

ON PROBABILISTIC CONSTRAINED PROGRAMMING

András PRÉKOPA

Technical University of Budapest
and

Mathematical Institute of the Hungarian Academy of Sciences

1 Introduction

The term probabilistic constrained programming means the same as chance constrained programming, i.e., optimization of a function subject to certain conditions where at least one is formulated so that a condition, involving random variables, should hold with a prescribed probability. The probability is usually not prescribed exactly but a lower bound is given instead which is in practice near unity.

The formulation of problem in connection with a stochastic system follows a line where first we consider the nonstochastic case, formulate the problem which loses its meaning as soon as certain parameters become random. Then the next step is the formulation of the right problem optimizing the system. This problem may be of different type depending on the system and our information. It is not possible to formulate one decision structure to cover all stochastic situations.

The deterministic problem from which we start is the following

$$\begin{aligned} & \min f(\mathbf{x}) \\ & \text{subject to} \\ & g_i(\mathbf{x}) \geq \beta_i, \quad i = 1, \dots, m, \\ & \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M, \end{aligned} \tag{1.1}$$

where \mathbf{x} is a vector of the n -dimensional space, $g_i(\mathbf{x})$, $f(\mathbf{x})$ are certain functions on which we shall impose some conditions later on when formulating the stochastic case. Suppose now that in (1.1) β_1, \dots, β_m are random variables and formulate the decision problem¹ of

¹The idea of formulating a decision problem by prescribing a lower bound for the probability of the fulfillment of the constraints was formulated first in [2].

the stochastic system as follows

$$\begin{aligned}
& \min f(\mathbf{x}) \\
& \text{subject to} \\
& P(g_i(\mathbf{x}) \geq \beta_i, \quad i = 1, \dots, m) \geq p, \\
& \quad \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M.
\end{aligned} \tag{1.2}$$

The nonnegative constraints, if any, are thought to be contained in the system of linear inequalities in (1.2). We introduce the following assumptions.

- A.1. The functions $g_1(\mathbf{x}), \dots, g_m(\mathbf{x})$ are defined in the closure \overline{K} of an open convex set K . These functions are concave and have continuous derivatives with respect to all variables in \overline{K} .
- A.2. The number p is between 0 and 1, $0 < p < 1$.
- A.3. If $\mathbf{x} \in \overline{K}$ and \mathbf{x} satisfies the constraints in (1.2), then \mathbf{x} is an internal point of \overline{K} , i.e., $\mathbf{x} \in K$.
- A.4. The function $f(\mathbf{x})$ is defined in an open convex set H containing the set of feasible solutions and we suppose that $f(\mathbf{x})$ is convex and has continuous derivatives with respect to all variables in every point of H .
- A.5. The random variables β_1, \dots, β_m have a continuous joint distribution which has continuous first order derivatives with respect to all variables in any point of the m -dimensional space of the form

$$(g_1(\mathbf{x}), \dots, g_m(\mathbf{x})), \quad x \in K.$$

We consider only such probabilities p , for which the set of feasible solutions is not empty. This set will be denoted by $D(p)$ in the sequel. Let $F(z_1, \dots, z_m)$ denote the joint probability distribution function of the random variables: β_1, \dots, β_m , i.e.,

$$F(z_1, \dots, z_m) = \mathbf{P}(\beta_1 \leq z_1, \dots, \beta_m \leq z_m)$$

and let further

$$G(\mathbf{x}) = F(g_1(\mathbf{x}), \dots, g_m(\mathbf{x})) = \mathbf{P}(g_1(\mathbf{x}) \geq \beta_1, \dots, g_m(\mathbf{x}) \geq \beta_m).$$

- A.6. For every \mathbf{x} satisfying the equality $G(\mathbf{x}) = p$, there corresponds a vector \mathbf{y} in the set of feasible solutions with the property that²

$$\nabla G(\mathbf{x})(\mathbf{y} - \mathbf{x}) > 0. \tag{1.3}$$

²This condition plays the same role here as the well-known Slater's condition in convex programming.

We remark that the gradient of a vector will be considered a row vector while all other vectors are columns. (The prime means the transpose.) Thus (1.3) has a meaning. As $G(\mathbf{y}) \geq p$ for all feasible vectors \mathbf{y} , it follows that

$$\nabla G(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0$$

provided the segment joining \mathbf{x} and \mathbf{y} is inside the set of feasible solutions. This holds if the feasible set is convex what we shall prove under certain assumptions. The purpose of the paper is the discussion of problem (1.2), to reduce it to a convex (or quasi-convex) programming problem and give an algorithm for finding the optimal solution.

2 Preliminary Lemmas

Let us introduce the notations

$$\begin{aligned} I &= \{1, 2, \dots, M\} \\ I(\mathbf{x}) &= \{i : i \in I, \mathbf{a}'_i \mathbf{x} = b_i\}, \end{aligned}$$

where \mathbf{x} is a feasible solution.

LEMMA 1 *Suppose that the set of feasible vectors is convex. If \mathbf{x} is feasible, $G(\mathbf{x}) = p$ and $u \geq 0$, $u_i \geq 0$, $i \in I$, further*

$$u \nabla G(\mathbf{x}) + \sum_{i \in I(\mathbf{x})} u_i \mathbf{a}'_i = \mathbf{0}', \quad (2.1)$$

then we have $u = 0$.

PROOF. Let \mathbf{y} be the vector satisfying assumption A.6. and multiply (2.1) by $\mathbf{y} - \mathbf{x}$. Then we have

$$0 = u \nabla G(\mathbf{x})(\mathbf{y} - \mathbf{x}) + \sum_{i \in I(\mathbf{x})} u_i \mathbf{a}'_i (\mathbf{y} - \mathbf{x}) \geq u \nabla G(\mathbf{x})(\mathbf{y} - \mathbf{x}) \geq 0.$$

Thus

$$u \nabla G(\mathbf{x})(\mathbf{y} - \mathbf{x}) = 0$$

and by A.6. we have $u = 0$. □

LEMMA 2 *Suppose that the set of feasible vectors is convex. Let \mathbf{x} be feasible and $G(\mathbf{x}) = p$. Then Kuhn–Tucker constraint qualification holds at the point \mathbf{x} .*

REMARK 1 If \mathbf{x} is such a feasible vector for which $G(\mathbf{x}) > p$, then the fulfillment of the fact that all other constraints in problem (1.2) are linear.

PROOF OF LEMMA 2. Let \mathbf{h} be a vector with the property that

$$\begin{aligned}\nabla G(\mathbf{x})\mathbf{h} &\geq 0, \\ \mathbf{a}'_i\mathbf{h} &\geq 0, \quad \text{for } i \in I(\mathbf{x}).\end{aligned}\tag{2.2}$$

We define the curve

$$\boldsymbol{\psi}(t) = \mathbf{x} + t[\mathbf{h} + t(\mathbf{y} - \mathbf{x})], \quad t \geq 0,\tag{2.3}$$

where \mathbf{y} is the feasible vector satisfying A.6. If $t > 0$, then by (1.3) and the first row of (2.2) we have

$$\nabla G(\mathbf{x})t[\mathbf{h} + t(\mathbf{y} - \mathbf{x})] > 0.$$

Thus by continuity, for small positive t values we also have

$$\begin{aligned}G(\mathbf{x} + t[\mathbf{h} + t(\mathbf{y} - \mathbf{x})]) - G(\mathbf{x}) \\ = \nabla G(\mathbf{x} + \delta t[\mathbf{h} + t(\mathbf{y} - \mathbf{x})])t[\mathbf{h} + t(\mathbf{y} - \mathbf{x})] > 0,\end{aligned}$$

where $0 < \delta < 1$. Thus there exists a $T > 0$ such that

$$G(\boldsymbol{\psi}(t)) \geq p \text{ for } 0 \leq t \leq T.$$

The curve (2.3) has the required property and the linear constraints can be treated in the customary way.

If the set of feasible solutions is convex, then by Lemma 2, the Kuhn–Tucker conditions are necessary³ that a certain feasible \mathbf{x}^* be optimal. The Kuhn–Tucker conditions in our case are the following:

$$\begin{aligned}-\nabla f(\mathbf{x}^*) + \lambda^*G(\mathbf{x}^*) + \sum_{i=1} \lambda_i^* \mathbf{a}'_i &= \mathbf{0}', \\ \lambda^*(G(\mathbf{x}^*) - p) + \sum_{i=1} \lambda_i^*(\mathbf{a}'_i\mathbf{x}^* - b_i) &= 0, \\ \lambda^* \geq 0, \quad \lambda_i^* \geq 0, \quad i = 1, \dots, M.\end{aligned}\tag{2.4}$$

The following lemma is an adaptation of the theory of quasi-concave⁴ programming of ARROW and ENTHOVEN [1].

LEMMA 3 *If $G(\mathbf{x})$ is quasi-concave in its domain of definition and the Kuhn–Tucker conditions (2.4) hold at \mathbf{x}^* , then \mathbf{x}^* is an optimal solution to problem (1.2).*

³Assumptions 1,3,4 and the fact that $D(p)$ is not empty imply that the set of feasible solutions is closed. The necessity of the Kuhn–Tucker conditions can be proved by using different analytic assumptions. It is not difficult to see that our assumptions are enough for a proof.

⁴A function $G(\mathbf{x})$ defined over a convex set H is quasi-concave if for every pair $\mathbf{x}_1, \mathbf{x}_2 \in H$ and $\lambda(0 \leq \lambda \leq 1)$, we have

$$G(\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2) \geq \min(G(\mathbf{x}_1), G(\mathbf{x}_2)).$$

It is easy to see that the set $\{\mathbf{x} : \mathbf{x} \in H, G(\mathbf{x}) \geq a\}$ is convex for every real a (for which the set is not empty) iff $G(\mathbf{x})$ is quasi-concave on H .

PROOF. The set of feasible solutions is now automatically convex. Let \mathbf{x} be an arbitrary feasible vector. Let us multiply the first row of (2.4) by $\mathbf{x} - \mathbf{x}^*$. Then we have

$$\begin{aligned}
0 &= -\nabla f(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] + \lambda^* \nabla G(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] + \sum_{i=1} \lambda_i^* \mathbf{a}'_i[\mathbf{x} - \mathbf{x}^*] \\
&= -\nabla f(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] + \lambda^* G(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] + \sum_{i \in I(\mathbf{x}^*)} \lambda_i^* \mathbf{a}'_i[\mathbf{x} - \mathbf{x}^*] \\
&\geq -\nabla f(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] + \lambda^* \nabla G(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*].
\end{aligned} \tag{2.5}$$

The second term on the right hand side vanishes if $G(\mathbf{x}^*) > p$, because in this case the second row of (2.4) implies $\lambda^* = 0$. If on the other hand $G(\mathbf{x}^*) = p$, then by the quasi-concavity of $G(\mathbf{x})$ we have

$$G(\mathbf{x}^*) + \lambda(\mathbf{x} - \mathbf{x}^*) \geq G(\mathbf{x}^*) = p \quad \text{for } 0 \leq \lambda \leq 1.$$

Thus we obtain

$$G(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] \geq 0. \tag{2.6}$$

Relations (2.5), (2.6) together imply

$$\nabla f(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] \geq 0,$$

hence it follows from the concavity of $f(\mathbf{x})$ that for all feasible \mathbf{x} the inequality

$$f(\mathbf{x}) - f(\mathbf{x}^*) \geq \nabla f(\mathbf{x}^*)[\mathbf{x} - \mathbf{x}^*] \geq 0 \tag{2.7}$$

holds. □

LEMMA 4 *If the function $F(z_1, \dots, z_m)$ is quasi-concave in a convex set of the m -dimensional space containing all points of the type*

$$(g_1(\mathbf{x}), \dots, g_m(\mathbf{x})), \quad \mathbf{x} \in L,$$

where L is a convex subset of \overline{K} , then the function

$$F(g_1(\mathbf{x}), \dots, g_m(\mathbf{x}))$$

is quasi-concave in L .

PROOF. Let $\mathbf{x}_1 \in L$, $\mathbf{x}_2 \in L$ and $0 \leq \lambda \leq 1$. Then taking into account the monotonicity of F with respect to all variables, we have

$$\begin{aligned}
&F(g_1(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2), \dots, g_m(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2)) \\
&\geq F(\lambda g_1(\mathbf{x}_1) + (1 - \lambda) g_1(\mathbf{x}_2), \dots, \lambda g_m(\mathbf{x}_1) + (1 - \lambda) g_m(\mathbf{x}_2)) \\
&\geq \min[F(g_1(\mathbf{x}_1), \dots, g_m(\mathbf{x}_1)), F(g_1(\mathbf{x}_2), \dots, g_m(\mathbf{x}_2))].
\end{aligned}$$

This proves the lemma. □

3 Concavity and Quasi-concavity Theorems for Probability Distribution Functions

Let $F(z_1, \dots, z_m)$ be an arbitrary probability distribution function of m variables. Let us subdivide the variables into two disjoint groups and denote by \mathbf{x} and \mathbf{y} , respectively the vectors formed by the variables belonging to these groups. Though \mathbf{x} does not necessarily consist of the variables of z_1, \dots, z_m , we introduce the notation

$$F(z_1, \dots, z_m) = F(\mathbf{x}, \mathbf{y}).$$

We say that $F(\mathbf{x}, \mathbf{y})$ is concave with respect to \mathbf{x} in a convex set E of \mathbb{R}^m in the positive direction if for any pair

$$\mathbf{z}_1 = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{y} \end{pmatrix} \in E, \quad \mathbf{z}_2 = \begin{pmatrix} \mathbf{x}_2 \\ \mathbf{y} \end{pmatrix} \in E,$$

where $\mathbf{x}_1 \leq \mathbf{x}_2$, the function F is concave between \mathbf{z}_1 and \mathbf{z}_2 .

THEOREM 1 *Let E be a closed rectangle in \mathbb{R}^m , i.e., the Cartesian product of the finite or infinite closed intervals I_1, \dots, I_m . Suppose that $F(z_1, \dots, z_m)$ is concave in E in the positive direction with respect to any subset of the variables containing $m - 1$ elements. Then $F(z_1, \dots, z_m)$ is quasi-concave in E .*

PROOF. Let \mathbf{x} and \mathbf{y} be two disjoint sets of variables of k_1 and k_2 elements, where

$$\begin{aligned} 1 \leq k_i \leq m - 1, \quad i = 1, 2, \\ k_1 + k_2 = m. \end{aligned}$$

Let $\mathbf{z}'_1 = (\mathbf{x}'_1, \mathbf{y}'_1)$, $\mathbf{z}'_2 = (\mathbf{x}'_2, \mathbf{y}'_2)$ be two points of E with the property that $\mathbf{x}'_1 \leq \mathbf{x}'_2$, $\mathbf{y}'_1 \geq \mathbf{y}'_2$. We shall prove that F is concave between \mathbf{z}'_1 and \mathbf{z}'_2 .

In fact, the function $F(\mathbf{x}, \mathbf{y})$ is concave in E in the positive direction with respect to both \mathbf{x} and \mathbf{y} hence if $0 \leq \lambda \leq 1$, we have

$$\begin{aligned} & F(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2, \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2) \\ & \geq \lambda F(\mathbf{x}_1, \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2) + (1 - \lambda) F(\mathbf{x}_2, \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2) \\ & \geq \lambda^2 F(\mathbf{x}_1, \mathbf{y}_1) + \lambda(1 - \lambda) F(\mathbf{x}_1, \mathbf{y}_2) + \lambda(1 - \lambda) F(\mathbf{x}_2, \mathbf{y}_1) + (1 - \lambda)^2 F(\mathbf{x}_2, \mathbf{y}_2). \end{aligned} \tag{3.1}$$

Now $F(\mathbf{x}, \mathbf{y})$ is a probability distribution function; let $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ be random vectors, the joint distribution of which is $F(\mathbf{x}, \mathbf{y})$, i.e.,

$$\mathbf{P}(\boldsymbol{\xi} \leq \mathbf{x}, \boldsymbol{\eta} \leq \mathbf{y}) = F(\mathbf{x}, \mathbf{y}).$$

It is easy to see that

$$\mathbf{P}(\mathbf{x}_1 \leq \boldsymbol{\xi} \leq \mathbf{x}_2, \mathbf{y}_2 \leq \boldsymbol{\eta} \leq \mathbf{y}_1) = F(\mathbf{x}_2, \mathbf{y}_1) - F(\mathbf{x}_1, \mathbf{y}_1) - F(\mathbf{x}_2, \mathbf{y}_2) + F(\mathbf{x}_1, \mathbf{y}_2),$$

thus

$$F(\mathbf{x}_2, \mathbf{y}_1) + F(\mathbf{x}_1, \mathbf{y}_2) \geq F(\mathbf{x}_1, \mathbf{y}_1) + F(\mathbf{x}_2, \mathbf{y}_2). \quad (3.2)$$

Substituting (3.2) into (3.1) we obtain

$$F(\lambda \mathbf{x}_1 + (1 - \lambda) \mathbf{x}_2, \lambda \mathbf{y}_1 + (1 - \lambda) \mathbf{y}_2) \geq \lambda F(\mathbf{x}_1, \mathbf{y}_1) + (1 - \lambda) F(\mathbf{x}_2, \mathbf{y}_2). \quad (3.3)$$

If \mathbf{z}_1 and \mathbf{z}_2 are arbitrary points of E , then either $\mathbf{z}_1 \leq \mathbf{z}_2$ or $\mathbf{z}_2 \leq \mathbf{z}_1$ or \mathbf{z} can be subdivided into two disjoint groups \mathbf{x}, \mathbf{y} as in the beginning of the proof. In the first two cases $F(\mathbf{z})$ is quasi-concave between \mathbf{z}_1 and \mathbf{z}_2 . This follows from the fact that F is a probability distribution function which is non-decreasing in any positive direction. In the third case $F(\mathbf{z})$ is concave between \mathbf{z}_1 and \mathbf{z}_2 as it was proved before. A concave function is also quasi-concave, thus the theorem is proved. \square

We now consider the case of the two-variate normal probability distribution. In the non-degenerate case the probability density is given by

$$\begin{aligned} \phi(x_1, x_2; r) &= -\frac{1}{2(1-r^2)} \left[\frac{(x_1 - m_1)^2}{\sigma_1^2} - 2r \frac{(x_1 - m_1)(x_2 - m_2)}{\sigma_1 \sigma_2} + \frac{(x_2 - m_2)^2}{\sigma_2^2} \right] \\ &= \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1-r^2}} e, \end{aligned} \quad (3.4)$$

where $|r| < 1$. We are considering the case $m_1 = m_2 = 0, \sigma_1 = \sigma_2 = 1$, i.e., the case where the variables have standard normal distribution, r is the correlated coefficient. From now on $\phi(x_1, x_2; r)$ is therefore defined by

$$\phi(x_1, x_2; r) = -\frac{1}{2(1-r^2)} [x_1^2 - 2rx_1x_2 + x_2^2] = \frac{1}{2\pi\sqrt{1-r^2}} e. \quad (3.5)$$

The probability distribution function belonging to the frequency function (3.5) is defined by

$$\Phi(x_1, x_2; r) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \phi(t_1, t_2; r) dt_2 dt_1. \quad (3.6)$$

We adopt the notation $\Phi(x_1, x_2; r)$ for the joint distribution of ξ_1, ξ_2 also in the case when $|r| = 1$ and ξ_1, ξ_2 have standard normal distribution.

THEOREM 2 *If $r \geq 0$ then $\Phi(x_1, x_2; r)$ is quasi-concave in the nonnegative orthant $x_1 \geq 0, x_2 \geq 0$.*

PROOF. We apply Theorem 1 in such a way that we prove that $\Phi(x_1, x_2; r)$ is concave with respect to x_1 if x_2 is fixed and with respect to x_2 if x_1 is fixed in the nonnegative orthant.

For continuous distribution functions $F(x_1, x_2)$ it is well-known that

$$\frac{\partial F(x_1, x_2)}{\partial x_1} = F(x_2 | x_1) f_1(x_1), \quad (3.7)$$

where $F(x_2 | x_1)$ is the conditional probability distribution function of ξ_2 given that $\xi_1 = x_1$ and $f_1(x)$ is the probability density function of ξ_1 . Applying this to $\Phi(x_1, x_2; r)$, we obtain

$$\frac{\partial \Phi(x_1, x_2; r)}{\partial x_1} = \Phi(x_2 | x_1) \phi(x_1), \quad (3.8)$$

where

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

It is also well-known that the conditional distribution function $\Phi(x_2 | x_1)$ is given by the following formula

$$\Phi(x_2 | x_1) = \Phi\left(\frac{x_2 - rx_1}{\sqrt{1-r^2}}\right), \quad (3.9)$$

where

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt.$$

Thus from (3.8) and (3.9) we obtain

$$\frac{\partial \Phi(x_1, x_2; r)}{\partial x_1} = \Phi\left(\frac{x_2 - rx_1}{\sqrt{1-r^2}}\right) \phi(x_1). \quad (3.10)$$

Taking the second derivative with respect to x_1 , we get

$$\frac{\partial^2 \Phi(x_1, x_2; r)}{\partial x_1^2} = \phi\left(\frac{x_2 - rx_1}{\sqrt{1-r^2}}\right) \frac{-r}{\sqrt{1-r^2}} \phi(x_1) - x_1 \phi(x_1) \Phi\left(\frac{x_2 - rx_1}{\sqrt{1-r^2}}\right). \quad (3.11)$$

Since $r \geq 0$ and $x_1 \geq 0$, the second derivative with respect to x_1 is nonpositive. Similarly

$$\frac{\partial^2 \Phi(x_1, x_2; r)}{\partial x_2^2} \leq 0 \quad \text{if } x_2 \geq 0$$

thus if $x_1 \geq 0$, $x_2 \geq 0$, the function $\Phi(x_1, x_2; r)$ is concave in both variables. This completes the proof. \square

THEOREM 3 *Let $-1 < r \leq 0$. Then $\Phi(x_1, x_2; r)$ is concave in x_i as a function of one variable if*

$$x_i \geq \sqrt{\frac{\phi(1)}{2\Phi(1) + \phi(1)}}, \quad i = 1, 2, . \quad (3.12)$$

The lower bound is exact in the sense that the smallest α with the property that for every r satisfying $-1 < r \leq 0$, $\Phi(x_1, x_2; r)$ is concave with respect to x_1 for a fixed x_2 and with respect to x_2 for a fixed x_1 , provided $x_1 \geq \alpha$, $x_2 \geq \alpha$, is equal to the right hand side of (3.12).

PROOF. Let us introduce the notation $\rho = -r$. To prove the first part of the theorem we consider the right hand side of (3.11) and show that it is non-positive whenever (3.12) holds. For this it is necessary and sufficient that the inequality

$$\frac{\Phi\left(\frac{x_2 + \rho x_1}{\sqrt{1 - \rho^2}}\right)}{\phi\left(\frac{x_2 + \rho x_1}{\sqrt{1 - \rho^2}}\right)} x_1 \geq \frac{\rho}{\sqrt{1 - \rho^2}} \quad (3.13)$$

holds for (3.12). Continuing this, (3.13) is certainly satisfied for $x_1 \geq \alpha$, $x_2 \geq \alpha$ if we have

$$\frac{\Phi\left(\alpha \frac{1 + \rho}{\sqrt{1 - \rho^2}}\right)}{\phi\left(\alpha \frac{1 + \rho}{\sqrt{1 - \rho^2}}\right)} \alpha \geq \frac{\rho}{\sqrt{1 - \rho^2}}. \quad (3.14)$$

Let us introduce the notation

$$\sqrt{\frac{1 + \rho}{1 - \rho}} = u.$$

Then $u \geq 1$ and

$$\rho = \frac{u^2 - 1}{u^2 + 1}, \quad \frac{\rho}{1 + \rho} = \frac{u^2 - 1}{2u^2}.$$

Thus if we multiply both sides of (3.14) by $\sqrt{1 - \rho^2}/(1 + \rho) = 1/u$, we obtain

$$\frac{\Phi(\alpha u)}{\phi(\alpha u)} \frac{\alpha}{u} \geq \frac{u^2 - 1}{2u^2}. \quad (3.15)$$

If $\alpha = 0$, then (3.15) does not hold for every $u \geq 1$. Let $\alpha > 0$ and introduce the notation $v = \alpha u$. Then (3.15) is equivalent to

$$\alpha^2 \geq \frac{v^2}{2 \frac{\Phi(v)}{\phi(v)} v + 1}, \quad (3.16)$$

in the sense that (3.16) must be satisfied for every $v \geq \alpha$. The smallest α satisfying this is

$$\alpha = \sqrt{\frac{\phi(1)}{2\Phi(1) + \phi(1)}}. \quad (3.17)$$

In fact the function of the variable $v > 0$

$$g(v) = \frac{1}{v^2} \left[2 \frac{\Phi(v)}{\phi(v)} v + 1 \right] \quad (3.18)$$

is strictly decreasing if $0 < v \leq 1$ and is strictly increasing if $v \geq 1$, since

$$g'(v) = \frac{2(v^2 - 1)}{v^2\phi(v)}[\phi(v) + v\Phi(v)], \quad v > 0. \quad (3.19)$$

Thus $1/g(v)$ has a maximum at $v = 1$ if $v \geq 0$. If we choose $\alpha = 1/g(1)$, then $\alpha < 1$ and (3.16) will be satisfied for $v \geq \alpha$. That α cannot be decreased, is obvious. This proves the first part of the theorem.

As the smallest difference between the left hand side and the right hand side in (3.13) is attained if $x_1 = x_2 = \alpha$ and according to our proof this smallest difference is 0 if and only if α has the value (3.17), the remaining part of the theorem is also true. \square

REMARK 2 To every ρ ($0 \leq \rho < 1$) or in other terms to every u ($u \geq 1$) there corresponds a unique smallest α such that $\Phi(x_1, x_2; r)$ is concave in x_1 and in x_2 if $x_1 \geq \alpha$, $x_2 \geq \alpha$. This α can be obtained from (3.15) if we choose α so that the equality hold or from (3.16) by substituting $\alpha = v/u$ on the left hand side and choose v so that the equality hold. It is easy to check that there is one and only one v of this kind. In fact, we seek for a v satisfying

$$u^2 = 2\frac{\Phi(v)}{\phi(v)}v + 1 \quad (3.20)$$

and the continuous function on the right hand side is equal to 1 if $v = 0$, strictly increasing for $v > 0$ and tends to infinity if $v \rightarrow \infty$. Now with this v we have the corresponding α by

$$\alpha = \frac{v}{\sqrt{2\frac{\Phi(v)}{\phi(v)}v + 1}}. \quad (3.21)$$

The largest α (i.e., the worst case) belongs to $v = 1$ in which case (3.21) coincides with (3.17). Then

$$r = -\rho = -\frac{u^2 - 1}{u^2 + 1} = -\frac{\Phi(1)}{\Phi(1) + \phi(1)} \approx 0.7766. \quad (3.22)$$

REMARK 3 If ξ_1 and ξ_2 have standard normal distribution where $|r| = 1$, i.e., the joint distribution is degenerated, then the joint distribution

$$\mathbf{P}(\xi_1 \leq x_1, \xi_2 \leq x_2)$$

is concave in the nonnegative orthant. In fact, if $r = 1$, then $\xi_1 = \xi_2$ with probability 1, hence

$$\Phi(x_1, x_2; 1) = \mathbf{P}(\xi_1 \leq x_1, \xi_2 \leq x_2) = \mathbf{P}(\xi_1 \leq \min(x_1, x_2)) = \Phi(\min(x_1, x_2)).$$

If $x_1 \geq 0$, $x_2 \geq 0$, then the function $\min(x_1, x_2)$ is concave. Since $\Phi(z)$ is an increasing concave function of the variable $z \geq 0$, it follows that

$$\Phi(\min(x_1, x_2))$$

is concave in the nonnegative orthant. Let now $r = -1$. Then $\xi_2 = -\xi_1$ with probability 1. Thus we have for $x_1 \geq 0, x_2 \geq 0$,

$$\begin{aligned}\Phi(x_1, x_2; -1) &= \mathbf{P}(\xi_1 \leq x_1, \xi_2 \leq x_2) \\ &= \mathbf{P}(-x_2 \leq \xi_1 \leq x_1) = \Phi(x_1) - \Phi(-x_2) = \Phi(x_1) + \Phi(x_2) - 1.\end{aligned}$$

Here the right hand side is concave in the nonnegative orthant and the assertion follows.

THEOREM 4 *If ξ_1, ξ_2 have an arbitrary joint normal distribution where ξ_1 and ξ_2 have expectation 0 and variance 1, then the set of points x_1, x_2 satisfying the inequality*

$$\Phi(x_1, x_2; r) \geq p \tag{3.23}$$

is convex if p is a fixed probability for which $p < 1$ and the following inequality

$$p \geq \Phi\left(\sqrt{\frac{\phi(1)}{2\Phi(1) + \phi(1)}}\right) \approx 0.6387. \tag{3.24}$$

PROOF. It follows from (3.23) that

$$\begin{aligned}\Phi(x_1) &\geq \Phi(x_1, x_2; r) \geq p, \\ \Phi(x_2) &\geq \Phi(x_1, x_2; r) \geq p,\end{aligned}$$

thus by (3.23) we have

$$\begin{aligned}x_1 &\geq \sqrt{\frac{\phi(1)}{2\Phi(1) + \phi(1)}}, \\ x_2 &\geq \sqrt{\frac{\phi(1)}{2\Phi(1) + \phi(1)}}.\end{aligned} \tag{3.25}$$

By Theorems 1, 2, 3 and Remark 3, the function $\Phi(x_1, x_2; r)$ is quasi-concave in the region (3.25). \square

THEOREM 5 *If ξ_1, ξ_2 have a joint normal distribution with $0 \leq r \leq 1$ or $r = -1$, where the random variables ξ_1, ξ_2 are standardized, then the set of points*

$$\Phi(x_1, x_2; r) \geq p \tag{3.26}$$

is convex where p is a fixed probability satisfying the inequality

$$\frac{1}{2} \leq p < 1.$$

PROOF. Since

$$\begin{aligned}\Phi(x_1) &\geq \Phi(x_1, x_2; r) \geq p \geq \frac{1}{2}, \\ \Phi(x_2) &\geq \Phi(x_1, x_2; r) \geq p \geq \frac{1}{2},\end{aligned}$$

it follows that $x_1 \geq 0, x_2 \geq 0$. By Theorems 1, 2 and Remark 3, the function $\Phi(x_1, x_2; r)$ is quasi-concave in the nonnegative orthant, hence the theorem. \square

We can give a convexity theorem which is not universal in the correlation coefficient but depends on it.

THEOREM 6 *If ξ_1, ξ_2 have a joint normal distribution, with $-1 < r \leq 0$, where ξ_1, ξ_2 are standardized, then the set of points*

$$\Phi(x_1, x_2; r) \geq p \tag{3.27}$$

is convex where p is a fixed probability satisfying the inequality

$$\Phi \left(\frac{v}{\sqrt{2 \frac{\Phi(v)}{\phi(v)} v + 1}} \right) \leq p < 1. \tag{3.28}$$

Here v is defined by (3.20).

The proof of this theorem is similar to that of Theorem 4.

4 Algorithm for the Solution of Problem (1.2)

Let us introduce the notation

$$G(\mathbf{x}) = \mathbf{P}(g_i(\mathbf{x}) \geq \beta_i, \quad i = 1, \dots, m). \tag{4.1}$$

Problem (1.2) can be formulated with this notation in the following manner

$$\begin{aligned} & \min f(\mathbf{x}) \\ & \text{subject to} \\ & G(\mathbf{x}) \geq p, \\ & \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M. \end{aligned} \tag{4.2}$$

For the solution of this problem we apply Zoutendijk's procedure (Procedure P 2 in [7], p. 74.) Only the formal procedure is the same, the reason why the procedure solves the problem is different and the convergence proof, uses ideas taken from Zoutendijk's work.

Let δ be an arbitrary but throughout the procedure fixed positive number. Let \mathbf{x}_1 be an arbitrary point of the set of feasible solutions. We assume that we already know one and shall use such a point as an initial solution. Suppose we already obtained $\mathbf{x}_1, \dots, \mathbf{x}_k$, where all of them are feasible vectors. Define the linear programming problem

$$\begin{aligned} & \min y \\ & \text{subject to} \\ & G(\mathbf{x}_k) + \nabla G(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) + \delta y \geq p, \\ & \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M, \\ & \nabla f(\mathbf{x}_k)(\mathbf{x} - \mathbf{x}_k) \leq y. \end{aligned} \tag{4.3}$$

The set defined by these constraints is non-empty because it contains e.g., $\mathbf{x} = \mathbf{x}_k, y = 0$. On the other hand the set of \mathbf{x} vectors determined by the linear constraints is bounded by supposition. Thus y is bounded from below hence Problem (4.3) always has a finite optimum.

If $\mathbf{x}_k, y = 0$ is an optimal solution to problem (4.3), then the procedure terminates. If $\mathbf{x}_k, y = 0$ is not optimal and \mathbf{x}_k^* is an optimal solution to problem (4.3), then we consider the halfline

$$\mathbf{x}_k + \lambda(\mathbf{x}_k^* - \mathbf{x}_k), \quad \lambda \geq 0 \quad (4.4)$$

and minimize $f(\mathbf{x})$ on that segment of this halfline which belongs to the set of feasible solutions. In other words, let μ_k be the largest λ satisfying

$$\begin{aligned} G(\mathbf{x}_k + \lambda(\mathbf{x}_k^* - \mathbf{x}_k)) &\geq p, \\ \mathbf{a}'_i(\mathbf{x}_k + \lambda(\mathbf{x}_k^* - \mathbf{x}_k)) &\geq b_i, \quad i = 1, \dots, M \end{aligned} \quad (4.5)$$

and let λ_k be a λ with the property that

$$f(\mathbf{x}_k + \lambda(\mathbf{x}_k^* - \mathbf{x}_k)) \geq f(\mathbf{x}_k + \lambda_k(\mathbf{x}_k^* - \mathbf{x}_k)), \quad \text{for } 0 \leq \lambda \leq \mu_k. \quad (4.6)$$

Then we define \mathbf{x}_{k+1} by the following equality

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k(\mathbf{x}_k^* - \mathbf{x}_k). \quad (4.7)$$

In the first case where we stop at \mathbf{x}_k , \mathbf{x}_k will be an optimal solution to problem (1.2). If, on the other hand, the procedure is infinite, then the sequence $f(\mathbf{x}_k)$ will converge to the absolute minimum of $f(\mathbf{x})$ over the set of feasible solutions of problem (1.2). These will be proved under certain conditions in the next section.

Consider now the case $m = 2$. The gradient of $G(\mathbf{x})$ can be obtained easily if the joint distribution of β_1 and β_2 is normal. In this case we may suppose that β_1 and β_2 are standardized because otherwise instead of the event

$$\begin{aligned} g_1(\mathbf{x}) &\geq \beta_1, \\ g_2(\mathbf{x}) &\geq \beta_2, \end{aligned}$$

we would consider its equivalent form

$$\begin{aligned} \frac{g_1(\mathbf{x}) - E(\beta_1)}{D(\beta_2)} &\geq \frac{\beta_1 - E(\beta_1)}{D(\beta_1)}, \\ \frac{g_2(\mathbf{x}) - E(\beta_2)}{D(\beta_2)} &\geq \frac{\beta_2 - E(\beta_2)}{D(\beta_2)}, \end{aligned}$$

where $E(\beta_i)$ is the expectation and $D(\beta_i)$ is the dispersion of the random variable β_i . Now if r is the correlation coefficient of β_1 and β_2

$$r = \frac{E(\beta_1\beta_2)}{D(\beta_1)D(\beta_2)} \quad (4.8)$$

and $|r| < 1$, then the joint distribution function of β_1 and β_2 equals

$$\Phi(z_1, z_2; r) = \int_{-\infty}^{z_1} \int_{-\infty}^{z_2} \phi(u_1, u_2; r) du_2 du_1, \quad (4.9)$$

where $\phi(u_1, u_2; r)$ is given by (3.5), and has a continuous gradient in the entire two-dimensional space. Thus, if at a point \mathbf{x} the vectors $\nabla g_1(\mathbf{x})$, $\nabla g_2(\mathbf{x})$ exist, then $\nabla G(\mathbf{x})$ also exists and using (3.10), we have for

$$G(\mathbf{x}) = \Phi(g_1(\mathbf{x}), g_2(\mathbf{x}); r) \quad (4.10)$$

that

$$\begin{aligned} \nabla G(\mathbf{x}) &= \left(\frac{\partial \Phi}{\partial z_1} \frac{\partial g_1}{\partial x_i} + \frac{\partial \Phi}{\partial z_2} \frac{\partial g_2}{\partial x_i}, \quad i = 1, \dots, n \right) \\ &= \left(\Phi \left(\frac{g_2(\mathbf{x}) - r g_1(\mathbf{x})}{\sqrt{1-r^2}} \right) \phi(g_1(\mathbf{x})) \frac{\partial g_1(\mathbf{x})}{\partial x_i} \right. \\ &\quad \left. + \Phi \left(\frac{g_1(\mathbf{x}) - r g_2(\mathbf{x})}{\sqrt{1-r^2}} \right) \phi(g_2(\mathbf{x})) \frac{\partial g_2(\mathbf{x})}{\partial x_i}, \quad i = 1, \dots, n \right). \end{aligned} \quad (4.11)$$

Thus

$$\begin{aligned} \nabla G(\mathbf{x}) &= \Phi \left(\frac{g_2(\mathbf{x}) - r g_1(\mathbf{x})}{\sqrt{1-r^2}} \right) \phi(g_1(\mathbf{x})) \nabla g_1(\mathbf{x}) \\ &\quad + \Phi \left(\frac{g_1(\mathbf{x}) - r g_2(\mathbf{x})}{\sqrt{1-r^2}} \right) \phi(g_2(\mathbf{x})) \nabla g_2(\mathbf{x}). \end{aligned} \quad (4.12)$$

The functions $\Phi(z)$, $\phi(z)$, $\Phi(z_1, z_2; r)$ are tabulated (see [4], [5]) which helps in the practical application. Problem (4.3) reads now as

$$\begin{aligned} &\min y \\ &\text{subject to} \\ &\Phi(g_1(\mathbf{x}_k), g_2(\mathbf{x}_k); r) \\ &= \Phi \left(\frac{g_2(\mathbf{x}_k) - r g_1(\mathbf{x}_k)}{\sqrt{1-r^2}} \right) \phi(g_1(\mathbf{x}_k)) \nabla g_1(\mathbf{x}_k) (\mathbf{x} - \mathbf{x}_k) \\ &\quad + \Phi \left(\frac{g_1(\mathbf{x}_k) - r g_2(\mathbf{x}_k)}{\sqrt{1-r^2}} \right) \phi(g_2(\mathbf{x}_k)) \nabla g_2(\mathbf{x}_k) (\mathbf{x} - \mathbf{x}_k) + \delta y \geq p, \\ &\mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M, \\ &\nabla f(\mathbf{x}_k) (\mathbf{x} - \mathbf{x}_k) \leq y. \end{aligned} \quad (4.13)$$

If the random variables β_1, \dots, β_m have a joint distribution different from the normal and we denote it by $F(z_1, \dots, z_m)$, then the procedure is defined in the same way.

The convergence of the procedure will be proved in the next sections. Here we prove only a theorem stating the optimality of \mathbf{x}_k if the procedure terminates with it.

THEOREM 7 *Under the assumptions of Section 1, the optimum value y_{opt} of problem (4.3) is 0 if and only if \mathbf{x}_k is an optimal solution to problem (1.2).*

PROOF. Suppose that in problem (4.3) we have $y_{\text{opt}} = 0$. In this case for all vectors \mathbf{x} , y satisfying the constraints of (4.3) the inequality $y \geq 0$ holds. Let $G(\mathbf{x}_k) = p$. Then

$$y \geq 0 \tag{4.14}$$

is a consequence of the system of homogeneous linear inequalities (of the variables \mathbf{z} , y):

$$\begin{aligned} \nabla G(\mathbf{x}_k)\mathbf{z} + \delta y &\geq 0, \\ \mathbf{a}'_i \mathbf{z} &\geq 0, \quad i \in I(\mathbf{x}_k), \\ -\nabla f(\mathbf{x}_k)\mathbf{z} + y &\geq 0. \end{aligned} \tag{4.15}$$

Hence by the theorem of Farkas there exist nonnegative numbers u , w , u_i , $i \in I(\mathbf{x}_k)$ such that

$$\begin{aligned} u\nabla G(\mathbf{x}_k) + \sum_{i \in I(\mathbf{x}_k)} u_i \mathbf{a}'_i - w\nabla f(\mathbf{x}_k) &= \mathbf{0}', \\ u\delta + w &= 1. \end{aligned} \tag{4.16}$$

By Lemma 1, $w = 0$ cannot hold because in this case we would have $u = 0$ which contradicts the second row in (4.16). Dividing by w everywhere in the first equality of (4.16) we obtain the Kuhn–Tucker conditions. Thus by Lemma 3, \mathbf{x}_k is an optimal solution to problem (1.2). In the case where $G(\mathbf{x}_k) > p$, the proof is quite similar. The difference is that $y \geq 0$ is a consequence of the conditions in the last two rows of (4.15) and $u = 0$ in both rows in (4.16).

Suppose now that \mathbf{x}_k is an optimal solution to problem (1.2). Then the Kuhn–Tucker conditions (2.4) hold. Thus there exist nonnegative numbers u , w , u_i , $i \in I(\mathbf{x})$ such that (4.16) is valid. Here $u = 0$ if $G(\mathbf{x}_k) > p$. Thus (4.14) is a consequence of the inequalities in (4.15) (the first is to be omitted if $u = 0$). This tells us that $y_{\text{opt}} \geq 0$ in problem (4.3). The vector $\mathbf{x} = \mathbf{x}_k$, $y = 0$ is, however, feasible in (4.3) hence $y_{\text{opt}} = 0$. \square

5 Auxiliary Lemma for the Convergence of the Procedure

In this section we prove two lemmas. The notations used here are independently used of the other sections. We consider a closed bounded convex set K and a function $F(\mathbf{x})$ defined in an open set containing K . We suppose that $F(\mathbf{x})$ has a continuous gradient in its domain of definition.

LEMMA 5 *Let $\mathbf{y}_1, \mathbf{y}_2, \dots$ be a sequence of points of K and $\mathbf{t}_1, \mathbf{t}_2, \dots$ be a bounded sequence of vectors. Let further $\gamma_1, \gamma_2, \dots$ be a sequence of positive numbers and suppose that*

$$\mathbf{y}_k + \gamma_k \mathbf{t}_k \in K, \quad 0 \leq \gamma \leq \gamma_k, \quad k = 1, 2, \dots; \tag{5.1}$$

there exists an $\varepsilon > 0$ such that

$$\nabla F(\mathbf{y}_k)\mathbf{t}_k \geq \varepsilon, \quad k = 1, 2, \dots \tag{5.2}$$

Let $0 < \varepsilon_1 < \varepsilon$ and suppose that $\gamma_k \rightarrow 0$ if $k \rightarrow \infty$. Then except for at most a finite number of subscripts k we have

$$\nabla F(\mathbf{y}_k + \gamma \mathbf{t}_k) \mathbf{t}_k \geq \varepsilon_1 \quad \text{if} \quad 0 \leq \gamma \leq \gamma_k. \quad (5.3)$$

PROOF. Contrary to (5.3) let us suppose that we have for infinitely many subscripts k

$$\nabla F(\mathbf{y}_k + \gamma'_k \mathbf{t}_k) \mathbf{t}_k < \varepsilon_1 \quad \text{with} \quad 0 < \gamma'_k \leq \gamma_k. \quad (5.4)$$

Then by (5.2) and (5.4) we have for this infinitely many subscripts

$$[\nabla F(\mathbf{y}_k) - \nabla F(\mathbf{y}_k + \gamma'_k \mathbf{t}_k)] \mathbf{t}_k \geq \varepsilon - \varepsilon_1 > 0.$$

This is, however, a contradiction because $\gamma'_k \rightarrow 0$ if $k \rightarrow \infty$, the sequence \mathbf{t}_k is bounded and $\nabla F(\mathbf{x})$ is uniformly bounded in K . \square

LEMMA 6 Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of points of K , $\mathbf{s}_1, \mathbf{s}_2, \dots$ a bounded sequence of vectors and $\lambda_1, \lambda_2, \dots$ a sequence of positive numbers such that

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \lambda_k \mathbf{s}_k, \quad \mathbf{x}_k + \lambda \mathbf{s}_k \in K, \quad 0 \leq \lambda \leq \lambda_k, \quad k = 1, 2, \dots \quad (5.5)$$

Suppose that

$$F(\mathbf{x}_{k+1}) = F(\mathbf{x}_k + \lambda_k \mathbf{s}_k) \geq F(\mathbf{x}_k + \lambda \mathbf{s}_k), \quad 0 \leq \lambda \leq \lambda_k, \quad k = 1, 2, \dots \quad (5.6)$$

Let $\mathbf{y}_1, \mathbf{y}_2, \dots; \mathbf{t}_1, \mathbf{t}_2, \dots; \gamma_1, \gamma_2, \dots$ be subsequences of the above sequences formed by the selection of corresponding elements. Suppose that there exists an $\varepsilon > 0$ such that

$$\nabla F(\mathbf{y}_i) \mathbf{t}_i \geq \varepsilon, \quad i = 1, 2, \dots \quad (5.7)$$

Then we have

$$\sum_{i=1}^{\infty} \gamma_i < \infty. \quad (5.8)$$

PROOF. Consider a fixed k and let i be that subscript for which $\mathbf{x}_k = \mathbf{y}_i$. Then

$$\begin{aligned} F(\mathbf{x}_{k+1}) - F(\mathbf{x}_1) &= \sum_{j=1}^k [F(\mathbf{x}_{j+1}) - F(\mathbf{x}_j)] \\ &= \sum_{j=1}^k [F(\mathbf{x}_j + \lambda_j \mathbf{s}_j) - F(\mathbf{x}_j)] \\ &\geq \sum_{r=1}^i [F(\mathbf{y}_r + \gamma_r \mathbf{t}_r) - F(\mathbf{y}_r)] \\ &\geq \sum_{r=1}^i [F(\mathbf{y}_r + \gamma'_r \mathbf{t}_r) - F(\mathbf{y}_r)], \end{aligned} \quad (5.9)$$

where γ'_r is the largest γ with the property that

$$\begin{aligned} \gamma &\leq \gamma_r, \\ \nabla F(\mathbf{y}_r + \gamma \mathbf{t}_r) \mathbf{t}_r &\geq \varepsilon_1, \end{aligned} \tag{5.10}$$

where $0 < \varepsilon_1 < \varepsilon$. Continuing (5.9),

$$\begin{aligned} \sum_{r=1}^i [F(\mathbf{y}'_r + \gamma'_r \mathbf{t}_r) - F(\mathbf{y}_r)] &= \sum_{r=1}^i F(\mathbf{y}_r + h_r \gamma'_r \mathbf{t}_r) \gamma'_r \mathbf{t}_r \\ &\geq \varepsilon_1 \sum_{r=1}^i \gamma'_r, \end{aligned} \tag{5.11}$$

where $0 < h_r < 1$. It follows from this that

$$\sum_{r=1}^{\infty} \gamma'_r < \infty. \tag{5.12}$$

Now we prove that $\gamma_r = \gamma'_r$ except for at most a finite number of subscripts. In fact, if for an r we have $\gamma'_r < \gamma_r$, then

$$\nabla F(\mathbf{y}_r + \gamma'_r \mathbf{t}_r) \mathbf{t}_r = \varepsilon_1$$

and

$$[\nabla F(\mathbf{y}_r) - \nabla F(\mathbf{y}_r + \gamma'_r \mathbf{t}_r)] \mathbf{t}_r \geq \varepsilon - \varepsilon_1 > 0.$$

This cannot hold infinitely many times because it contradicts to the uniform continuity of $\nabla F(\mathbf{x})$ in K . Thus the Lemma is proved. \square

6 Convergence of the Procedure

We give a proof for the convergence of the procedure for the solution of problem (4.2) where we may disregard the special meaning of the function $G(\mathbf{x})$. We use only the assumption that $G(\mathbf{x})$ has continuous gradient in K of Section 1 and that $G(\mathbf{x})$ is quasi-concave in K . Of course A.6. is also maintained. We consider the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ generated by the subsequent solutions of problem (4.3) in the described manner. If this sequence is finite, i.e., $y_{\text{opt}} = 0$ at the problem (4.3), then by the Theorem 7, \mathbf{x}_k is an optimal solution to problem (1.2). If, however, this sequence is infinite, then, as the set of feasible solutions is bounded, the sequence $\mathbf{x}_1, \mathbf{x}_2, \dots$ has a convergent subsequence what we denote by $\mathbf{y}_1, \mathbf{y}_2, \dots$. Let

$$\mathbf{y}^* = \lim_{k \rightarrow \infty} \mathbf{y}_k \tag{6.1}$$

and consider the (4.3) type problem belonging to \mathbf{y}^* , i.e., the problem

$$\begin{aligned} &\min y \\ &\text{subject to} \\ &\nabla G(\mathbf{y}^*) + \nabla G(\mathbf{y}^*)(\mathbf{x} - \mathbf{y}^*) + \delta y \geq p, \\ &\mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M, \\ &\nabla f(\mathbf{y}^*)(\mathbf{x} - \mathbf{y}^*) \leq y. \end{aligned} \tag{6.2}$$

If in (5.2) we have $y_{\text{opt}} = 0$, then \mathbf{y}^* is an optimal solution to problem (4.2). Thus, using an indirect proof, we suppose that in (5.2) $y_{\text{opt}} = -\delta < 0$. In this case there exists⁵ a neighborhood $N(\mathbf{y}^*)$ of \mathbf{y}^* such that if $\mathbf{z} \in N(\mathbf{y}^*)$, then the corresponding $y_{\text{opt}} \leq -\delta/2$.

Let $\mathbf{s}_k = \mathbf{x}_k^* - \mathbf{x}_k$ (see equality (4.7)) and let \mathbf{t}_k and γ_k be a subsequences of the sequences \mathbf{s}_k, λ_k , respectively, selected in the same way as \mathbf{y}_k is selected from \mathbf{x}_k . If k is large enough then $\mathbf{y}_k \in N(\mathbf{y}^*)$ hence by the last row of (6.2),

$$-\nabla f(\mathbf{y}_k)\mathbf{t}_k \geq \frac{\delta}{2}. \quad (6.3)$$

According to Lemma 6,

$$\sum_{k=1}^{\infty} \gamma_k < \infty \quad (6.4)$$

which implies

$$\gamma_k \rightarrow 0, \text{ if } k \rightarrow \infty.$$

Thus by Lemma 5, we have for every δ_1 , satisfying $\mathbf{y}_k + \gamma\mathbf{t}_k, \gamma \geq 0$.

$$-\nabla f(\mathbf{y}_k + \gamma\mathbf{t}_k)\mathbf{t}_k \geq \delta_1, \quad \text{if } 0 \leq \gamma \leq \gamma_k, \quad (6.5)$$

except for at most a finite number of subscripts k . It follows from (6.5) that going ahead from \mathbf{y}_k in the direction \mathbf{t}_k , the function f decreases still at the point $\mathbf{y}_k + \gamma\mathbf{t}_k$, hence the constraints are those which put a stop to the progress in the direction \mathbf{t}_k . Consider the constraints of problem (1.2) along the halfline

$$\mathbf{y}_k + \gamma\mathbf{t}_k, \quad \gamma \geq 0,$$

i.e.,

$$\begin{aligned} G(\mathbf{y}_k + \gamma\mathbf{t}_k) &\geq p, \\ \mathbf{a}'_i(\mathbf{y}_k + \gamma\mathbf{t}_k) &\geq b_i, \quad i \in I. \end{aligned} \quad (6.6)$$

The second row allows also $\gamma = 1$ by construction of problem (4.3) thus it is the first row of (6.6) which is violated first, going ahead with λ , except for at most a finite number of subscripts. The reason for this is the relation (6.4)

Thus we have the following equality:

$$G(\mathbf{y}_k + \gamma\mathbf{s}_k) = p, \quad \text{if } k \geq k_0. \quad (6.7)$$

We also know that

$$G(\mathbf{y}_k) \geq p. \quad (6.8)$$

⁵The existence of such a neighborhood $N(\mathbf{y}^*)$ can be proved easily under the condition that either $\nabla G(\mathbf{y}^*)$ is not a zero vector or it is a zero vector but $G(\mathbf{y}^*) > p$. If on the other hand $\nabla G(\mathbf{y}^*)$ is the zero vector and $G(\mathbf{y}^*) = p$, then \mathbf{y}^* is an optimal solution to Problem (4.2).

We write now problem (4.3) corresponding to the vector \mathbf{y}_k instead of \mathbf{x}_k . We have

$$\begin{aligned}
& \min y \\
& \text{subject to} \\
& G(\mathbf{y}_k) + \nabla G(\mathbf{y}_k)(\mathbf{x} - \mathbf{y}_k) + \delta y \geq p, \\
& \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i \in I, \\
& \nabla f(\mathbf{y}_k)(\mathbf{x} - \mathbf{y}_k) \leq y.
\end{aligned} \tag{6.9}$$

As the sequence \mathbf{y}_k converges and $G(\mathbf{x})$ is continuous in the set of feasible solutions, it follows that $G(\mathbf{y}_k)$ converges too. In view of (6.7) we have $\lim_{k \rightarrow \infty} G(\mathbf{y}_k) = p$. Inserting the optimal solution in the place of \mathbf{x} in (6.9) and taking into account that

$$\begin{aligned}
-y_{\text{opt}} &\geq \frac{\delta}{2}, \\
p - G(\mathbf{y}_k) &\geq -\varepsilon \text{ for large } k,
\end{aligned}$$

where ε is an arbitrary positive number, it follows that

$$\nabla G(\mathbf{y}_k) \mathbf{t}_k \geq \delta \frac{\delta}{2} - \varepsilon = \varepsilon_1. \tag{6.10}$$

We may choose ε so small that the ε_1 on the right hand side in (6.10) is positive. Since $\gamma_k \rightarrow 0$ if $k \rightarrow \infty$, by Lemma 5 it follows that for any $0 < \varepsilon_2 < \varepsilon_1$,

$$\nabla G(\mathbf{y}_k + \gamma \mathbf{t}_k) \mathbf{t}_k \geq \varepsilon_2, \quad 0 \leq \gamma \leq \gamma_k. \tag{6.11}$$

This contradicts to (6.7) and (6.8) because they imply that at some point between \mathbf{y}_k and $\mathbf{y}_k + \gamma_k \mathbf{t}_k$, the left hand side of (6.11) is equal to 0.

Thus we have proved that

$$\lim_{k \rightarrow \infty} f(\mathbf{x}_k) = \min_{x \in D(p)} f(\mathbf{x}). \tag{6.12}$$

REMARK 4 The present theory applies also to cases where instead of the problem

$$\begin{aligned}
& \min f(\mathbf{x}) \\
& \text{subject to} \\
& G(\mathbf{x}) \geq p, \\
& \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i \in I,
\end{aligned} \tag{6.13}$$

we have to solve the following

$$\begin{aligned}
& \min f(\mathbf{x}) \\
& \text{subject to} \\
& G_i(\mathbf{x}) \geq p_i, \quad i = 1, \dots, N, \\
& \mathbf{a}'_i \mathbf{x} \geq b_i, \quad i = 1, \dots, M,
\end{aligned} \tag{6.14}$$

where the functions $G_1(\mathbf{x}), \dots, G_N(\mathbf{x})$ are quasi-concave. The condition concerning the domain of definition of the functions $G_1(\mathbf{x}), \dots, G_N(\mathbf{x})$ and the more general formulation of condition A.6. is immediate. Similarly, the lemmas of Section 2 and the proof of the convergence of the method applied to problem (6.14) does not present any difficulty.

REMARK 5 If $G(\mathbf{x})$ has the special meaning (4.1) and $m = 2$, then the theorems of Section 3 can be used to ensure the required property of $G(\mathbf{x})$. Only Assumption 6 in the first section cannot be ruled this way but it depends on the special structure of the functions $g_i(\mathbf{x})$.

7 Generalization to an Arbitrary Number of Joint Constraints

In principle one can solve also problems of type (1.2) with $m > 2$. Without a proof we mention that any multivariate normal probability distribution function of standardized variables and nonsingular correlation matrix

$$\Phi(z_1, \dots, z_m; R) \quad (7.1)$$

is quasi-concave in the closed convex set $z_i \geq d$, $i = 1, \dots, m$ if d is large enough. Let this d be the minimum of the numbers having that property and let p_0 be defined by the equality

$$\Phi(d) = p_0. \quad (7.2)$$

Let $p \geq p_0$, then

$$\{\mathbf{z} : \Phi(z_1, \dots, z_m; R) \geq p\} \subset \{\mathbf{z} : z_i \geq d, i = 1, \dots, m\}, \quad (7.3)$$

moreover, if $p > p_0$, then every point of the set on the left hand side is an internal point of the set on the right hand side. This follows from the inequality

$$\Phi(z_i) \geq \Phi(z_1, \dots, z_m; R) \geq p, \quad i = 1, \dots, m. \quad (7.4)$$

We consider the problem

$$\begin{aligned} & \min f(\mathbf{x}) \\ & \text{subject to} \\ & \mathbf{P}(\mathbf{d}'_i \mathbf{x} \geq \beta_i, i = 1, \dots, m) \geq p, \\ & \mathbf{a}'_i \mathbf{x} \geq b_i, i \in I. \end{aligned} \quad (7.5)$$

Let

$$\mu_i = E(\beta_i), \quad \sigma_i^2 = E[(\beta_i - \mu_i)^2]. \quad (7.6)$$

Then (7.5) can be written as

$$\begin{aligned} & \min f(\mathbf{x}) \\ & \text{subject to} \\ & \Phi\left(\frac{1}{\sigma_1}(\mathbf{d}'_1 \mathbf{x} - \mu_1), \dots, \frac{1}{\sigma_m}(\mathbf{d}'_m \mathbf{x} - \mu_m); R\right) \geq p, \\ & \mathbf{a}'_i \mathbf{x} \geq b_i, i \in I, \end{aligned} \quad (7.7)$$

where R is the correlation matrix of the random variables β_1, \dots, β_m . The solution of problem (7.7) is similar to that of the case $m = 2$. First we mention that (see, e.g. [8])

$$\begin{aligned} \frac{\partial \Phi(z_1, \dots, z_m; R)}{\partial z_i} &= \Phi(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_m \mid z_i) \phi(z_i) \\ &= \Phi \left(\frac{z_1 - r_{1i} z_i}{\sqrt{1 - r_{1i}^2}}, \dots, \frac{z_m - r_{mi} z_i}{\sqrt{1 - r_{mi}^2}}; R_i \right) \phi(z_i), \end{aligned} \quad (7.8)$$

where in the last row the first factor is an $m - 1$ -dimensional normal distribution function with the correlation matrix R_i , having entries

$$\frac{r_{jk} - r_{1j} r_{1k}}{\sqrt{1 - r_{1j}^2} \sqrt{1 - r_{1k}^2}}, \quad j, k = 1, \dots, i - 1, i + 1, \dots, m. \quad (7.9)$$

If

$$\mathbf{d}'_i = (d_{i1}, \dots, d_{in}), \quad i = 1, \dots, m, \quad (7.10)$$

then the gradient of the function

$$G(\mathbf{x}) = \Phi \left(\frac{1}{\sigma_i} (\mathbf{d}'_i \mathbf{x} - \mu_i), \quad i = 1, \dots, m; R \right) \quad (7.11)$$

is equal to

$$G(\mathbf{x}) = \left(\sum_{i=1}^m \frac{\partial \Phi(z_1, \dots, z_m; R)}{\partial z_i} \frac{d_{ij}}{\sigma_i}, \quad j = 1, \dots, n \right), \quad (7.12)$$

where

$$z_i = \frac{1}{\sigma_i} (\mathbf{d}'_i \mathbf{x} - \mu_i), \quad i = 1, \dots, m. \quad (7.13)$$

The linear functions (7.13) can be replaced by concave functions.

8 Application to the Nutrition Problem

Problem (7.5) with the cost function

$$f(\mathbf{x}) = \mathbf{c}'\mathbf{x} \quad (8.1)$$

has a direct application to the nutrition problem. In fact, suppose that the number of nutrients is $m + M$ and the necessary nutrient levels of the first m of them are random variables, i.e., vary from unit to unit in the population for which we want to obtain an optimal nutrition program. These levels are denoted by β_1, \dots, β_m . The remaining levels are constant, i.e., they are the same for each unit of the population. Thus each unit of the population has its own minimum nutrient levels

$$\beta_1, \dots, \beta_m, b_1, \dots, b_M, \quad (8.2)$$

where β_1, \dots, β_m take a certain realization at each unit. The first condition in (7.5) means that at least 100% of the population should get his minimum amounts β_1, \dots, β_m from the first m nutrients while everybody should get the constant minimum levels b_1, \dots, b_M from the remaining nutrients. Unfortunately, in lack of the necessary data, we are unable to present the solution of a realistic problem on this field. The marginal distributions of β_1, \dots, β_m are already investigated but there is no trace in the literature of the correlations between these random variables.

References

- [1] ARROW, K. J. and A. C. ENTHOVEN (1961). Quasi-concave Programming. *Econometrica*, **29**, 779–800.
- [2] CHARNES, A., W. W. COOPER and G. H. SYMONDS (1958). Cost Horizons and Certainly Equivalents: An Approach to Stochastic Programming of Heating Oil, *Management Science*, **4**.
- [3] DANTZIG, G. B. (1963). *Linear Programming and Extensions*, Princeton University Press, Princeton, New Jersey.
- [4] GUPTA, SH. S. (1963). Probability Integrals of Multivariate Normal and Multivariate t , *Annals of Math. Stat.*, **34**, 792–828.
- [5] GUPTA, SH. S. (1963). Bibliography on the Multivariate Normal Integrals and Related Topics, *Annals of Math. Stat.*, **34**, 829–838.
- [6] MILLER, B. L. and H. M. WAGNER (1965). Chance Constrained Programming with Joint Constraints, *Operations Research* **13**, 930–945.
- [7] ZOUTENDIJK, G. (1960). *Methods of Feasible Directions*, Elsevier Publishing Company, Amsterdam.
- [8] WILKS, S. S. (1962). *Mathematical Statistics*, Wiley, New York, London.