

0895-7177(95)00009-7

A Functional Equation that Arises in Problems of Scheduling with **Priorities and Lateness/Earliness Penalties**

F. S. ROBERTS

Department of Mathematics and Center for Operations Research Rutgers University, New Brunswick, NJ 08903, U.S.A.

(Received June 1994; accepted July 1994)

Abstract—We consider a functional equation that arises in problems of scheduling with priorities and earliness/lateness penalties. We solve the equation and note how it can be used to analyze the invariance of conclusions about optimality of a schedule if the scale used to measure priority is replaced by another acceptable scale.

Keywords-Scheduling, Functional equations, Measurement theory, Penalty functions, Meaningful statements.

1. INTRODUCTION

There has been a great deal of interest in problems of scheduling when we apply penalties for late arrival, and, more recently, interest in such problems when we also apply penalties for early arrival. A complication that makes such problems especially difficult is when the penalty depends upon the priority of an item being scheduled. In this case, the scales of measurement used to measure priority enter into the picture and it is necessary to consider possible admissible transformations of scale (for example by changing units or zero points). Mahadev, Pekeč, and Roberts [1] raise the question of determining whether or not the conclusion of optimality in a scheduling problem can change if we apply an admissible transformation of scale in priority measurement. They give examples to show that this can happen and conditions under which it does not. In measurement theory, we call a statement whose truth does not change after admissible transformations of scales a meaningful statement. (For discussion of this concept, see for example Luce et al. [2] and Roberts [3-5].) In this paper, we investigate a functional equation that plays a role in giving conditions under which the conclusion of optimality of a schedule is a meaningful conclusion.

In the next section, we present and solve the functional equation. In Section 3, we formulate the scheduling problem and apply the solution to the equation.

2. THE FUNCTIONAL EQUATION

Suppose that h_1 and h_2 are two functions from the positive reals to the positive reals. We say that h_i is semilinear if for every $\alpha > 0$ and β , if t > 0 and $\alpha t + \beta > 0$, then

$$h_i(\alpha t + \beta) = K_i(\alpha, \beta)h_i(t) + L_i(\alpha, \beta), \qquad K_i(\alpha, \beta) > 0.$$
(1)

Typeset by A_MS -TEX

The author thanks the National Science Foundation for its support under Grant Number SES-9211492 to Rutgers University and also thanks J. Aczél, N. V. R. Mahadev, and A. Pekeč for their helpful comments.

We say that the pair (h_1, h_2) is pair semilinear if h_1 and h_2 are both semilinear and K_i is the same function for i = 1, 2 and L_i is the same function for i = 1, 2.

Equation (1) arises in measurement theory if we apply what Luce [6] has called the "principle of theory construction": If a law relates a dependent variable to an independent variable, then an admissible transformation of the independent variable should lead to an admissible transformation of the dependent variable. This equation arises when both the independent and dependent variables are "interval scales" (see Section 3). This "principle" was first applied by Luce to determine the general forms of the possible laws of psychophysics.¹ Avoiding the principle, Roberts and Rosenbaum [13] observe how one comes up with the same functional equation under different assumptions. The same equation arises in understanding the general form of scientific laws, not just those of psychophysics, in understanding the possible merging functions in various decision making contexts, and in understanding the possible index numbers in economics. See Roberts [5] for a summary of such applications.

The functional equation (1) was solved by Luce [6] under the assumption that h_i is continuous, that it is defined from the reals to the reals, and that it holds even if $t \leq 0$ and $\alpha t + \beta \leq 0$. Roberts [3] gives another proof of Luce's result. The assumption of continuity can readily be replaced by the assumption that h_i is increasing in t. Under either assumption, the possible h_i are then all the functions of the form $h_i(t) = a_i t + b_i$, where a_i and b_i are real constants. Aczél, Roberts, and Rosenbaum [14] obtain the same result after dropping the assumption that h_i is continuous or increasing in t. However, the proof is not readily translatable to one for functions from the positive reals to the positive reals and in addition the possibility that $\alpha t + \beta \leq 0$ is used in a crucial way in the proof. In this paper, we show how to get around both of these difficulties.

THEOREM 1. Suppose h_i is a function from the positive reals to the positive reals and h_i is semilinear. Then $h_i(t) = a_i t + b_i$, where $a_i \ge 0$, $b_i \ge 0$ and either $a_i > 0$ or $b_i > 0$, and conversely any function satisfying these conditions is semilinear. If (h_1, h_2) is pair semilinear, then in addition $a_1 = a_2$ and $b_1 = b_2$, and conversely these conditions imply that (h_1, h_2) is pair semilinear.

PROOF. The proof mimics the solution to functional equation (1) in Aczél, Roberts, and Rosenbaum [14], where $t \leq 0$ and $\alpha t + \beta \leq 0$ are allowed. We prove the first part of the theorem for $h = h_i$. Define W(t) = h(t) - h(1). Choose t, α, β so that $t > 0, \alpha > 0, \alpha t + \beta > 0$. Then by (1),

$$h(\alpha + \beta) = K(\alpha, \beta)h(1) + L(\alpha, \beta).$$

Combining this with (1), we get

$$h(\alpha t + \beta) - h(\alpha + \beta) = K(\alpha, \beta)W(t).$$

Adding h(1) - h(1) to the left hand side gives us

$$W(\alpha t + \beta) = K(\alpha, \beta)W(t) + W(\alpha + \beta),$$
(2)

for t > 0, $\alpha > 0$, $\alpha t + \beta > 0$, and in particular for t > 0, $\alpha > 0$, $\beta \ge 0$. Hence, for t > 0, $\alpha > 0$, $\alpha' > 0$, $\beta \ge 0$, $\beta' \ge 0$,

$$W(\alpha \alpha' t + \alpha \beta + \beta') = K(\alpha \alpha', \alpha \beta + \beta')W(t) + W(\alpha \alpha' + \alpha \beta + \beta')$$

¹Luce no longer refers to this as a "principle" and, following criticisms of Rozeboom [7,8], points out (e.g., Luce [9]) that it is not universally applicable. Luce [personal communication, August 20, 1985] now argues that this should not be an assumption, but it should be derived from deeper principles of theory construction. Attempts in this direction, at least for the related concept of dimensional invariance, are described in Krantz *et al.* [10], Luce [11], and Luce *et al.* [2]. See also Luce [12].

and also

$$W(\alpha \alpha' t + \alpha \beta + \beta') = K(\alpha, \beta')W(\alpha' t + \beta) + W(\alpha + \beta')$$

= $K(\alpha, \beta')K(\alpha', \beta)W(t) + K(\alpha, \beta')W(\alpha' + \beta) + W(\alpha + \beta').$

If aW(t) + b = cW(t) + d for t > 0, we conclude that either a = c or W(t) is a constant (d-b)/(a-c). Thus, either W(t) is constant (in which case h(t) is constant and we are done) or for α , $\alpha' > 0$ and β , $\beta' \ge 0$,

$$K(\alpha \alpha', \alpha \beta + \beta') = K(\alpha, \beta') K(\alpha', \beta).$$
(3)

Let $M(\alpha) = K(\alpha, 0)$ and $E(\beta) = K(1, \beta)$. Letting $\beta = \beta' = 0$ in (3), we obtain for all $\alpha, \alpha' > 0$,

$$M(\alpha \alpha') = M(\alpha)M(\alpha'). \tag{4}$$

Letting $\alpha = \alpha' = 1$ in (3), we obtain for all β , $\beta' \ge 0$,

$$E(\beta + \beta') = E(\beta')E(\beta).$$
(5)

Now M, E > 0 since K > 0. Hence, letting $\alpha = 1, \beta = 0$ in (3) gives us for all $\alpha' > 0$ and $\beta' \ge 0$,

$$K(\alpha',\beta') = M(\alpha')E(\beta').$$
(6)

Letting $\alpha' = 1$, $\beta' = 0$ in (3) gives us

$$K(\alpha, \alpha\beta) = M(\alpha)E(\beta).$$

However, by (6), $K(\alpha, \alpha\beta) = M(\alpha)E(\alpha\beta)$, so for all $\alpha > 0$, $\beta \ge 0$, since $M(\alpha) > 0$, we conclude that $E(\alpha\beta) = E(\beta)$. Hence, we have $E(2\beta) = E(\beta)$. However, by (5), $E(2\beta) = E(\beta)E(\beta)$. Hence, it follows that $E(\beta) = E(\beta)^2$. Since E > 0, we conclude that $E(\beta) \equiv 1$. Thus, by (6), for all $\alpha > 0$, $\beta \ge 0$, $K(\alpha, \beta) = M(\alpha)$. Thus, (2) becomes

$$W(\alpha t + \beta) = M(\alpha)W(t) + W(\alpha + \beta)$$
(7)

for all t > 0, $\alpha > 0$, $\beta \ge 0$. Letting $\beta = 0$ in (7) gives us for all t > 0, $\alpha > 0$,

$$W(\alpha t) = M(\alpha)W(t) + W(\alpha).$$
(8)

Letting $\alpha = 1$ in (7) gives us for all $t > 0, \beta \ge 0$,

$$W(t+\beta) = M(1)W(t) + W(1+\beta) = W(t) + W(1+\beta)$$
(9)

because, by (4), $M(1 \times 1) = M(1)M(1)$ and M(1) > 0, so M(1) = 1. Now consider $\alpha, \alpha', \alpha'' > 0$. By (9) and (8), we have

$$W((\alpha + \alpha')\alpha'') = W(\alpha\alpha'') + W(1 + \alpha'\alpha'')$$

= $M(\alpha)W(\alpha'') + W(\alpha) + W(\alpha'\alpha'') + W(2)$
= $M(\alpha)W(\alpha'') + W(\alpha) + M(\alpha')W(\alpha'') + W(\alpha') + W(2)$
= $Y.$

However, we also have

$$\begin{split} W((\alpha + \alpha')\alpha'') &= M(\alpha + \alpha')W(\alpha'') + W(\alpha + \alpha') \\ &= M(\alpha + \alpha')W(\alpha'') + W(\alpha) + W(1 + \alpha') \\ &= M(\alpha + \alpha')W(\alpha'') + W(\alpha) + W(\alpha') + W(2) = Y'. \end{split}$$

Since Y = Y', we get

$$M(lpha+lpha')W(lpha'')=[M(lpha)+M(lpha')]W(lpha'')$$

Thus, unless W is identically 0, in which case h is a constant and we are done, we conclude that

$$M(\alpha + \alpha') = M(\alpha) + M(\alpha').$$
⁽¹⁰⁾

Now this plus (4) both hold for α , $\alpha' > 0$. It is a well-known result [15] that if (10) and (4) hold for all α , $\alpha' > 0$, then either $M(\alpha) = \alpha$ for all $\alpha > 0$ or $M(\alpha) \equiv 0$. However, $M(\alpha)$ is known to be > 0 since K > 0, and we conclude that $M(\alpha) = \alpha$ for all $\alpha > 0$. It follows from (8) that

$$W(\alpha t) = \alpha W(t) + W(\alpha)$$

for all α , t > 0. Now note that $W(\alpha t) = W(t\alpha)$, so

$$\alpha W(t) + W(\alpha) = tW(\alpha) + W(t),$$

or

$$W(t)(\alpha - 1) = W(\alpha)(t - 1)$$

for all α , t > 0. In particular, letting $\alpha = 2$, we get

$$W(t) = W(2)(t-1) = at + c$$

for all t > 0. It follows that

$$h(t) = at + b$$

for all t > 0. Moreover, a must be nonnegative for otherwise h(t) = at + b < 0 for t sufficiently large. Also, b must be nonnegative for otherwise h(t) = at + b < 0 for t sufficiently small. Now a and b cannot both be zero, since otherwise h(t) = at + b = 0 and we have assumed that h > 0. This completes the proof in the case where h_i is semilinear, except for the converse part, which is straightforward.

Now suppose that (h_1, h_2) is pair semilinear. From the semilinearity result,

$$h_1(\alpha t + \beta) = a_1\alpha t + a_1\beta + b_1,$$
 $h_2(\alpha t + \beta) = a_2\alpha t + a_2\beta + b_2.$

Since this must hold for all t > 0, it follows by equation (1) that

$$K_1(\alpha,\beta) = a_1\alpha, \quad K_2(\alpha,\beta) = a_2\alpha.$$

Since $K_1 = K_2$ and $\alpha > 0$, we conclude that $a_1 = a_2$. But then

$$L_1(\alpha,\beta) = a_1\beta + b_1,$$
 $L_2(\alpha,\beta) = a_2\beta + b_2 = a_1\beta + b_2.$

Since $L_1 = L_2$, we conclude that $b_1 = b_2$. The proof of the converse part is straightforward.

3. APPLICATION TO SCHEDULING

To see how functional equation (1) arises in scheduling, let us define a specific scheduling problem. We have n items that need to be transported from an origin to a destination. Let t_i , i = 1, ..., n, denote some measure of the priority of i and d_i denote the desired arrival time of i. We assume that each d_i is a positive integer. At any given arrival time, there is a certain available positive integer capacity c for transportation. A schedule u is then an assignment of an integer arrival time u_i to each item i, subject to the constraint that at any given arrival time, at most c items are scheduled to arrive. A penalty is based on the vectors $\mathbf{t} = (t_1, t_2, \ldots, t_n)$,

80

 $\mathbf{d} = (d_1, d_2, \dots, d_n), \mathbf{u} = (u_1, u_2, \dots, u_n)$ and the integer c. (We use the convention that bold-face letters indicate vectors.) This penalty is denoted $P(\mathbf{u}; \mathbf{t}, \mathbf{d}, c)$ and we seek to minimize it. We shall assume that the penalty is summable in the sense that it can be expressed as a sum $\sum_{i=1}^{n} g(t_i, u_i, d_i)$ and separable in the sense that g can be expressed as

$$g(t, a, d) = \begin{cases} h_1(t)f(a, d), & \text{if } a \ge d, \\ h_2(t)f(a, d), & \text{if } a < d. \end{cases}$$

for functions h_1 and h_2 defined on the positive reals into the positive reals and f(a, d) defined on $N \times N$ into the reals. It is often useful to assume that the penalty function P is symmetric in the sense that $h_1 = h_2$, and *t*-increasing in the sense that each h_i is increasing in t. Further information about any of these concepts can be found in Mahadev, Pekeč, and Roberts [1,16] or in many of the references to the scheduling literature found there. Sample references that in turn include many references are Baker and Scudder [17], Hall and Posner [18], and Garey, Tarjan, and Wilfong [19].

If $\varphi(t)$ is a function, we shall let $\varphi(t)$ denote the vector $(\varphi(t_1), \varphi(t_2), \dots, \varphi(t_n))$. The question of whether or not the conclusion that schedule **u** is optimal is a meaningful conclusion is equivalent to the question: If $P(\mathbf{u}; \mathbf{t}, \mathbf{d}, c)$ is minimum for this scheduling problem defined by $\mathbf{t}, \mathbf{d}, c$, is $P(\mathbf{u}; \varphi(\mathbf{t}), \mathbf{d}, c)$ also minimum for the scheduling problem defined by $\varphi(\mathbf{t}), \mathbf{d}, c$, when φ is an admissible transformation of scale?

In measurement theory, we say that a scale is a *ratio scale* if the admissible transformations of scale correspond exactly to multiplication by a positive constant, i.e., they are transformations of the form $\varphi(t) = \alpha t, \alpha > 0$. These transformations correspond to change of unit. We say that a scale is an interval scale if the admissible transformations correspond exactly to multiplication by a positive constant and addition of another constant, i.e., they are transformations of the form $\varphi(t) = \alpha t + \beta, \alpha > 0$. These transformations correspond to change of unit and change of zero point. We say that a scale is an ordinal scale if the admissible transformations correspond exactly to the (strictly) monotone increasing transformations. For further information about scales of measurement, see the books by Krantz et al. [10], Luce et al. [2], Suppes et al. [20], or Roberts [3], and see the paper by Roberts [5]. We will be interested here in the question of whether or not the conclusion that u is optimal for a scheduling problem is meaningful if priorities are measured on an interval scale. In the situation where scales must be positive, not every change of zero point is allowed. Given t > 0, we only allow transformations $\alpha t + \beta$ that result in positive numbers. Thus, in the case of scheduling with positive priorities measured on an interval scale, we consider transformations of the form $\varphi(t) = \alpha t + \beta$ where $\alpha > 0, t > 0$, and $\alpha t + \beta > 0$. This is exactly the situation under which we consider functional equation (1). If the penalty function is summable and separable, then $P(\mathbf{u}; \varphi(\mathbf{t}), \mathbf{d}, c)$ involves terms of the form $h_i(\alpha t + \beta)$. Under Luce's principle of theory construction (see Section 2), if $h_i(t)$ also defines an interval scale, then an admissible transformation $\alpha t + \beta$ of t should lead to an admissible transformation $K_i(\alpha, \beta)h_i(t) + L_i(\alpha, \beta)$ of $h_i(t)$, where $K_i(\alpha,\beta)$ is the change of unit (depending upon α and β) and $L_i(\alpha,\beta)$ is the change of zero point. This gives us equation (1). We are not going to argue for the principle of theory construction here or argue that equation (1) has to hold. Rather, we are going to argue that if it does hold and some other conditions hold, then certain conclusions about optimality of a schedule under interval scale priority measurement are meaningful. Surprisingly, it turns out that these conclusions are meaningful even under ordinal scale priority measurement.

Mahadev, Pekeč, and Roberts [16] study a very specific scheduling problem, one where there are only two desired arrival times and all items but one have the same desired arrival times. Here, $\mathbf{d} = (d, d, \dots, d, k)$. They prove the following theorem.

THEOREM 2². Consider a scheduling problem $\mathbf{t}, \mathbf{d}, c$ with $\mathbf{d} = (d, d, \dots, d, k), c = 1$, and either

²Technically, Mahadev, Pekeč, and Roberts make the assumption that $t_1 \ge t_2 \ge \cdots \ge t_{n-1}$. However, this assumption is not needed for this theorem to hold, only to define what they call a (d, k)-problem.

- (i) k > d or
- (ii) n/2 < d and k < d.

Suppose that the penalty function is summable, separable, symmetric, and t-increasing and f(a,d) = D|a-d| + E for D > 0. Then the statement that **u** is optimal for this problem is meaningful if priorities are measured on an ordinal scale.

Let us say that a summable, separable penalty function is *pair semilinear* if (h_1, h_2) is pair semilinear. By the observations in Theorem 1, if the penalty function is semilinear, then $a_1 = a_2$, $b_1 = b_2$ and $h_1(t) = h_2(t) = At + B$. In case A > 0, we have symmetry and *t*-increasingness. Thus, these hypotheses of Theorem 2 do not need to be assumed. In case A = 0, h_1 and h_2 are constants and so for all \mathbf{u} , $P(\mathbf{u}; \mathbf{t}, \mathbf{d}, c) = P(\mathbf{u}; \varphi(\mathbf{t}), \mathbf{d}, c)$. In this case, optimality of \mathbf{u} is clearly preserved under any transformation φ . Thus, we have the following corollary of Theorem 2.

THEOREM 3. Consider a scheduling problem t, d, c with $d = (d, d, \dots, d, k), c = 1$, and either

- (i) k > d or
- (ii) n/2 < d and k < d.

Suppose that the penalty function is summable, separable, and pair semilinear, and f(a,d) = D|a-d| + E for D > 0. Then the statement that u is optimal for this problem is meaningful if priorities are measured on an ordinal scale.

It is interesting to note that Theorem 3 fails if we change the definition of a pair semilinear penalty function to be that each of h_1 and h_2 are individually semilinear. Suppose that $h_1(t) = 3 + t$, $h_2(t) = 3t$, f(a, d) = |a - d|, $\mathbf{d} = (2, 2, 3)$, $\mathbf{t} = (1, 1, 1)$, c = 1. Then one can show that the schedule $\mathbf{u} = (2, 1, 3)$ is optimal. However, if $\varphi(t) = 7t$ and \mathbf{t} is replaced by $\varphi(\mathbf{t}) = (7, 7, 7)$, then \mathbf{u} is no longer optimal because $\mathbf{v} = (2, 4, 3)$ has a smaller penalty.

For the special case d = 1, Mahadev, Pekeč, and Roberts [16] obtain the following result.

THEOREM 4³. Consider a scheduling problem t, d, c with d = (1, 1, ..., 1, k), $k \neq 1$, and c = 1. Suppose that the penalty function is summable and separable, $h_1(t)$ is increasing in t, and f(a,d) = D|a-d| for D > 0. Then the statement that u is optimal for this problem is meaningful if priorities are measured on an ordinal scale.

Let us say that a summable, separable penalty function is 1-semilinear if h_1 is semilinear. The next result follows from Theorem 4 in the same way that Theorem 3 follows from Theorem 2. We use the observation in Theorem 1 that semilinearity of h_1 implies $h_1(t) = At + B$, $A \ge 0$.

THEOREM 5. Consider a scheduling problem t, d, c with d = (1, 1, ..., 1, k), $k \neq 1$, and c = 1. Suppose that the penalty function is summable, separable and 1-semilinear, and f(a, d) = D|a-d| for D > 0. Then the statement that u is optimal for this problem is meaningful if priorities are measured on an ordinal scale.

REFERENCES

- N.V.R. Mahadev, A. Pekeč and F.S. Roberts, On the Scheduling Problem with Priorities and Lateness/Earliness Penalties: When is an Optimal Solution not Optimal?, Research Report 19-94, Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ, (1994).
- R.D. Luce, D.H. Krantz, P. Suppes and A. Tversky, Foundations of Measurement, Vol. III, Academic Press, New York, (1990).
- 3. F.S. Roberts, Measurement Theory, with Applications to Decisionmaking, Utility, and the Social Sciences, Addison-Wesley, Reading, MA, (1979).
- F.S. Roberts, Applications of the theory of meaningfulness to psychology, Journal of Mathematical Psychology 29, 311–332, (1985).
- F.S. Roberts, Limitations on conclusions using scales of measurement, In Operations Research and the Public Sector, (Edited by A. Barnett, S.M. Pollock and M.H. Rothkopf), pp. 621-671, Elsevier, Amsterdam, (1994).
- 6. R.D. Luce, On the possible psychophysical laws, Psych. Rev. 69, 81-95, (1959).

82

³Technically, Mahadev, Pekeč, and Roberts make the assumption that $t_1 \ge t_2 \ge \cdots \ge t_{n-1}$. However, as with Theorem 2, this assumption is not needed for this theorem to hold.

A Functional Equation

- 7. W.W. Rozeboom, The untenability of Luce's principle, Psych. Rev. 69, 542-547, (1962).
- 8. W.W. Rozeboom, Comment, Psych. Rev. 69, 552, (1962).
- 9. R.D. Luce, Comments on Rozeboom's criticisms of 'On the possible psychophysical laws', *Psych. Rev.* 69, 548-555, (1962).
- 10. D.H. Krantz, R.D. Luce, P. Suppes and A. Tversky, Foundations of Measurement, Vol. I, Academic Press, New York, (1971).
- 11. R.D. Luce, Dimensionally invariant numerical laws correspond to meaningful qualitative relations, *Phil. of Sci.* 45, 1-16, (1978).
- R.D. Luce, 'On the possible psychophysical laws' revisited: Remarks on cross-modal matching, *Psych. Rev.* 97, 66–77, (1990).
- F.S. Roberts, Z. Rosenbaum, Scale type, meaningfulness, and the possible psychophysical laws, Math. Soc. Sci. 12, 77-95, (1986).
- J. Aczél, F.S. Roberts, and Z. Rosenbaum, On scientific laws without dimensional constants, J. Math. Anal. & Applic. 119, 389-416, (1986).
- 15. J. Aczél, Lectures on Functional Equations and their Applications, Academic Press, New York, (1966).
- N.V.R. Mahadev, A. Pekeč and F.S. Roberts, On the Scheduling Problem with Priorities and Lateness/Earliness Penalties: Some Meaningful Conclusions with Nonconstant Arrival Times, Research Report 20-94, Rutgers Center for Operations Research, Rutgers University, New Brunswick, NJ, (1994).
- K.R. Baker and G.W. Scudder, Sequencing with earliness and tardiness penalties: A review, Operations Research 38, 22-36, (1990).
- 18. N.G. Hall and M.E. Posner, Earliness-tardiness scheduling problems I: Weighted deviation of completion times about a common due date, *Operations Research* **39**, 836-846, (1991).
- M.R. Garey, R.E. Tarjan and G.T. Wilfong, One-processor scheduling with symmetric earliness and tardiness penalties, *Mathematics of Operations Research* 13, 330–348, (1988).
- P. Suppes, D.H. Krantz, R.D. Luce and A. Tversky, Foundations of Measurement, Vol. II, Academic Press, New York, (1989).