_j >> 1$ and $q_k >> 1$ are to be adjusted to guarantee a first order of magnitude perturbation in the constraints.

(ii) Determine the dispersion of $\underline{\eta}$ in Equation (11) to produce a first order perturbation error in the constraints of Equation (12), by considering that;

$$\underline{Y} = \underline{M}\ (\bar{\underline{X}} + \underline{\eta}) + \underline{V} = \bar{\underline{Y}} + \underline{M}\ \underline{\eta}$$

or

$$\underline{Y} - \bar{\underline{Y}}\ \underline{\Delta}\ \underline{\varepsilon}^Y = \underline{M}\ \underline{\eta}$$

resulting in:

$$\sum_{j=1}^{n} M_{ij}^2\ \bar{P}_{jj} = E[(\varepsilon_i^Y)^2],\ i = 1,2,\ldots,\ m, \tag{29}$$

where it is assumed $E[(\varepsilon_i^y)^2] = q_i(3R_i)$, so that:

$$\sum_{j=1}^{n} M_{ij}^2 \bar{P}_{jj} = q_i(3R_i), \quad i = 1,2\ldots, m, \tag{30}$$

which is a condition to be used for evaluating those $\bar{P}_{jj}$ that are present in Equation (29) (that is, those $\bar{P}_{jj}$ for which at least one of the $M_{ij} \neq 0$). Noticing that $\bar{\underline{X}}(p_c)$ is also to be taken as an increment of first order, then for those $\bar{X}_j(p_c) \neq 0$ another condition that can be used is:

$$3\bar{P}_{jj} = \bar{X}_j^2(p_c) \ . \tag{31}$$

Eliminating those $\bar{P}_{jj}$ that appear in Equation (30) and for which $\bar{X}_j(p_c) \neq 0$, one gets a condition for determining p_c and those $\bar{P}_{jj}$ that eventually remained in Equation (30). Depending upon the problem, either redundancy or insufficiency of equations may occur to determine the unknown variable of this condition. In the first case, a least square fitting can be used. In the second, an additional criterion has to be adopted to complete this condition. After determining p_c, the determination of the $\bar{P}_{jj}$ can be completed using again the Equation (31).

4. Relationship with the Projection of the Gradient

A deterministic version of the projection of the gradient method results if the following particularizing assumptions are simultaneously considered.

(i) In Equation (15), one takes:

$$\underline{\bar{P}} = \alpha \ \underline{I}_n \ , \ \alpha > 0 \tag{32}$$

(ii) In this same Equation (15), one takes the ratios R_k/α, k=1,2,...,m, to zero in the limit, resulting in a correspondent value $\underline{K}_d$, given by:

$$\underline{K}_d = \underline{M}^T(\underline{M}\ \underline{M}^T)^{-1} \tag{33}$$

Taking the result of Equation (33) in Equation (13), it results:

$$\underline{\hat{X}} = (\underline{I}_n - \underline{K}_d \underline{M})\ \underline{\bar{X}} + \underline{K}_d\ \underline{Y} = \underline{P}_d\ \underline{\bar{X}} + \underline{K}_d\ \underline{Y} \tag{34}$$

where $\underline{P}_d \triangleq (\underline{I}_n - \underline{K}_d\ \underline{M})$ is now the projection matrix, exactly as defined for the deterministic version.

The first Kuhn - Tucker condition is obtained directly from Equation (34), if constraint satisfaction is assumed ($\underline{Y} = 0$) and in Equation (10) one takes $p > 0$, resulting in:

$$(\underline{I}_n - \underline{K}_d\ \underline{M})\ \nabla f(\underline{\bar{x}}) = 0, \tag{35}$$

which corresponds to the situation where the objective function gradient vector is orthogonal to the active constraints.

The second Kuhn - Tucker condition results from Equation(22), with $p > 0$ and with $\underline{K}^a$ taken to the deterministic limit, $\underline{K}_d^a$:

$$- \underline{K}_d^{aT}\ \underline{\nabla}_f^{\ T}(\underline{\bar{x}}) \geq 0. \tag{36}$$

5. Numerical Example

As an example, for a preliminary testing of the algorithm of section 3.3, the following problem was considered:

$$\text{Minimize} \quad : f(\underline{x}) = x_3 \tag{37}$$

$$\text{Subject to} : h_1(\underline{x}) = x_1 x_2 + x_3 = 0;\ h_2(\underline{x}) = x_1^2 + x_2^2/4 - 1 = 0 \tag{38}$$

This problem has as exact solutions:

$$x_1 = \sqrt{2}/2,\ x_2 = \sqrt{2}\ ,\ x_3 = -1;\ x_1 = -\sqrt{2}/2,\ x_2 = -\sqrt{2},\ x_3 = -1.$$

Two cases were tested, with different tolerances for the constraints errors:

case 1: $|h_1| \leq .001$, $|h_2| \leq .01$

case 2: $|h_1| \leq .01$, $|h_2| \leq .001$.

The numerical results obtained are shown in Tables 1 and 2, respectively. The parameters needed in the implementation of the algorithm were calibrated as follows:

case 1: $\gamma = 1$; $\alpha = 0$; $q_1 = 1.+06$, $q_2 = q_1 R_1/R_2$ (see Equation (29))

case 2: $\gamma = 1$; $\alpha = 0$; $q_1 = 1.+04$, $q_2 = q_1 R_1/R_2$

ITERATION	x_1	x_2	x_3	h_1	h_2
0	2.00000 + 00	2.00000 + 00	2.00000 + 00	6.00000 + 00	4.00000 + 00
1	1.28571 + 00	8.57147 - 01	-2.85705 - 01	8.16334 - 01	8.36715 - 01
2	9.39497 - 01	9.82138 - 01	-9.65990 - 01	-4.32733 - 02	1.23804 - 01
3	8.01137 - 01	1.25957 + 00	-1.04747 + 00	-3.83863 - 02	3.84482 - 02
4	7.26776 - 01	1.38785 + 00	-1.01819 + 00	-9.54004 - 03	9.73368 - 03
5	7.08680 - 01	1.41187 + 00	-1.00100 + 00	-4.35726 - 04	5.71348 - 04

TABLE 1

FIRST CASE

ITERATION	x_1	x_2	x_3	h_1	h_2
0	2.00000 - 00	2.00000 + 00	2.00000 + 00	6.00000 + 00	4.00000 + 00
1	1.28570 + 00	8.57187 - 01	-2.85652 - 01	8.16436 - 01	8.36725 - 01
2	9.39481 - 01	9.82138 - 01	-9.65993 - 01	-4.32927 - 02	1.23773 - 01
3	8.01108 - 01	1.25954 + 00	-1.04751 + 00	-3.84774 - 02	3.83863 - 02
4	7.26734 - 01	1.38781 + 00	-1.01821 + 00	-9.64230 - 03	9.64517 - 03
5	7.08640 - 01	1.41181 + 00	-1.00100 + 00	-5.36044 - 04	4.72445 - 04
6	7.07283 - 01	1.41387 + 00	-1.00011 + 00	-1.02879 - 04	3.89684 - 06

TABLE 2

SECOND CASE

To adjust the parameters which characterize the dispersion of η (the values of $\overline{P}_{jj}$ in Equation (30)), a criterion to give equal opportunity of each $\overline{P}_{jj}$ contributing to the first order perturbation was adopted (e.g. [7]). Following this criterion, the determination of $\overline{P}_{jj}$, for a given j, resulted from solving with a least square fitting equations of the type:

$$M_{ij}^2 \; \overline{P}_{jj} = q_i \; (3R_i)/n_i, \; i = 1,2,\ldots,m \;, \qquad (39)$$

where n_i is the number of $M_{ij} \neq 0$ for a given value of the index i.

The analysis of the results shows a satisfactory performance of the method for the problem tested. Convergence is reached with a very reduced number of iterations. The criterion to check convergence (Equation (24)) worked well in both cases. Other numerical tests done have showed that there are occasions when this criterion may be too restrictive. In all tests done it was found that the parameters q_i are important to the performance of the algorithm, specially in terms of number of iterations necessary; however it was verified that there is no excess of sensitivity to their variation and that it is not difficult to adjust them to obtain a good solution.

6. CONCLUSIONS

A stochastic version of the projection of the gradient method was proposed. The stochastic approach led to a method which allows the direct treatment of constraints errors along the numerical solution. In this sense, the proposed method is certainly more general than any existing projection of the gradient deterministic version. However, as any numerical method, it is also problem dependent, and being a more general method is not sufficient to guarantee always a better performance. Its stochastic characteristic adds complexity to the associated numerical algorithm and thus a careful judgment of necessity has to be made to evaluate the convenience of adopting the method for a given problem.

The algorithm suggested in Section 3.3 is a direct consequence of the search strategy adopted in Section 3.2. It should be considered a result still under analysis and testing. Other strategies may be found which result more well-suited to the method, leading to other alternatives in terms of the numerical algorithm.

The numerical application presented in Section 5 was done with the purpose of a preliminary numerical testing to assess the general functioning characteristics of the method and to illustrate how the direct treatment of the errors in the constraints affects the final solution.

Aside from the need of investigating alternative algorithms, there are many other aspects of the stochastic method which should be better investigated. Among these are included those relative to the calibration of the parameter p and of the dispersion of $\underline{n}$, in Equation (10). Future investigations shall take advantage of the vast and well-succeeded experience with the deterministic projection of the gradient. This shall be combined with an effort to explore the possibility of using results already available in linear estimation theory, specially in adaptive state estimation.

REFERENCES

1 BAMBACE, L.A.W. Implementação e Teste de um Procedimento Subótimo para a Obtenção do Controle em Problemas de Otimização Dinâmica. M.Sc. Thesis, INPE-2508-TDL/098 - Brazil, 1982.

2 JAZWINSKI, A.H. Stochastic Processes and Filtering Theory. Academic Press, New York, 1970.

3 KUSHNER, H.J. and GAVIN, T. Stochastic Approximation Type Methods for Constrained Systems: Algorithms and Numerical Results. IEEE Trans. Aut. Control, AC-19, 4, 349-357, 1974.

4 LUENBERGER, D.G. Introduction to Linear and Nonlinear Programming. Addison-Wesley, Reading, Mass, 1973.

5 POLJAK, B.T. Nonlinear Programming Methods in the Presence of Noise. Mathematical Programming, 14, 87-97, 1978.

6 RIOS-NETO, A. Estimação Linear Ótima Aplicada a Geração de Método Direto de Busca em Otimização de Parâmetros. Rev.Br.C.Mec, III, 1, 11-14, 1981.

7 RIOS-NETO, A. and BAMBACE, L.A.W. Optimal Linear Estimation and Suboptimal Numerical Solutions of Dynamical Systems Control Problems. Proc. Int. Symp. Spacecraft Flight Dynamics, Darmstadt, FRG (ESA SP-160), Aug. 1981.

8 ROSEN, J.B. The Gradient Projection Method for Nonlinear Programming. Part I, Linear Constraints. J.Soc.Indust. Appl. Math, 8, 1, 181-217, 1960.

9 ROSEN, J.B. The Gradient Projection Method for Nonlinear Programming. Part II, Nonlinear Constraints. J.Soc. Indust. Appl. Math, 9, 4, 515-532, 1961.

10 RUSZCZYNSKI, A. Feasible Direction Methods for Stochastic Programming Problems. Mathematical Programming,19,220-229,1980.