A Simple Solution of the van der Waerden Permanent Problem

MORGAN A. HANSON
The Florida State University, Department of Statistics, Tallahassee, FL 32306-3033, U.S.A.

(Received: 22 August 1990; accepted: 18 June 1991)

Abstract. The van der Waerden permanent problem was solved using mainly algebraic methods. A much simpler analytic proof is given using a new concept in optimization theory which may be of importance in the general theory of mathematical programming.

Key words. Invexity, weak type I functions, van der Waerden permanent problem.

1. Introduction

The *permanent* of an $n \times n$ matrix $X \equiv [x_{ij}]$ is defined to be

$$per(X) = \sum_{\sigma \in S_n} x_{1,\sigma(1)} x_{2,\sigma(2)} \cdots x_{n,\sigma(n)} \,.$$

where S_n is the group of permutations of the integers $\{1, 2, 3, \ldots, n\}$.

The matrix X is *doubly stochastic* if all of its elements are nonnegative and each of its rows and columns sum to 1.

Arising from a problem devised by van der Waerden in 1926 [1], it was conjectured that the minimum value of the permanent of a doubly stochastic $n \times n$ matrix is attained when all the elements x_{ij} are equal to $\frac{1}{n}$. The problem is trivial for $n = 2$.

In 1959 Marcus and Newman [2] proved it for $n = 3$; in 1968 Eberlein and Mudholkar [3] proved it for $n = 4$; and in 1969 Eberlein [4] proved it for $n = 5$.

Finally, in 1981 proofs for the general case were found independently by Falikman [5] and Egorychev [6].

All these proofs for $n \geq 3$ are complicated. The proofs are mainly of an algebraic nature. In this paper a much simpler analytic proof is given through the use of a new concept in optimization theory which may be of importance in the theory of mathematical programming since it represents a substantial departure from the usual convexity assumptions in mathematical programming theory.

2. Sufficient Conditions for Optimality

Consider the general problem:

$$\text{Minimize } F(x) \tag{2.1}$$

$$\text{subject to } G(x) \leq 0 \,, \tag{2.2}$$

Journal of Global Optimization **1**: 287–293, 1991

where $x \in C \subset R^n$, F is a differentiable scalar function over C and G is an m-dimensional differentiable vector function over C.

The Kuhn–Tucker necessary conditions for a minimum at u, where u satisfies (2.2), are that there exists a vector $y \in R^m$ such that

$$\nabla F(u) + \nabla y'G(u) = 0\,, \tag{2.3}$$

$$y'G(u) = 0\,, \tag{2.4}$$

$$\text{and } y \geq 0\,, \tag{2.5}$$

where ∇ is the n-dimensional differential operator with respect to u.

Hanson [7] showed that the Kuhn–Tucker conditions are also sufficient for a global minimum if F and G belong to a certain class of functions now known as invex functions. Hanson and Mond [8] generalized this class to a class called Type I functions. In this paper the concept is generalized a little further, and we say that the functions $F(x)$ and $G(x)$ are respectively Weak Type I objective and Weak Type I constraint functions over C respectively at u with respect to the kernel $\eta(x, u)$ if there exist a sequence of n-dimensional vectors $z_{(k)}$, $k = 1, 2, \ldots$, and an n-dimensional function $\eta(x, z_{(k)})$, where u is a limit point of the set $\{z_k\}$, such that for $x \in C$, $z_{(k)} \in C$

$$F(x) - F(u) \geq \lim_{\|u - z_{(k)}\| \to 0} \eta'(x, z_{(k)}) \nabla F(z_{(k)})$$

and

$$-G_i(u) \geq \lim_{\|u - z_{(k)}\| \to 0} \eta'(x, z_{(k)}) \nabla G_i(z_{(k)})\,, \quad i = 1, 2, \ldots, m\,,$$

and it is assumed that the limits exist.

To simplify the notation we shall write z for $z_{(k)}$.

THEOREM 1. *If in problem (2.1)–(2.2), $F(x)$ and $G(x)$ are Weak Type I functions, with bounded second derivatives in a neighborhood of u on the constraint set (2.2) at a feasible point u with a common kernel $\eta(x, u)$, and the Kuhn–Tucker conditions are satisfied at u, then u is a global minimum for the problem.*

Proof. Let u and y be vectors satisfying the Kuhn–Tucker conditions (2.3)–(2.5). By the mean value theorem we have

$$\frac{\partial}{\partial z_j}[F(z) + y'G(z)] = \frac{\partial}{\partial z_j}[F(u) + y'G(u)] + (z - u)'\nabla \frac{\partial}{\partial z_j}[F(u^j) + y'G(u^j)]$$

for some vector u^j between z and u, $j = 1, 2, \ldots, n$,

$$= (z - u)'\nabla \frac{\partial}{\partial z_j}[F(u^j) + y'G(u^j)]\,,\; j = 1, 2, \ldots, n\,.$$

by (2.3). (2.6)

For any feasible x,

$$F(x) - F(u) \geqslant \lim_{\|u-z\| \to 0} \eta'(x, z) \nabla F(z),$$

since F is a Weak Type I objective function

$$= \lim_{\|u-z\| \to 0} \sum_{j=1}^{n} \eta_j(x, z) \frac{\partial}{\partial z_j} [F(z)]$$

$$= \lim_{\|u-z\| \to 0} \sum_{j=1}^{n} \eta_j(x, z)$$

$$\times \left[- \frac{\partial}{\partial z_j} [y'G(z)] + (z-u)'\nabla \frac{\partial}{\partial z_j} [F(u^j) + y'G(u^j)] \right]$$

by (2.6) .

$$= \lim_{\|u-z\| \to 0} \sum_{j=1}^{n} \eta_j(x, z)$$

$$\times \left[- \frac{\partial}{\partial z_j} (y'G(z)) \right]$$ since the second derivatives

of F and G are bounded in a neighborhood of u ,

$$= \lim_{\|u-z\| \to 0} \eta'(x, z)[-\nabla y'G(z)]$$

$\geqslant y'G(u)$, since G is a Weak Type I constraint function, and $y \geqslant 0$,

$= 0$, by (2.4), which proves the theorem.

Hence if a common kernel $\eta(x, u)$ can be found for the objective and constraint functions in the van der Waerden problem at a point u which satisfies the Kuhn–Tucker conditions, then u will be a global minimum for the problem.

3. The van der Waerden Problem

An $n \times n$ doubly stochastic matrix X can be written in the form $X = \Sigma_{i=1}^{n!} \theta_i P_i$ where $\Sigma_{i=1}^{n!} \theta_i = 1$, $\theta_i \geqslant 0$, and P_i is the i-th permutation matrix of order n (See Birkhoff [9]).

So the van der Waerden problem is the following nonlinear programming problem:

$$\text{Minimize } per\left(\sum_{i=1}^{n!} \theta_i P_i \right) \tag{3.1}$$

$$\text{subject to } \sum_{i=1}^{n!} \theta_i = 1 \tag{3.2}$$

$$\text{and } \theta_i \geqslant 0 , \quad i = 1, 2, \ldots , n! . \tag{3.3}$$

$$\text{Write } per\left(\sum_{i=1}^{n!} \theta_i P_i\right) = \Phi(\theta)\,, \tag{3.4}$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_{n!})$.

The scalar function $\Phi(\theta)$ is homogeneous of degree n, and by Euler's theorem for homogeneous functions

$$\theta' \nabla \Phi(\theta) = n \Phi(\theta)\,. \tag{3.5}$$

In place of problem (3.1)–(3.3) we consider the equivalent problem:

$$\text{Minimize } f(\theta) \equiv \Phi(\theta) - C_n\left(\sum_{i=1}^{n!} \theta_i - 1\right) + M_n(\theta_{n!+1} - 1)^2 \tag{3.6}$$

$$\text{subject to } g_1(\theta) \equiv \left(\sum_{i=1}^{n!} \theta_i - 1\right)^2 \leq 0\,, \tag{3.7}$$

$$\text{and } g_{i+1}(\theta) \equiv -\theta_i \leq 0\,, \quad i = 1, 2, \ldots, n!\,, \tag{3.8}$$

where C_n is a constant to be specified later, and the scalar $\theta_{n!+1}$ which obviously is forced to have the value 1 at minimum, is introduced for convenience later. M_n is a large positive constant.

Since $\Sigma_{i=1}^{n!} \theta_i = 1$, by (3.7), the term $-C_n(\Sigma_{i=1}^{n!} \theta_i - 1)$ introduced into the objective function (3.6) has no effect on the minimal value of the objective function, but does affect the value of the derivative of the objective function, which will be required later.

We now redefine θ to be the $(n!+1)$-dimensional vector $(\theta_1, \theta_2, \ldots, \theta_{n!}, \theta_{n!+1})'$, and ∇ to be the $(n!+1)$-dimensional differential operator with respect to θ. The function $\Phi(\theta)$ remains $\Phi(\theta_1, \theta_2, \ldots, \theta_{n!})$.

Let ω be the $(n!+1)$-dimensional vector $(1/n!, 1/n!, \ldots, 1/n!, 1)'$. We then have

$$\begin{aligned} \omega' \nabla f(\omega) &= \omega' \nabla \Phi(\omega) - C_n \sum_{i=1}^{n!} \omega_i + 2M_n \omega_{n!+1}(\omega_{n!+1} - 1) \\ &= 0\,, \end{aligned}$$

where we define C_n to have the value

$$\omega' \nabla \Phi(\omega) = \frac{1}{n!} \sum_{i=1}^{n!} \frac{\partial \Phi(\omega)}{\partial \theta_i} = \frac{\partial \Phi(\omega)}{\partial \theta_i}\,, \quad i = 1, 2, \ldots, n!\,.$$

Problem (3.6)–(3.8) has the form of problem (2.1)–(2.2) where $G(\theta) = (g_1(\theta), g_2(\theta), \ldots, g_{n!+1}(\theta))$. So by theorem 1, if a vector θ can be found which satisfies the Kuhn–Tucker conditions, and if a suitable η can be found, then θ will be a global minimum in problem (3.6)–(3.8). It will be shown that ω is such a vector.

For problem (3.6)–(3.8) the Kuhn–Tucker condition (2.3) at ω is

$$\frac{\partial \Phi(\omega)}{\partial \theta_i} - C_n + \sum_{l=1}^{n!+1} v_l \frac{\partial g_l(\omega)}{\partial \theta_i} = 0 , \quad i = 1, 2, \ldots, n! ,$$

$$\text{and } 2M_n(\omega_{n!+1} - 1) = 0 .$$

That is, respectively,

$$0 + \sum_{l=1}^{n!+1} v_l \frac{\partial g_l(\omega)}{\partial \theta_i} = 0 , \quad i = 1, 2, \ldots, n! , \tag{3.9}$$

$$\text{and } 0 = 0 . \tag{3.10}$$

The Kuhn–Tucker condition (2.4) at ω is

$$v_l g_l(\omega) = 0 , \quad l = 1, 2, \ldots, n! + 1 . \tag{3.11}$$

The Kuhn–Tucker condition (2.5) at ω is

$$v_l \geq 0 , \quad l = 1, 2, \ldots, n! + 1 . \tag{3.12}$$

So all the Kuhn–Tucker conditions (3.9)–(3.12) are satisfied at ω if we put

$$v_l = 0 , \quad l = 1, 2, \ldots, n! + 1 . \tag{3.13}$$

We now define the $(n! + 1)$-dimensional kernel $\eta(\theta, \omega)$. Divide the constraint set (3.7)–(3.8) into two subsets A and B:

$$A = \{\theta \mid f(\theta) - f(\omega) < 0\} ,$$

$$B = \{\theta \mid f(\theta) - f(\omega) \geq 0\} .$$

Define

$$\eta(\theta, z) = \begin{cases} \dfrac{f(\theta) - f(\omega)}{z'\nabla f(z)} \, z & \text{if } \theta \in A , \\ 0 & \text{if } \theta \in B , \end{cases} \tag{3.14}$$

where z is of the form $(1/n + \epsilon,\ 1/n! - \epsilon,\ 1/n!,\ 1/n!, \ldots, 1/n!, 1-\epsilon)$ where $0 < \epsilon < 1/n!$ and $\epsilon \to 0$. So z satisfies the constraints (3.7)–(3.8), and

$$\begin{aligned} z'\nabla f(z) &= z'\nabla\Phi(z) - C_n \sum_{i=1}^{n!} z_i + 2M_n(z_{n!+1} - 1) z_{n!+1} , \\ &= z'\nabla\Phi(z) - \omega'\nabla\Phi(\omega) + 2M_n(-\epsilon)(1 - \epsilon) , \\ &= n(\Phi(z) - \Phi(\omega)) - 2M_n\epsilon(1 - \epsilon) , \text{ by Euler's theorem} , \\ &= n(z - \omega)'\nabla\Phi(z^*) \\ &\qquad - 2M_n\epsilon(1 - \epsilon) , \text{ for some } z^* \text{ between } z \text{ and } \omega , \\ &= \epsilon\left[n\left(\frac{\partial \Phi(z^*)}{\partial z_1} - \frac{\partial \Phi(z^*)}{\partial z_2} \right) - 2M_n(1 - \epsilon) \right] , \end{aligned} \tag{3.15}$$

$$< 0 \text{ for sufficiently large } M_n . \tag{3.16}$$

In A we have

$$(i)\quad \lim_{\|\omega - z\| \to 0} \eta'(\theta, z)\nabla f(z) = \lim_{\|\omega - z\| \to 0} \frac{f(\theta) - f(\omega)}{z'\nabla f(z)} z'\nabla f(z)$$
$$= f(\theta) - f(\omega) .$$

So, in A, $f(\theta)$ is Weak Type I with respect to $\eta(\theta, \omega)$ at ω.

$$(ii)\quad \lim_{\|\omega - z\| \to 0} \eta'(\theta, z)\nabla g_1(z) = \lim_{\|\omega - z\| \to 0} \frac{f(\theta) - f(\omega)}{z'\nabla f(z)} z'\nabla g_1(z) ,$$
$$= \lim_{\|\omega - z\| \to 0} \frac{f(\theta) - f(\omega)}{z'\nabla f(z)} \left(\sum_{i=1}^{n!} z_i\right) 2\left(\sum_{i=1}^{n} z_i - 1\right)$$
$$= 0 ,$$
$$= -g_1(\omega) , \text{ by (3.7)} .$$

So, in A, $g_1(\theta)$ is Weak Type I with respect to $\eta(\theta, \omega)$ at ω.

$$(iii)\quad \lim_{\|\omega - z\| \to 0} \eta'(\theta, z)\nabla g_{i+1}(z) = \lim_{\|\omega - z\| \to 0} \frac{f(\theta) - f(\omega)}{z'\nabla f(z)} z'\nabla g_{i+1}(z) ,$$
$$i = 1, 2, \ldots, n!$$
$$= \lim_{\|\omega - z\| \to 0} \frac{f(\theta) - f(\omega)}{z'\nabla f(z)} (-z_i) , \quad i = 1, 2, \ldots, n! ,$$
$$\leqslant 0, \text{ since } f(\theta) - f(\omega) < 0 , \; z'\nabla f(z) < 0$$
$$\text{and } - z_i \leqslant 0 , \quad i = 1, 2, \ldots, n! ,$$
$$\leqslant -g_{i+1}(\omega) , \text{ by (3.8)} .$$

So, in A, $g_{i+1}(\theta)$ $i = 1, 2, \ldots, n!$, is Weak Type I with respect to $\eta(\theta, \omega)$ at ω.

In B we have

$$(i)\quad \lim_{\|\omega - z\| \to 0} \eta'(\theta, z)\nabla f(z) = 0 , \leqslant f(\theta) - f(\omega) , \text{ by definition of } B .$$

So, in B, $f(\theta)$ is Weak Type I with respect to $\eta(\theta, \omega)$ at ω.

$$(ii)\quad \lim_{\|\omega - z\| \to 0} \eta'(\theta, z)\nabla g_l(z) = 0 , \quad l = 1, 2, \ldots, n! + 1 ,$$
$$\leqslant -g_l(\omega) , \quad l = 1, 2, \ldots, n! + 1 , \text{ by (3.7)–(3.8)} .$$

So, in B, $g_l(\theta)$, $l = 1, 2, \ldots, n! + 1$, is Weak Type I with respect to $\eta(\theta, \omega)$ at ω.

So it has been shown that all the functions in problem (3.6)–(3.8) are Weak Type I with respect to the same η at $\omega = (1/n!, 1/n!, \ldots, 1/n!, 1)'$; and since the Kuhn–Tucker conditions are satisfied at ω the minimum value of $f(\theta)$ is $\Phi(\omega)$ which equals $per(\Sigma_{i=1}^{n!} \omega_i P_i)$ in the equivalent problem (3.1)—(3.3). Since each element in the matrix $\Sigma_{i=1}^{n!} \omega_i P_i$ consists of the sum of $(n-1)!$ of the ω_i's each of which has the value $1/n!$ at minimum, then it follows that each element has the value $(n-1)!/n! = n^{-1}$ at minimum.

That is, $n^{-1}(1, \ldots, 1)'$ is a global minimum in the van der Waerden Problem.

References

1. van der Waerden, B.L. (1926) *Aufgabe 45, Jber. Deutsch. Math. Verein* **35**, 117.
2. Marcus, M. and M. Newman (1959), On the Minimum of the Permanent of a Doubly Stochastic Matrix, *Duke Math. J.* **26**, 61–72.
3. Eberlein, P.J. and G.S. Mudholkar (1968), Some Remarks on the van der Waerden Conjecture, *J. Combinatorial Theory* **5**, 386–396.
4. Eberlein, P.J. (1969), Remarks on the van der Waerden Conjecture II, *Linear Algebra and Appl.* **2**, 311–320.
5. Falikman, D.I. (1981), A Proof of the van der Waerden Conjecture on the Permanent of a Doubly Stochastic Matrix, *Mat. Zametki* **29**, 931–938 (in Russian).
6. Egorychev, G.P. (1981), The Solution of van der Waerden's Problem for Permanents, *Dokl. Akad. Sci. SSSR* **258**, 1041–1044 (in Russian); *Advances in Math* **42**, 299–305.
7. Hanson, M.A. (1981), On Sufficiency of the Kuhn–Tucker Conditions, *J. Math. Anal. Appl.* **80**, 545–550.
8. Hanson, M.A. and B. Mond (1987), Necessary and Sufficient Conditions in Constrained Optimization, *Math. Prog.* **37**, 51–58.
9. Birkhoff, G. (1946), Tres observaciones sobre el algebra lineal, *Rev. Universidad nacional de Tucuman, Ser. A.* **5**, 147–150.