RR-75-26

NON-LINEAR PROGRAMMING APPROACHES TO NATIONAL SETTLEMENT SYSTEM PLANNING

Yuri Evtushenko
Ross D. MacKinnon

July 1975

Abstract

Three rather aggregate approaches to modelling interregional migration processes within a national urban settlement systems context are described. General, modified penalty function methods of non-linear programming are developed and then adapted for application to the simplest of the three migration models. The numerical convergence properties of the procedure are discussed. Some of the numerical results for a Canadian urban system case study are interpreted. Finally, some extensions to the procedures used in this study as well as alternative approaches to the same or similar problems are suggested.

Non-Linear Programming Approaches to National Settlement System Planning

Yuri Evtushenko*
Ross D. MacKinnon**

This paper has three primary objectives. The first is to outline some models which attempt to identify, in an aggregate way, strategies for achieving certain desirable population trajectories by manipulating the migration parameters of a simple linear model. The second is to present a family of numerical optimization methods suitable for solving these types of problems. The third is to describe some preliminary numerical results using these methods in a national settlement system context.

The particular numerical results, while of some interest, should not be taken too seriously as the goal and cost functions are to a large extent fictitious. The results are illustrative of the type of indications one might obtain from a more thorough empirical study rather than actual' prescriptions for an urban policy making agency. One more general purpose of such a study is to determine whether such results would be meaningful to policy makers and, if not, how the methodology could be adjusted to provide more useful insights.

It is explicitly recognized at the outset that these models are rather unrealistic in at least three respects: (1) the goal specification and weighting problem (social preference function) is assumed to be solved--moreover the system goals are defined in terms of desirable population trajectories for each of the regions, or perhaps some subset of the regions; (2) the cost function of influencing migration patterns is given; (3) the precise instruments whereby migration patterns can be changed are not considered; in a formal sense, migration rates themselves are instrument variables, whereas in most, if not all, societies the controllable variables are less directly related to population distributions.

In summary, then, we assume a rather simple closure between target, state and instrument variables, not because

---

*IIASA and Computing Centre, Moscow, USSR.

**IIASA and University of Toronto, Canada.

we have confidence that such a structure is realistic, but rather to determine whether a characterization of the national urban settlement planning in this way is likely to provide insights into the more complex process which it obviously is.

The paper is divided into four sections. First, the mathematical models and the empirical data are described. Second, new penalty function methods of non-linear programming are summarized and the characteristics of their computational performance in the national settlement system's application are discussed. Third, the numerical results--the population and control trajectories and their sensitivity to parameter changes--are presented and evaluated. Finally, the methodology and the results are summarized and evaluated and some promising areas of future research are indicated.

1. Model Specification

In situations where the goals, costs, and causal structure are poorly understood, it is perhaps naive to attempt to specify a dynamic normative model. One could argue with some justification that much analysis is needed before the plan synthesis step is taken. Although such a position is arguable, it is not the one adopted here because a simple linear sequential approach to problem formulation and resolution is deemed to be inappropriate. We argue that with poorly understood systems, in particular, a dialectic between analysis and synthesis is potentially more rewarding. Thus, an attempt to develop formal planning strategies with albeit preliminary statements of system behaviour may aid in the development of both descriptive and normative aspects of modelling. In summary then, it is recognized that the problem statement is overly simple. In an immediate practical planning context, it is postulated that even such simple frameworks may give a government agency some insights into the orders of magnitude and the spatial and temporal distribution of effort necessary to move the system towards specified goals: the trade-offs between different parameters in the system (goals, costs and migration rates), the sensitivity of system performance to changes in any or all of these parameters, and the range of alternative population distributions which are plausible. Perhaps even more important than these considerations is the hope that these initial experiments with optimization methods in national settlement system management will stimulate more comprehensive and ultimately more realistic attempts.

A number of models are formulated, although numerical results are presented only for the simplest form. All of the models are essentially more explicit statements of those described in MacKinnon (1975a). Three classes of models are formulated:

(1) Forward Linkage Models. Controls are in the form of in-migrants to the system who, together with the previously existing population, subsequently migrate between the regions of the system. In their purest form, these models do not attempt to change the nature of the interregional migration propensities.

(2) Backward Linkage Models. Controls are in the form of stimuli for people to move to specific locations within the system. New job and housing vacancies are perhaps the most obvious examples of such stimuli. Again the proportional distribution of origins for a given destination is assumed to be constant.

(3) Variable Structure Models. The elements of the interregional migration matrix are themselves control variables.

It is, of course, quite probable that effective policies would represent combinations of these three classes of systems. Examples of each of the three types are now presented, followed by a discussion of the problem of defining a suitable objective function.

1.1 Forward Linkage Models

For all of the models presented, the state of the system at time i is the population distribution vector x_i. Changes in population distribution from one time period to the next, for the first category of models, are related to the distribution and magnitude of births, deaths, intra-system migration flows, and migration from outside the system. More formally

$$x_{i+1} = (N + M)\, x_i + u_i \quad , \qquad i = 0,1,\ldots,I-1 \quad , \qquad [1]$$

where N is a diagonal matrix of rates of natural increase;
M is a $K \times K$ matrix of interregional migration propensities (i.e., M^{jk} is the probability that a person currently residing in region k will migrate to region j during one time interval);

u_i is the number of people migrating to each region from outside the system. (These in-migrants may be people coming from foreign areas or from non-urban locations in the nation.)

In the simplest formulation, u_i is assumed to be controllable. That is, it is possible for the federal government to direct in-migrants to any of the locations either by regulation or by providing subsidies or imposing tax penalties. While it is clear that few national governments have the will or the ability to control the system so directly, it is of some interest to determine how effective such direct controls would be were they feasible*.

Imposing controls usually implies the incurrence of costs. Avoidance of costs is characteristic of the management of many systems. Costs may be included either in the objective function or as a part of the constraint set. In this formulation, the latter alternative is adopted:

$$\sum_{i=0}^{I-1} \sum_{k=1}^{K} r_i^k u_i^k \leq B \quad , \qquad [2]$$

where r_i^k is the cost, perhaps discounted, of directing a person to region k in the i^{th} time period; and B is the total allowable budget over the entire planning period.

In part, r_i^k represents directly measurable costs, but it should probably be interpreted more generally as including not only monetary costs but the bureaucratic, even psychological, effort necessary to induce an in-migrant to locate in a specific region. Of interest is the shape of the trade-off between effectiveness as measured by the objective function and different values of B.

---

*It would be possible to generalize the results somewhat by assuming that the effectiveness of the intended control could be described according to a probability distribution. While certainly a more realistic representation, this approach has not been adopted at this stage because it adds a level of complexity that could make the results less readily interpretable. Moreover, it would require the estimation of another set of parameters in a model which is already overextended in this regard.

Another constraint on control is the total in-migrant pool available in each year. Although it is conceivable that this pool itself is partially controllable, it is assumed here that the stream of in-migrants is, at best, only predictable, perhaps by another model; at worst, it is an unpredictable, exogenous variable so that extensive analysis would have to be undertaken to determine the range of responses and outcomes which would be implied under different conditions. These constraints, one for each time period, are expressed in the following way:

$$\sum_{k=1}^{K} u_i^k = \hat{u}_i \qquad i = 0,1,\ldots,I-1 \quad , \qquad [3]$$

where $\hat{u}_i$ is the total in-migration pool available in the i^{th} time period. The final constraint set in this formulation consists of the conventional non-negativity conditions:

$$u_i^k \geq 0 \qquad k = 1,\ldots,K \quad , \qquad [4]$$

$$i = 0,\ldots,I-1 \quad .$$

In other words, control is exercised with respect only to in-migrants, and not out-migrants. Out-migration may be incorporated by allowing columns of the matrix M not summing to one or equivalently by including out-migration rates in death rates.

1.2 Backward Linkage Systems

The growth and distribution of the population of many urban systems can be controlled only very marginally by directing in-migrants, either because in-migrants represent a very small proportion of total population or because there are severe economic, social and political restrictions in controlling their destinations. In such cases, the federal government may wish to introduce stimuli in specific temporal and spatial sequences in order to steer the system as closely as possible towards population distribution goals. Consider the case where job vacancies are the stimulus to which migrants tend to respond. One appropriate model may replace [1] with:

$$x_{i+1} = [I + N]\ x_i + [I - \tilde{M}]\ V_i \qquad [5]$$

$$V_i = V_{u_i} + \tilde{M}V_{i-1} + V_{z_i} \quad , \qquad [6]$$

where V_i is the distribution of job vacancies in the i^{th} time period;
V_{u_i} is the distribution of government stimulated job vacancies;
V_{z_i} is the distribution of spontaneously occurring vacancies (arising from retirement, economic growth, etc.);
$\tilde{M}$ is a migration matrix with elements $\tilde{M}^{jk}$, the probability that a job vacancy in region k will be filled by someone living in j.

The budget constraints corresponding to [2] and [3] would, of course, have V_{u_i} terms instead of u_i, as would the non-negativity conditions [4]. This model is discussed in more detail in MacKinnon (1975a).

It may, in fact, be more appropriate to use this model to control the distribution of a particularly important subgroup. That is, what stimuli must be imposed on the system in order for the distribution of teachers or doctors to come as close as possible to some "equitable" distribution, taking into account the likely origins of those teachers and doctors*? Because of time and data restrictions, this model has not been implemented. However, no computational difficulties are anticipated.

1.3 Variable Structure Models

Perhaps the most interesting class of models consists of those in which the rules of change can themselves be controlled within certain limitations. That is, the propensities with which people tend to migrate between regions can be changed. Thus the system dynamics could be represented by:

$$x_{i+1} = (N + M_i)\ x_i + z_i \quad , \qquad [7]$$

---

*This suggestion was made by Nathan Keyfitz, Harvard University.

where M_i is the time varying, controllable migration matrix;

z_i is an exogenously given net in-migration vector.

Ideally, these changes should be related to some specific control variables attached to which are certain cost functions. Thus changes should be made which are feasible within the context of budgetary limitations. However, in the absence of such information, it is assumed that the difficulties of changes in migration rates are directly proportional to the relative magnitude of such changes. For each element M_i^{jk}, there could be the following constraint:

$$(1 - \lambda) M_{i-1}^{jk} \leq M_i^{jk} \leq (1 + \lambda) M_{i-1}^{jk} \quad . \qquad [8]$$

That is, M_i^{jk} must lie within specified deviations of previous values of the parameter. Only gradual changes in system structure are possible.

A somewhat more complicated constraint with a similar intent is:

$$\sum_{j=1}^{K} \sum_{k=1}^{K} (M_i^{jk} - \delta_i x_i^j x_i^k)^2 \leq \sigma \quad . \qquad [9]$$

This would be suitable in the case where the migration rates M_i^{jk} were known to be statistically related to the population distribution of the system in addition to some other uncontrollable variables (incorporated in δ_i). A gravity or spatial interactance model would be of this type. The constraint simply insures that the controls are plausible within the context of this known statistical relationship. Although for short periods and small values of λ, constraint [8] would effectively impose this condition; but for longer planning periods [8] may not be sufficient.

Other constraints must be placed on M_i^{jk} if they are to be interpreted as transition probabilities:

$$\sum_{j=1}^{K} M_i^{jk} = 1.0 \qquad \text{for all } k = 1,\ldots,K$$
$$i = 0,\ldots,I-1 \qquad [10]$$
$$0 \leq M_i^{jk} \leq 1.0 \quad .$$

Although this problem appears to be feasible, if somewhat cumbersome with respect to data requirements and computational demands, no numerical results have been computed at this date.

1.4 Specification of a Criterion Function

To complete all of the models formulated above, the inclusion of an objective function is necessary. We have chosen to express this in terms of deviations from prescribed population trajectories. A quadratic loss function has some appeal (see Nykamp and Somermeyer, 1974), although its symmetry, weighting positive and negative deviations equally, is unlikely to be fully satisfactory. In our numerical analyses, two alternative forms are used, the symmetric, quadratic and the asymmetric, exponential:

$$\text{MIN} \sum_{i=1}^{I} \sum_{k=1}^{K} a^k (x_i^k - \hat{x}_i^k)^2 \qquad [11]$$

$$\text{MIN} \sum_{i=1}^{I} \sum_{k=1}^{K} \left(a^k \exp \left(\beta^k \left(\frac{x_i^k - \hat{x}_i^k}{\hat{x}_i^k} \right) \right) - 1 \right) \quad , \qquad [12]$$

where $\hat{x}_i^k$ is the population target for region k at time i, the set $\hat{x}_1^k, \hat{x}_2^k, \ldots, \hat{x}_I^k$ defining the desired population trajectory for region k;

a^k is the importance associated with attaining the population trajectory of region k;

β^k is defined in such a way as to take goal asymmetry into account. For example, if region k is growing too rapidly, exceeding the population target $\hat{x}_i^k$ should be penalized more severely than falling short of the target; thus, if

$x_i^k > \hat{x}_i^k \quad , \qquad \beta^k = \bar{\beta}^k$

and if $x_i^k \leq \hat{x}_i^k \quad , \qquad \beta^k = \bar{\bar{\beta}}^k$

where $\bar{\bar{\beta}}^k < 0$ and $|\bar{\beta}^k| > |\bar{\bar{\beta}}^k|$. An analogous definition of β^k is made when k is a region which is deemed to be lagging in terms of growth. Some regions may have symmetric loss functions, i.e. $|\bar{\beta}^k| = |\bar{\bar{\beta}}^k|$.

Both formulations can be criticized on practical as well as theoretical grounds. The problems of objective estimation of these goals and parameters are formidable. The additive nature of both functions is highly questionable. Might not the failure to meet a goal in one region have implications for the importance attached to meeting goals in other regions? More formally, if [11] is expressed in matrix terms,

$$\sum_{i=1}^{I} (x_i^k - \hat{x}_i^k)' \, A(x_i^k - \hat{x}_i^k) \quad , \qquad [13]$$

we have been assuming that A is a diagonal matrix, whereas goal interdependencies would imply the existence of cross-product terms. We ignore this and other problems, not because we believe them to be unimportant, but rather to gain some experience with the properties of such systems in their simplest forms and to make some judgment as to the most promising areas of extension.

1.5 The Data

The "objective" data for the following experiments are taken from the 1966 and 1971 Censuses of Canada. The 1971 populations of the 22 Census Metropolitan Areas (C.M.A.'s) is the vector X(0), the "initial" state of the system. Estimates for the inter-C.M.A. migration rates are obtained from the 1966 population distribution vector and a matrix Q. The elements Q^{jk} represent the number of people in city j in 1971 who had moved from city k some time in the five-year time interval 1966-1971. Our estimates of the off-diagonal transition probabilities M^{jk} are $\dfrac{Q^{jk}}{X^k(1966)}$, $k \neq j$ and

$M^{jj} = 1 - \sum_{j \neq k} M^{jk}$. Although clearly there are problems with these estimates (as there are with most parameters which are defined in terms of both flow and stock variables), they are used here as reasonable first approximations for migration propensities.

The problems associated with some of the other parameters are even more severe. Population goals for each city ($\hat{x}_i^k$), their relative importance (a^k), and the costs influencing the direction of the stream of in-migrants (r_i^k) were generated quite subjectively by one of the authors who has some familiarity with the Canadian urban scene. They are intended to illustrate the method, rather than to indicate real goal and cost parameters. In an actual planning context, some of these could be generated more objectively whereas others would continue to be quite subjective. One of the purposes of such models is to demonstrate the implications of quite hypothetical goals and costs.

2. A Description of Numerical Methods

There are a number of methods which could be used to solve optimization problems such as those described in the previous section. In the research reported here, methods similar to those presented in Evtushenko (1975a and b) are used. However, since more general versions have since been derived, they are presented here without proof. Moreover, we show how these methods have been adapted to solve multistage programming problems.

This section can be summarized as follows. First, we present a general formulation of a non-linear programming problem. Next, we describe three variants of simplified penalty function methods, including the formulation of an important convergence theorem. Then, one of these is adapted to a general multistage programming problem and applied to one of the problems presented in the previous section. Finally, the nature of the numerical results from a computational point of view is discussed.

2.1 Modified Penalty Function Methods for Solving Non-Linear Programming Problems

We consider the following preliminary non-linear programming problem:

minimize $F(x)$

subject to constraint [14]

$$x \in X = \{x \in E_n | g(x) = 0, \quad h(x) \leq 0\}$$

where F, g, h are given functions defined on E_n, Euclidean n-space; $x = (x^1, x^2, \ldots, x^n)$ is a point in E_n; functions F, g, h define the mappings $F : E_n \to E_1$, $g : E_n \to E_e$, $h : E_n \to E_c$.

This is called a convex programming problem if $F(x)$, $h(x)$ are convex functions and $g(x)$ is affine.

The auxiliary exterior penalty function for [14] is defined as

$$P(x,t) = \mu(t)\ F(x) + \tau(t)\ S(x) \quad . \qquad [15]$$

Here $S(x) = 0$ if $x \in X$ and $S(x) > 0$ if $x \notin X$; $\mu(t)$, $\tau(t)$ are strictly positive, continuous scalar functions of scalar variable t which are defined for all $t_0 \leq t < \infty$ and satisfy the following conditions:

$$\int_{t_0}^{\infty} \mu(t)\ dt = \infty \quad , \qquad \lim_{t\to\infty} \tau(t) > 0 \quad . \qquad [16]$$

Introduce the following four sets:

$$X_* = \{x \in E_n : \min_{z\in X} F(z) = F(x) \quad , \qquad x \in X\}$$

$$X_0 = \{x \in E_n : g(x) = 0 \quad , \qquad h(x) < 0\}$$

$$G = \{x \in E_n : P(x,0) \leq \mu(0)\ F(x_*)\}$$

$$Z = \{x \in E_n : \min_{z \in E_n} P(s,T) = P(x,T)\} \quad .$$

Here T is some given positive number. For a numerical solution of [14], it is proposed to find the limit points as $t \to \infty$ of the solution of the Cauchy problem for the system

$$\dot{x} = -P_x = -\mu(t)\ F_x(x) - \tau(t)\ S_x(x) \quad , \quad x(0) = x_0 \quad , \quad (\dot{\cdot}) = d/dt \qquad [17]$$

where P_x, F_x, S_x are gradients of the respective functions.

Theorem 1. Let F(x) and S(x) be convex, continuously differentiable functions on E_n; the set G is compact, continuous functions $\mu(t)$, $\tau(t)$ satisfy [16], $\tau(t)/\mu(t) \to \infty$ as $t \to \infty$. Then as $t \to \infty$, the solutions [3] $x(x_0,t)$ converge to the solution set X_* for any $x_0 \in E_n$; at least one limit point of the sequence $\tau(t)\ S(x(x_0,t))/\mu(t)$ exists and is equal to zero.

If X_0 is non-empty, then for any $x_* \in X_*$, vectors $p_* \in E_e$ and $w_* \in E_c$ exist such that

$$w_* \geq 0, \quad w_*^j h^j(x_*) = 0 \quad , \quad 1 \leq j \leq c \quad ,$$

$$F_x(x_*) + \sum_{j=1}^{e} g_x^j(x_*)\ p_*^j + \sum_{j=1}^{c} h_x^j(x_*)\ w_*^j = 0 \quad .$$

We say that S(x) is separable in g and h if it has a form

$$S(x) = \sum_{j=1}^{e} \psi(|g^j(x)|) + \sum_{j=1}^{c} \psi(h_+^j(x)) \quad .$$

Here $h_+^j(x) = \mathrm{MAX}\ [0, h^j(x)]$ and $\psi(y)$ is a scalar-valued function of the single variable y, defined for all positive y. Suppose that this function is twice differentiable and satisfies the following conditions:

$$\psi(0) = 0, \quad \psi'(0) = d\psi(0)/dy = 0 \quad , \qquad [18]$$

$$d^2\psi(y)/dy^2 \geq \sigma > 0 \qquad \text{for all } y \geq 0 \quad .$$

It is easy to verify that if F(x), h(x) are convex differentiable functions, and g(x) is affine, then P(x,t) is also convex and differentiable in x.

The family of simplified penalty function methods described and used in this study should be contrasted with routine penalty function methods. (See for example, Fiacco and McCormick (1968).) The conventional methods prescribe some sequence of $t = t_1, t_2, \ldots, t_s$ such that $t_i \to \infty$ for all i. It is necessary to solve an unconstrained minimization problem using, for example, a gradient method of a similar form as [17] for fixed t_i. That is, the limit point must be found for each fixed t_i. Then t_i is changed and the new limit point is found. The limit of all such limit points solves the problem described in [14]. Theorem 1 permits us to avoid much calculation and find the limit point of [17] only one time, changing μ and τ as functions of t according to condition [16].

<u>Lemma 1</u>. If [14] is a convex programming problem, the set X_* is compact, and X_* and X_0 are non-empty sets, then for any $x_* \in X_*$, $x \in E_n$, $0 < \tau \leq T$ the following inequalities hold:

$$F(x_*) - \gamma/\tau \leq P(x,t) \leq P(x,T) \quad ;$$

where $$\gamma = \left[\sum_{j=1}^{e} (p_*^j)^2 + \sum_{j=1}^{c} (w_*^j)^2\right]/2\sigma \quad .$$

This lemma was proved in Skarin (1973) and Eremin (1967).

Consider the following maximin problem associated with problem [14]:

$$J = \max_{\tau \leq T} \min_{x \in E_n} P(x,t) \quad , \qquad [19]$$

where $T > 0$ is some fixed number.

A pair $(T,\tilde{x})$, where $\tilde{x} \in Z$, solves the maximin problem [19]. If $\tilde{x} \in Z$, $x_* \in X_*$ then $F(x_*) - \gamma/T \leq P(\tilde{x},T) \leq F(x_*)$. If function $F(x)$ is bounded from below ($F(x) \leq \delta$ for all $x \in E_n$) then

$$S(\tilde{x}) \leq \left[F(x_*) - \delta\right]/T \quad .$$

By making T sufficiently large we can thereby find an appropriate solution to [14] with any required accuracy. For solving the maximin problem [19], it is sufficient to solve the following problem: minimize $P(x,T)$ over all $x \in E_n$. Regrettably this unconstrained problem is extremely difficult to solve, since for large T the function $P(x,T)$ is ill-conditioned. It is more convenient (see Evtushenko, 1975a and b) to let the parameter τ vary continuously from zero to T and solve differential equations of the form

$$\dot{x} = - P_x(x,t) \ , \quad \dot{\tau} = y(\tau) \ S(x) \ \theta(T - \tau) \ , \quad x(0) = x_0 \ ,$$
$$\tau(0) = \tau_0 \ . \qquad [20]$$

Here $\theta(y) = 1$ if $y > 0$, and otherwise, $\theta(y) = 0$; $\mu(t) = 1$; $y(\tau)$ is a continuous positive function defined for all $0 \leq \tau_0 \leq \tau < T$ and satisfies inequality

$$0 < c < \int_{\tau_0}^{T} \frac{T - \tau}{y(\tau)} \, d\tau < \infty \quad . \qquad [21]$$

We can take, for example, $y(\tau) = T - \tau$ or $y(\tau) = 1$. In the computations described later, we set $y(\tau) = \tau$.

The simplest discrete version of this method is

$$x_{s+1} = x_s - \alpha_s P_x(x_s, \tau_s) \quad , \quad \tau_{s+1} = \tau_s + \alpha_s \, y(\tau_s) \, S(x_s) \, \theta(T - \tau_s)$$

$$s = 0,1,2,\ldots. \qquad [22]$$

Step length, α_s, is a monotonically decreasing sequence which satisfies the following conditions:

$$0 < \alpha_s \quad , \quad \alpha_s \to 0 \quad , \quad \lim_{k\to\infty} \sum_{s=0}^{k} \alpha_s \to \infty \quad . \qquad [23]$$

In the following computations, we set

$$\alpha_s = \frac{\alpha_0}{\sqrt{1 + s}} \quad .$$

Theorem 2. Let F and h be convex, continuously differentiable functions, $g(x)$ be an affine function, Z be a non-empty compact set, and the inequalities [21] hold. Then method [20] converges globally to solution set Z for any $x_0 \in E_n$. Discrete method [22] globally converges to Z if α_s is a monotonically decreasing sequence satisfying [23] and if α_0 is sufficiently small.

In Evtushenko (1975b), other methods are presented which are based on a transformation of [14] into an unconstrained maximin problem, by using generalized Lagrange multiplier techniques. We define the modified Lagrangian function $H(x,p,w)$ associated with problem [14] as

$$H(x,p,w) = F(x) + \sum_{j=1}^{e} p^j g^j(x) + \sum_{j=1}^{c} (w^j)^2 h^j(x) \quad ,$$

where $p \in E_e$, $w \in E_c$. The simplest gradient yields the following method

$$\dot{x} = - H_x \quad , \quad \dot{p} = H_p \quad , \quad \dot{w} = H_w \quad , \quad (\cdot) = d/dt$$

$$x(0) = x_0 \quad , \quad p(0) = p_0 \quad , w(0) = w_0 \quad . \qquad [24]$$

In Evtushenko (1975b), it is proved that the solution $x(t)$, $p(t)$, $w(t)$ of system [24] locally converges to that of [14] as $t \to \infty$ under some conditions.

2.2 Numerical Methods for Solving Multistage Optimal Control Problems

We now consider a dynamical system described by the difference equation

$$x_{i+1} = f_i(x_i, u_i) \qquad i = 0,1,\ldots I-1 \quad , \qquad [25]$$

where $x_i = (x_i^1, x_i^2, \ldots, x_i^K)$ which is a point in E_K. The control applied to the system at time i is $u_i = [u_i^1, u_i^2, \ldots, u_i^r]$ which is a point in E_r; f_i is a real-valued continuously differentiable vector function, $f_i = [f_i^1, f_i^2, \ldots, f_i^K]$, defined on the Cartesian product $E_K \times E_r$; I is the duration of the control process. The problem is to find control sequence $u_0, u_1, \ldots, u_{I-1}$ and a corresponding trajectory $x_1, x_2, \ldots, x_I$ determined by [25], which minimizes the cost function

$$\sum_{i=1}^{I} R_i(x_i, u_i) \qquad [26]$$

subject to constraints

$$g_i(x_i, u_i) = 0 \quad , \quad h_i(x_i, u_i) \le 0 \quad , \quad i = 1,2,\ldots,I \quad , [27]$$

where g_i, h_i are given functions,

$$h_i = (h_i^1, h_i^2, \ldots, h_i^e) \in E_c \quad ,$$

$$g_i = (g_i^1, g_i^2, \ldots, g_i^e) \in E_e \quad .$$

For solving the primal problem we shall use modified penalty function methods. For simplicity in formulas [18], setting $\psi(y) = y^2$, we obtain

$$P(x,u,\tau) = \sum_{i=1}^{I} P_i(x_i,u_i,\tau)$$

$$= \sum_{i=1}^{I} \left[R_i(x_i,u_i) + \tau\left\{\| g_i(x_i,u_i)\|^2 + \| h_{i+}(x_i,u_i)\|^2\right\}\right] \quad ,$$

where $h_{i+}^j(y) = \mathrm{MAX}\ [0, h_i^j(y)]$, $\| a\|^2 = \sum_{i=1}^{\ell} (a^i)^2$; here $a = [a^1, a^2, \ldots, a^\ell]$.

If functions f_i, F_i, S_i, h_i are differentiable, then using a common procedure (Polak, 1971), the following formulas for derivatives can be written:

$$\frac{dP}{du_i} = \frac{\partial P_i}{\partial u_i} + \frac{\partial f_i}{\partial u_i} p_{i+1}$$

where $\partial f_i/\partial u_i$ is $r \times n$ matrix whose j,s^{th} element is $\partial f_i^s/\partial u_i^j$. The n-dimensional multiplier vectors $p_1, p_2, \ldots, p_I$ satisfy the following difference equations:

$$p_I = \frac{\partial P_I}{\partial x} \quad , \quad p_i = \frac{\partial f_i}{\partial x_i} p_{i+1} + \frac{\partial P_i}{\partial x_I} \quad , \qquad [28]$$

where $\partial f_i/\partial x_i$ is $n \times n$ matrix whose j,s^{th} element is $\partial f_i^s/\partial x_i^j$.

Now for solving the primal ([25], [26], and [27]), we can use any method presented above. For example, using [20] yields

$$\frac{du_i(t)}{dt} = - \frac{d}{du_i} P(x(t),u(t),\tau(t)) \quad , \quad i = 0,1,\ldots,I-1$$

[29]

$$\frac{d\tau(t)}{dt} = \tau(t) \ \theta(T - \tau(t)) \ \frac{d}{d\tau} P(x_i(t),u_i(t),\tau(t)) \quad .$$

In performing numerical calculations, instead of this continuous system we shall solve the corresponding discrete version:

$$u_i(s + 1) = u_i(s) - \alpha_s \frac{d}{du_i} P(x(s),u(s),\tau(s))$$

[30]

$$\tau(s + 1) = \tau(s) + \alpha_s \ \tau(s) \ \theta(T - \tau(s)) \ \frac{d}{d\tau} P(x(s),u(s),\tau(s)) \quad .$$

In this system, initial control vector $u(0)$ is given. By solving system [25], we find the corresponding trajectory $x_i(0)$ $(i = 1,2,\ldots,I)$. After this, we solve the difference equation [28] from $i = I,I-1,\ldots,1$, simultaneously changing control variables in accordance with system [30]. After this first step, we find a new control vector $u(1)$.

Again solving [25], we find the corresponding state vector $x_i(1)$ $(i = 1,2,\ldots,I)$. Next, we move backward again and so on.

Theorem 2 can be easily reformulated for this particular case. The method in [24] gives the following system:

$$\frac{du_i(t)}{dt} = - \frac{dH(x(t),u(t),\lambda(t),w(t))}{du_i}$$

$$\frac{d\lambda_i^j(t)}{dt} = g_i^j(x_i(t),u_i(t)) \qquad [31]$$

$$\frac{dw_i^j(t)}{dt} = 2w_i^j(t)\ h_i^j(x_i,u_i) \quad ,$$

where $H = \sum_{i=1}^{I} \left[R_i(x_i,u_i) + \sum_{j=1}^{e} \lambda_i^j g_i^j(x_i,u_i) \right.$

$$\left. + \sum_{j=1}^{c} (w_i^j)^2\ h_i^j(x_i,u_i) \right] \quad .$$

This method locally converges to the solution of a primal problem. Convergence theorems for the multistage case could be formulated as in theorems 1 and 2, but they would be considerably more cumbersome.

2.3 A National Settlement System Problem in Modified Penalty Function Form

We now consider, explicitly, an optimal migration problem. In this case, as represented by [1], [2], [4], and [11], we have

$$x_{i+1}^k = \sum_{j=1}^{K} M^{kj} x_i^j + u_i^k \quad , \qquad i = 0,1,\ldots,I-1$$

$$\sum_{i=0}^{I-1} \sum_{k=1}^{K} b_i^k u_i^k \le B \quad .$$

Therefore the penalty function has the following form:

$$P = \sum_{i=1}^{I} \sum_{k=1}^{K} a^k \left(x_i^k - \hat{x}_i^k \right)^2 + \tau \left[\left(\sum_{k=1}^{K} \sum_{i=0}^{I-1} b_i^k u_i^k - B \right)_+^2 \right] \quad .$$

System [28] thus has the form:

$$p_I^j = 2a^j (x_I^j - \hat{x}_I^j)$$

$$p_i^j = 2a^j (x_i^j - \hat{x}_i^j) + \sum_{\ell=1}^{K} M^{\ell s} p_{i+1}^{\ell} \quad .$$

Method [30] has the form:

$$u_i^k(t+1) = u_i^k(t) - \alpha_t \left[p_{i+1}^k + 2\tau b^k \left(\sum_{j=1}^{K} \sum_{i=0}^{I-1} b_i^j u_i^j(t) - B \right)_+ \right]$$

$$\tau(t+1) = \tau(t) + \alpha_t \left[\left(\sum_{i=0}^{I-1} \sum_{k=1}^{K} b_i^k u_i^k(t) - B \right)_+^2 \right] \tau(t) \; \theta(T - \tau(t)) \quad .$$

2.4 Some Comments on the Numerical Performance of the Method

Unlike many mathematical programming methods the penalty function methods presented here require that the programmer use some judgment to increase the effectiveness of the solution methods. As has been shown, convergence is assured as $t \to \infty$ but the speed of convergence or the closeness to optimality for a given t critically depends on the initial solution and the choice of two parameters τ_0 and α_0.

Any intuition which the modeller can use concerning good initial values of $u_i (i = 0,1,\ldots,I-1)$ may significantly speed up convergence. For simplicity, the u_i^k terms may be set equal to zero except for a few--those, for example, which have large negative values of the difference $x_0^k - \hat{x}_1^k$. It should, of course, be emphasized that the initial solution need not be feasible.

The choices for the parameters τ_0 and α_0 are more subtle and some experimentation is usually necessary before good values can be selected. If τ_0 is "too small," the algorithm will be insensitive to violations of the constraints. If, on the other hand, τ_0 is "too large," a violation of the constraint set at step t will result in an over-reaction in the next time period; i.e., in the attempt to obtain feasibility, the objective function will be increased to unnecessarily large values. Figure 1 gives an example of two alternative values of τ_0 for the case where $B = 15 \times 10^5$. Clearly, J_t, the value of the objective function at step t, appears to be converging more rapidly for $\tau_0 = .002$ than for $\tau_0 = .1$. The latter value results in greater instability of subsequent solutions.

Similarly, "too large" values of α_0 will result in large changes in the control variables from one solution to the next. Large changes may result in "overshooting" the optimal solution, whereas small adjustments may result in slow convergence.

The fact that with both choices experiments are useful, and in most cases necessary, certainly implies that good solutions could be much more readily obtained with an interactive computer system.

In our limited experience with this method applied to a problem with twenty-two state variables and four time periods (i.e., effectively eighty-eight state variables), near-optimal solutions were obtained in eighty steps using about twenty seconds of C.P.U. time on a C.D.C. 6600 computer.

3. Interpretation of the Numerical Results

As a numerical example of the procedures outlined above, only one constraint, other than the non-negativity constraints, has been imposed. This, the budget constraint [2], limits the total effort which the planning agency can apply in order to control the system. Moreover, a single time-invariant population target has been postulated for each of the cities rather than more conventional trend lines. Arising directly from these specifications is a pattern of temporal

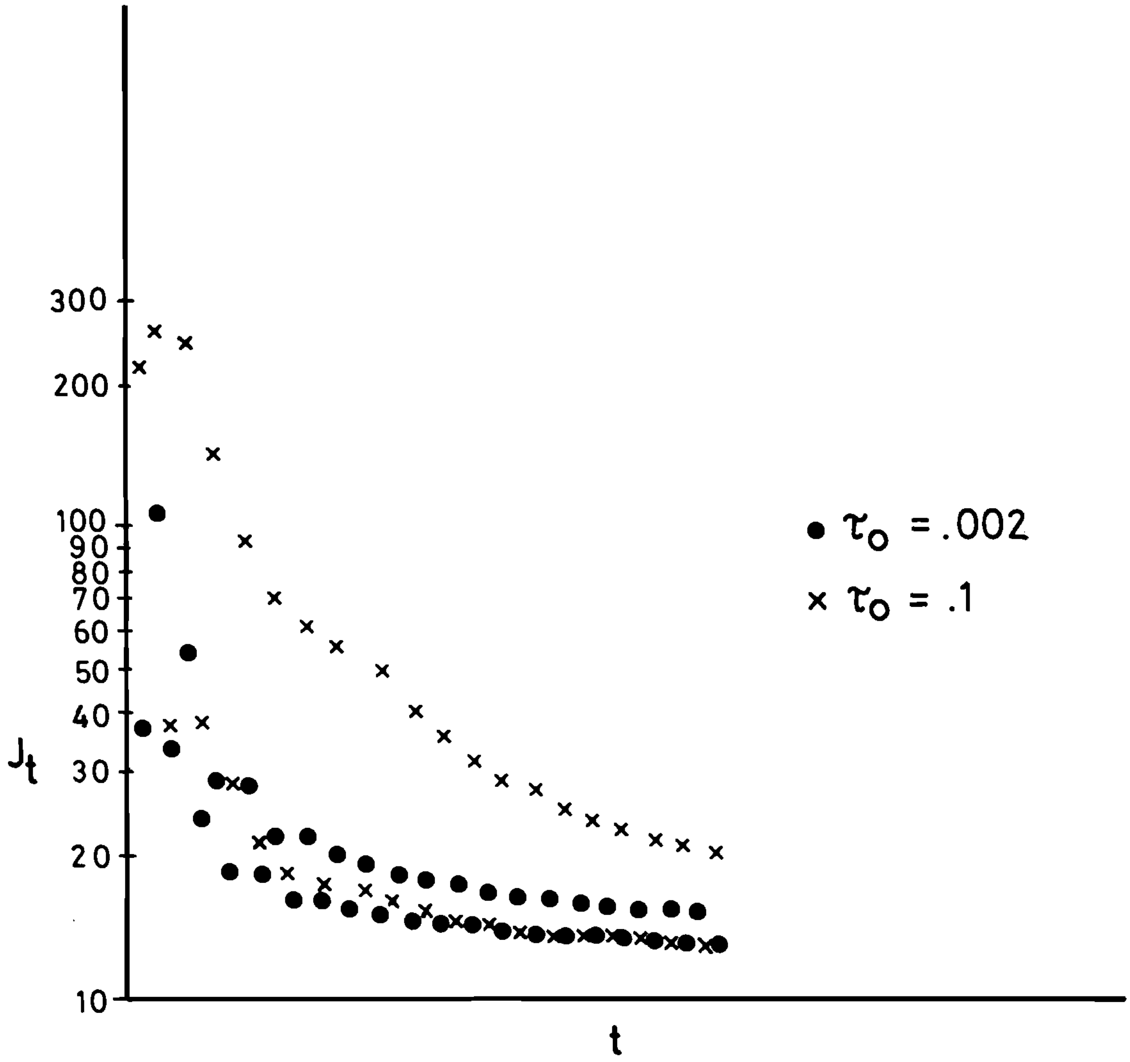


FIGURE 1. CONVERGENCE PROPERTIES FOR TWO DIFFERENT VALUES OF τ_o.

allocation of effort which is highly peaked in the initial time periods, falling off at an exponential rate in most cases. Inputs are made to the system in order to move the system close to its targets as soon as possible. The only important factor which encourages the temporal spread of controls is that some cities, having come close to their targets, immediately begin losing population. Thus, new inputs must be channelled into these cities to help make up for losses due to out-migration. Some metropolitan areas (e.g., Toronto, Calgary and Ottawa) which are either growing or only slowly declining due to net in-migration have highly peaked controls, while others (e.g., St. John's, Saint John and Winnipeg) have a more even distribution of controls.

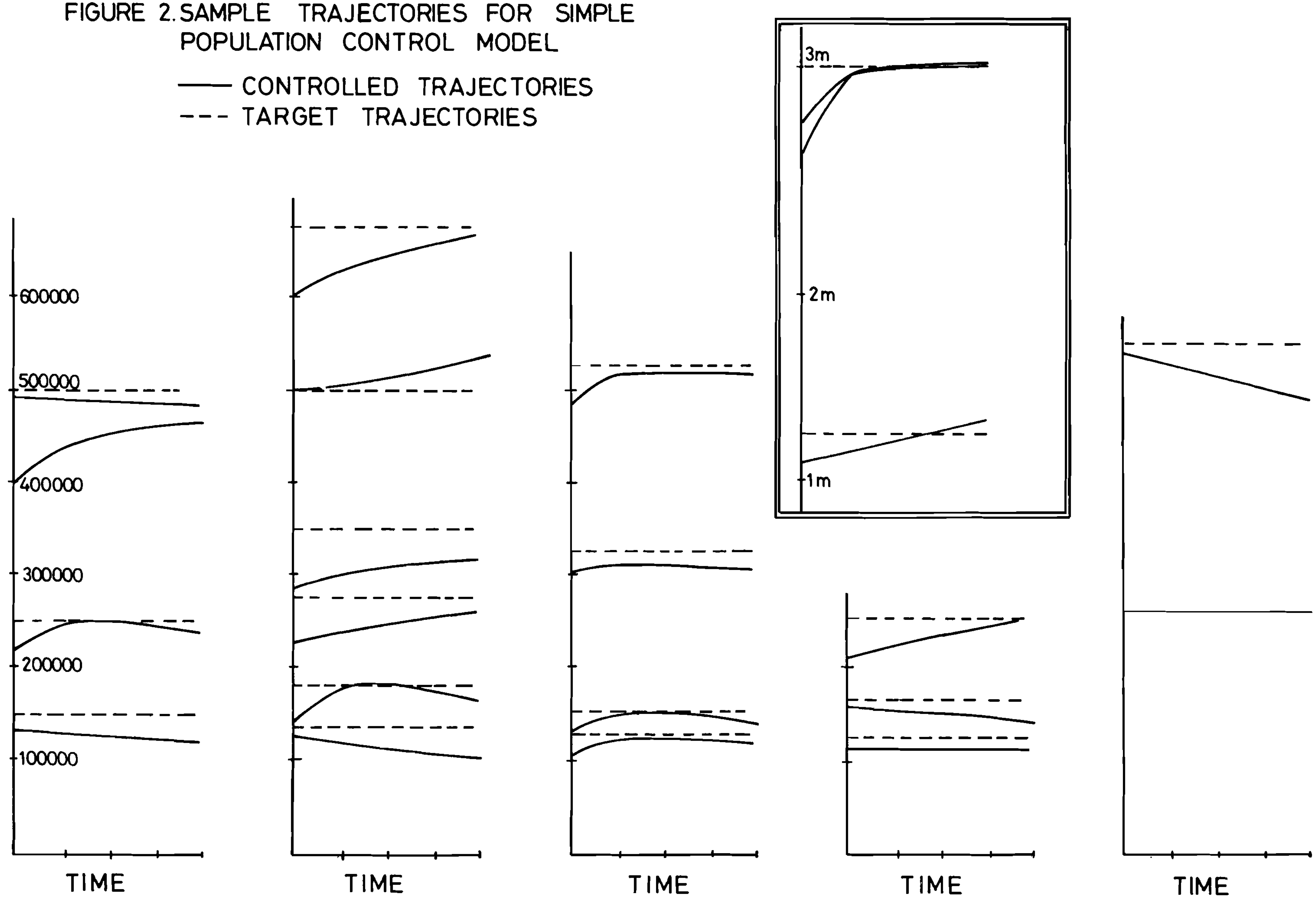
Figure 2 displays the trajectories for the state variables for $B = 20 \times 10^5$. Any thorough interpretation of the model would have to include a discussion of the interaction between "natural" system dynamics, controls imposed on the system, the postulated goals of the system and their relative importance, and the budgetary constraint. Each of these factors can directly influence the shape of the trajectory for any of the cities. For example, some of the cities are gaining population through inter-city migration; if they are close to their goals (e.g., Hamilton and Vancouver), no controls will be applied as this would involve overshooting targets in later periods. Other cities are initially far short of their targets: of these, some (e.g., Montreal, Halifax, Saint John, St. John's, Regina) are losers through migration; others such as Toronto and Ottawa are stable or net gainers. Both categories of cities have similar trajectories for sufficiently large values of B, even though the pattern of controls is markedly different. Other cities are losing, but for purposes of this study have relatively small weights attached to goal attainment (e.g., Saskatoon, Winnipeg). The trajectories of these cities are either constantly declining or concave upwards depending on their initial deviation from targets. Finally, there are some urban areas which are initially very close to their targets and are approximately in equilibrium with the rest of the system (e.g., Windsor, Edmonton, Thunder Bay, Sudbury). These cities, of course, have approximately flat trajectories with little or no controls.

The best use of the model in its present form is to demonstrate the ways in which different goals and goal weights trade off against each other and interact with cost functions, budgetary constraints and the inherent dynamics of the system.

FIGURE 2. SAMPLE TRAJECTORIES FOR SIMPLE POPULATION CONTROL MODEL

—— CONTROLLED TRAJECTORIES

- - - TARGET TRAJECTORIES



From these initial empirical results, it appears that there is a rather simple relationship between system goals and controls. That is, the regional interactions, as incorporated in the migration propensity matrix M, while taken into account, are rather unimportant compared with the direct control vectors u_i. Indirect consequences of these controls are rather minimal. This is directly related to the strong diagonal dominance of M. While, in principle, this makes planning rather simple (assuming u_i are controllable), in a sense it makes the interpretation of the model much less interesting.

Of considerable interest is the shape of the cost-effectiveness curve associated with the system; that is, the manner in which the objective function decreases with increasingly large applications of effort. Figure 3 shows very strikingly the diminishing marginal productivity of controls. The curve is extremely concave, reaching a saturation point at $B = 34.34 \times 10^5$ where additional budgetary allocations will not be used; that is, the constraint [2] will not be binding.

4. Closure

In a very real sense, this study is at a rather preliminary stage with only the simplest of all the models having been implemented. The other formulations all appear to be feasible although their data requirements and/or their computational complexities are greater. Of considerably greater policy interest are the "backward linkage" and "variable structure" cases. It is hoped that this computational approach will be adapted and extended to solve these more interesting problems.

Regarding the computational procedures, two comments should be made. First, it would be possible to increase the effectiveness of the method by adjoining it to another procedure (e.g., Newton's method) when the decrease in subsequent values of the objective function becomes small. That is, the penalty function methods used here give only approximately optimal solutions; if greater accuracy is required, the methods of this paper should be used to generate good starting solutions for other, more accurate (but more time-consuming) procedures. It is, however, unlikely that the reliability of the data in these applications is sufficient to warrant the use of more accurate methods. The second comment regarding computational matters is that there are

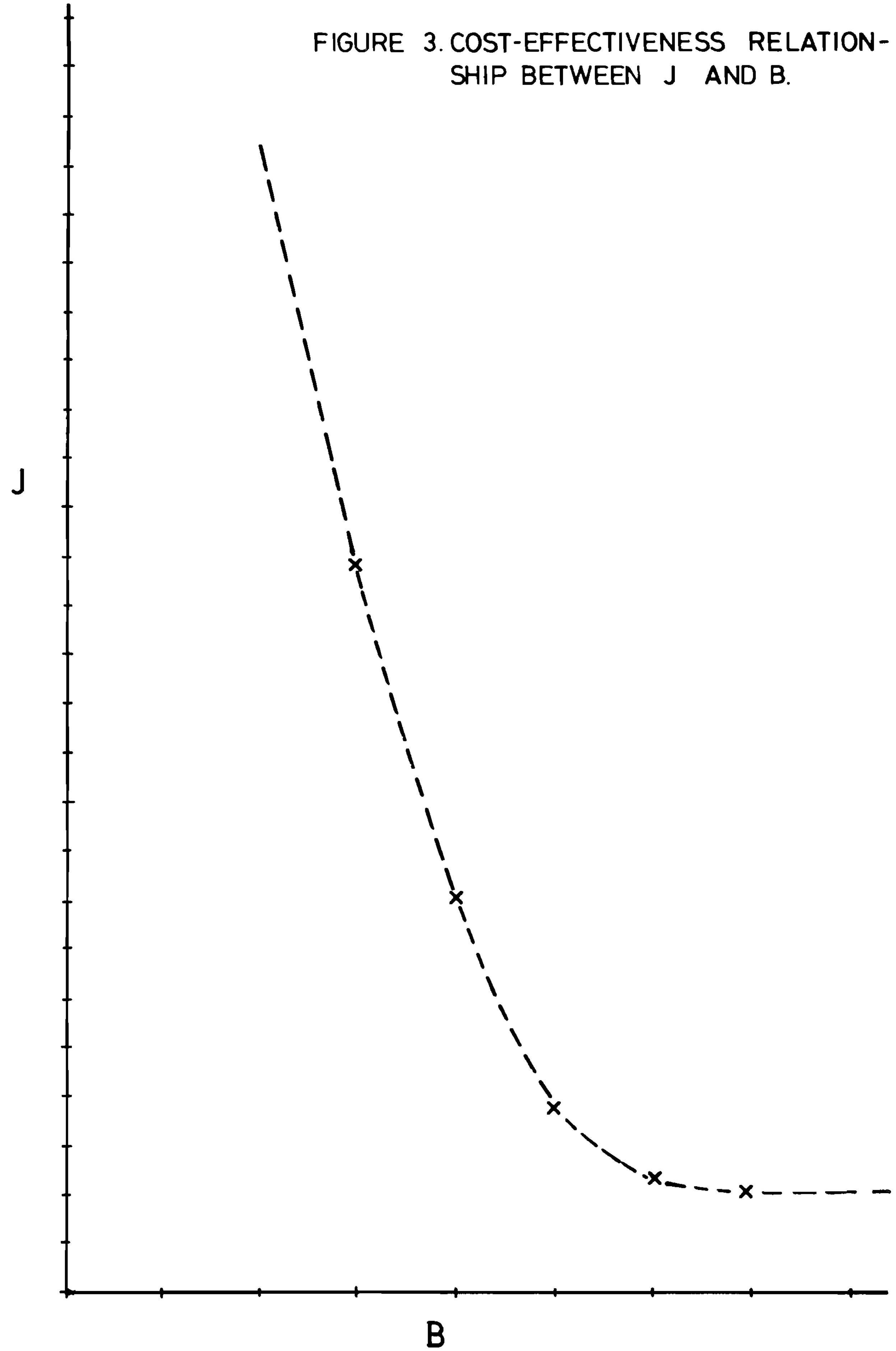


FIGURE 3. COST-EFFECTIVENESS RELATIONSHIP BETWEEN J AND B.

other methods of mathematical programming, control theory and even the classical calculus of variations which can be applied to this class of problems described. The problem for which numerical results have been obtained in this study appears to be particularly amenable to Lagrangian type of analysis*. Variable-structure, or bi-linear, systems methods could be applied to this problem formulation (see Mohler, 1974).

There are a number of ways in which the models presented here could be enriched; some of these are enumerated in MacKinnon (1975a). Most of the straightforward extensions imply a considerable increase in the dimensionality of the system; thus some control theory and dynamic programming approaches would appear to be inappropriate candidates for obtaining numerical results for such problems.

The uncertainty associated with the system parameters and even the general structure of system relationships should be explicitly considered. One method is, of course, extensive experimentation to determine the sensitivity of controls and goal attainment to _ad hoc_ changes in the parameters. Alternatively, the parameters themselves can be assumed to be random variates with known probability distributions. Although this approach may be feasible only for a limited number of parameters, it is apparent that some incorporation of uncertainty in the model would be useful.

Clearly then there are many ways in which the models could be elaborated and there are several alternative solution methods. In addition, there are some other urban and regional problems which could be formulated within a similar methodological framework. The diffusion of information throughout an urban system can be modelled by means of a matrix operator. Controls can be either locationally specific information inputs to the system or alterations in the matrix operator itself. The goal may be to obtain a reasonably equitable distribution of information throughout the system by a specified time. (See MacKinnon 1975b for a more formal representation of this problem.) A second possible application could be in urban land use structure. Land use changes within a city can be represented by a matrix operator where the states

---

*Professor Raman Mehra of IIASA is currently working out the details of this approach.

are mutually exclusive land use categories (e.g., Bourne, 1969). Controls may be in the form of changes in the matrix elements (via land use conversion regulations) or inputs to the system (via construction regulations on vacant land). The goals may be stated in terms of desirable mixes of land use for different parts of the city.

In conclusion then, it would appear that non-linear programming has become a computationally feasible and quite flexible approach for undertaking experiments which generate alternative national settlement system planning strategies. These experiments yield not precise prescriptions but rather general indications of the magnitude and direction of controls which are necessary to achieve specified goals. Clearly additional research is necessary on goal and cost function specification, and on a more detailed elaboration of the dynamics of the system, including an indication of the precise instrumental variables which can be used to effect changes in system inputs or system structure. Although much research remains to be done, it is argued that the approach presented here provides a promising, computationally feasible methodological framework within which these future research results can be placed.

Acknowledgements

The completion of this paper was made possible by the release of one week of Evtushenko's time by the Computing Centre in Moscow. MacKinnon's financial support by the Canada Council in the form of a leave fellowship is gratefully acknowledged. The rather difficult task of transforming the untidy manuscript into final copy was accomplished successfully and patiently by Linda Samide.

References

[1] Bourne, L.S. 1969. A spatial allocation - land use conversion model of urban growth. Journal of Regional Science. 9:2:261-272.

[2] Eremin, I.I. 1967. Penalty function techniques for convex programming. Kibernetika. 4:63-67 (in Russian).

[3] Evtushenko, Y. 1975a. Algorithms for solving non-linear programming problems. Optimization Techniques IFIP Technical Conference at Novosibirsk, July 1974. Berlin, Springer-Verlag, pp. 308-313.

[4] Evtushenko, Y. 1975b. Generalized Lagrange multipliers technique for non-linear programming. IIASA Research Report 75-13, Laxenburg, Austria.

[5] Fiacco, A.V. and G.P. McCormick. 1968. Non-Linear Programming: Sequential Unconstrained Minimization Techniques. New York, Wiley.

[6] MacKinnon, R.D. 1975a. Controlling migration processes of a Markovian type. Environment and Planning (in press).

[7] MacKinnon, R.D. 1975b. Geographical diffusion processes: Alternative methodological approaches of an operational type. Revista Geografica (in Portuguese).

[8] Mohler, R.R. 1973. Bilinear Control Processes with Application to Engineering, Ecology and Medicine. New York, Academic Press.

[9] Nykamp, P. and W.H. Somermeyer. 1974. Explicating implicit social preference functions. Economics of Planning. 11:3:101-119.

[10] Polak, E. 1971. Computation Methods in Optimization. New York, Academic Press.

[11] Skarin, V.D. 1973. Penalty function techniques for convex programming. Zhurnal vych. matematiki i matematicheskoy fizicki. 13:5:1186-1199 (in Russian).