ELSEVIER

Transportation Research Part A 38 (2004) 347–365

TRANSPORTATION RESEARCH PART A

www.elsevier.com/locate/tra

Optimal scheduling of rehabilitation activities for multiple pavement facilities: exact and approximate solutions

Yanfeng Ouyang [1], Samer Madanat *

Department of Civil and Environmental Engineering, University of California at Berkeley, 114 McLaughlin Hall, ITS, Berkeley, CA 94720-1712, USA

Received 2 August 2002; received in revised form 21 April 2003; accepted 7 October 2003

Abstract

This paper presents a mathematical programming model for optimal highway pavement rehabilitation planning which minimizes the life-cycle cost for a finite horizon. It extends previous researches in this area by solving the problem of multiple rehabilitation activities on multiple facilities, with realistic empirical models of deterioration and rehabilitation effectiveness. The formulation is based on discrete control theory. A nonlinear pavement performance model and integer decision variables are incorporated into a mixed-integer nonlinear programming (MINLP). Two solution approaches, a branch-and-bound algorithm and a greedy heuristic, are proposed for this model. It is shown that the heuristic results provide a good approximation to the exact optima, but with much lower computational costs.

1. Introduction

This paper addresses the problem of determining the optimal rehabilitation frequency and intensity on a system of pavements. This problem has received substantial attention from researchers in the field of infrastructure management. In this paper we focus exclusively on the continuous state, multiple-facility problem, and the case of deterministic deterioration and rehabilitation-only policy. The problem is defined as follows: given a deterministic deterioration process and rehabilitation effectiveness, what is the frequency and intensity of the rehabilitation activities that minimize the total discounted life-cycle cost over a planning horizon?

* Corresponding author. Tel.: +1-510-643-1084; fax: +1-510-642-1246.
E-mail addresses: yfouyang@uclink.berkeley.edu (Y. Ouyang), madanat@ce.berkeley.edu (S. Madanat).
[1] Tel.: +1-510-642-9907.

0965-8564/$ - see front matter
doi:10.1016/j.tra.2003.10.007

This paper extends the state of the art in two directions. On the modeling side, earlier research focused on the simpler problem of a single facility under either maintenance or a single rehabilitation over the planning horizon. More recent contributions addressed the problem of multiple rehabilitation activities. However, the models used were based on unrealistic models of pavement performance or rehabilitation effectiveness. The present paper extends previous research by solving the general problem of multiple rehabilitation activities on multiple facilities, and by incorporating realistic empirical models of deterioration and rehabilitation effectiveness.

This paper is organized as follows. We first present a review of the relevant literature on the deterministic problem and clarify the relation of this research to previous work. Then we formulate a mathematical optimization model for the problem and present two solution algorithms. Finally we present parametric studies and a comparison between these algorithms.

2. Literature review

The problem of optimal pavement management for the continuous time and continuous state case has been addressed by a number of researchers. A common approach is based on the theory of optimal control (Bryson and Ho, 1975). Friesz and Fernandez (1979) first formulated the problem of finding an optimal maintenance profile over time as an optimal control problem. Maintenance applications were represented by piecewise, continuous investments over time. They later (Fernandez and Friesz, 1981) extended their work to the problem of optimal timings of highway stage construction. Markow and Balta (1985) used optimal control to solve for the optimal timing of pavement rehabilitation. Because rehabilitation actions lead to discontinuous changes in pavement serviceability, which complicates the problem, the authors provided a solution only for the case of a single rehabilitation action.

These applications can be extended to solving for the optimum strategy for multiple actions when the number of actions is predetermined. Closed-form optimal solutions can be obtained analytically in such models.

To avoid the difficulty associated with the discontinuities in pavement condition caused by rehabilitation, Tsunokawa and Schofer (1994) used a trend-curve approximation method to solve the problem of optimal timings and intensity of pavement resurfacing activities. Their approach was to approximate the saw-tooth curve of pavement condition by a continuous function passing through the midpoints of the spikes. Solving the resulting standard optimal control problem, the value of resurfacing frequency and intensity can be calculated. However, as pointed out by Li and Madanat (2002), such approximation leads to problematic solutions. Li and Madanat (2002) used Tsunokawa's formulation and data to come up with a simpler approach to solve for optimal rehabilitation policy in steady state problems. One of their findings is that the discounted lifetime cost is not very sensitive to cycle time.

Another possible approach is to use mathematical programming. Taking into account the fact that rehabilitation only occurs at certain moments, e.g. the beginning of a budget year, it is realistic to discretize the planning horizon and formulate the rehabilitation problem as a mixed-integer mathematical program. Murakami and Turnquist (1985) used a dynamic model for allocating maintenance resources across several facilities over multiple time periods. Their model maximizes an overall measure of system performance under resource availability constraints, rather than life-cycle costs minimization. Al-Subhi et al. (1990) and Jacobs (1992) applied similar mixed-integer mathematical models to optimally schedule long-term bridge deck rehabilitation and replacement activities. Jacobs simplified the problem to a mixed-integer linear program by taking rehabilitation intensity as constant and the deterioration curve as piecewise linear. However, in reality, nonlinearities arise from various sources such as facility deterioration and rehabilitation effectiveness.

The formulation of the optimal infrastructure management problem must be based on realistic performance models that describe the behavior of the facilities. For our purposes, we are concerned with deterministic models of pavement deterioration and rehabilitation effectiveness. Friesz and Fernandez (1979) used simple multiplicative deterioration factors to describe the change of roughness over time, taking roughness as the index of condition. The flaw of this simple deterioration assumption is that it implies no deterioration when the roughness index is zero. Markow and Balta (1985) used an additive term together with the multiplicative part. Paterson (1987) reviewed existing empirical models of pavement deterioration over time. Paterson (1990a) also quantified the effectiveness of pavement maintenance and rehabilitation. Taking roughness as an index of condition, pavement overlay effectiveness was found to be a function of the overlay thickness (also called rehabilitation intensity) and the condition before overlay. His models will serve as the empirical basis of our study.

3. Model formulation

As described in literature, the roughness of a given highway pavement follows a saw-tooth trajectory over time as the pavement deteriorates and is resurfaced.

In practice, highway agencies make maintenance and rehabilitation decisions subject to budget constraints and resource availability. They allocate resources for rehabilitation activities at the beginning of a budgeting year. It is realistic, therefore, to discretize the planning horizon into predetermined temporal stages (e.g. years) and restrict rehabilitations to occur only at such time points. The deterioration process on the other hand, is represented as a continuous trend over time.

The aforementioned discretization leads to the application of discrete multiple-stage optimal control theory. A graphical representation of such models is shown in Fig. 1.

In Fig. 1, $A(t)$ represents the action at time t, $s(t)$ represents facility condition at time t, finally $f_t(A(t), s(t))$ is the function that determines condition $s(t+1)$ based on the previous condition $s(t)$ and the action $A(t)$.

3.1. Objective function and constraints

The highway agency decides for each pavement segment in each year whether to perform rehabilitation and how intense the rehabilitation should be. The goal of a highway agency is

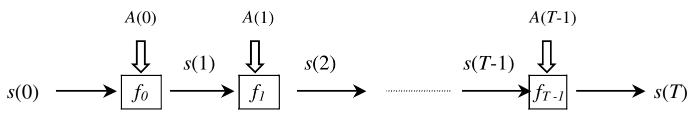

Fig. 1. Flow chart for a multi-period single-facility problem.

to minimize the net present value of life-cycle costs for the agency and the users over a finite horizon.

In a highway network of N pavement facilities, over a planning horizon containing T periods, the life-cycle cost minimization is formulated as follows:

$$\text{Min } J(N,T) = \sum_{t=0}^{T-1} \sum_{n=1}^{N} \left\{ \int_{t}^{t+1} C_n(s_n(u))\mathrm{e}^{-ru}\,\mathrm{d}u + X_{tn} \cdot M_{tn}(w_{tn})\mathrm{e}^{-rt} \right\} \tag{1.1}$$

$$\text{s.t.} \quad s_n[(t+1)^-] = F_n[s_n(t^+)] \tag{1.2}$$

$$s_n(t^-) - s_n(t^+) = G_n(w_{tn}, s_n(t^-)) \tag{1.3}$$

$$0 \leqslant w_{tn} \leqslant w_{\max}(s_n(t^-)) \cdot X_{tn} \tag{1.4}$$

$$\sum_{n=1}^{N} M_{tn}(w_{tn}) \leqslant B(t) \tag{1.5}$$

$$s_n(0) = s_{0n}, \quad s_n(T) \leqslant s_{Tn} \tag{1.6}$$

$$X_{tn} \in \{0,1\} \tag{1.7}$$

where

n index of pavement facility, $n = 1, 2, \ldots, N$

t discrete time for rehabilitation activities, taken to be a year, $t = 1, 2, \ldots, T$

t^+, t^- the moment right after/before time t, as shown in Fig. 2

X_{tn} binary variable for rehabilitation decision on pavement n at time t

w_{tn} non-negative rehabilitation intensity on pavement n at action time t

$s_n(t^+)$, $s_n(t^-)$ roughness of pavement facility n at t^+ and t^- respectively. In the absence of rehabilitation, $s_n(t^+) = s_n(t^-)$ and we simply write it as $s_n(t)$

$w_{\max}(s_n(t^-))$ the maximum effective rehabilitation intensity

$J(N,T)$ the present value of agency and user costs incurred on N pavement segments over a planning horizon of T years

$F_n(s_n(t^+))$ the pavement roughness function if no rehabilitation is taken

$G_n(w_{tn}, s_n(t^-))$ the rehabilitation effectiveness as a function of resurfacing intensity and the before-action pavement roughness

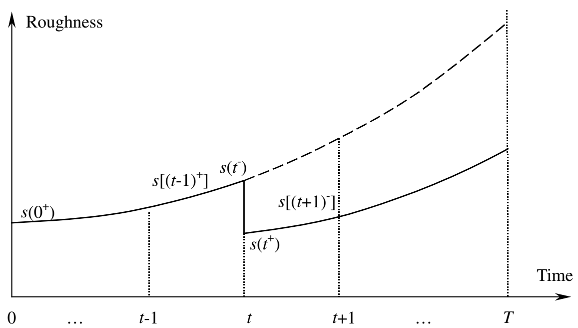

Fig. 2. Roughness deterioration and improvement under rehabilitation at t.

$C_n(s_n(t))$ the user cost rate on pavement n as a function of pavement roughness
$M_{tn}(w_{tn})$ the cost of rehabilitation for pavement n at time t, as a function of the rehabilitation intensity w_{tn}
s_{0n} the initial condition of pavement n at the beginning of the planning horizon
s_{Tn} the required pavement condition at the end of the planning horizon
r discount rate; the continuous discount factor e^{-rt} discounts the costs into present values
$B(t)$ the budget constraint in fiscal year t.

Formulation (1.1)–(1.7) is referred to in the literature as a mixed-integer nonlinear programming model, which involves both integer variables and nonlinear functions. The objective function $J(N,T)$ is simply the sum of total discounted user and agency costs over all facilities in the planning horizon. The discounted user cost is incurred continuously and is represented by the integral in (1.1). The agency cost, the second term in (1.1), is incurred at discrete times. Constraint (1.2) defines the continuous pavement deterioration process. Constraint (1.3) gives the rehabilitation effectiveness. Constraint (1.4) defines the effective range of rehabilitation intensity w_{tn} as well as its logical relationship with the binary decision variable X_{tn}. Constraint (1.5) restricts agency expenditure to the annual budget. Constraint (1.6) sets the boundary conditions of pavement roughness at the beginning and the end of the planning horizon.

The functions in this formulation, i.e., $C_n(s_n(t))$, $M_{tn}(w_{tn})$, $F_n(s_n(t^+))$, $G_n(w_{tn}, s_n(t^-))$, and $w_{\max}(s_n(t^-))$, are derived from empirical studies and are described in the next sections. Additional information about the constraints is also provided there. The subscript n is often dropped for convenience, referring to the performance of a general facility.

3.1.1. Pavement deterioration

Pavement deterioration is a continuous process between rehabilitations. In the absence of rehabilitation, the deterioration rate $\frac{ds}{dt}$ is generally believed to be a function of pavement condition s. Tsunokawa and Schofer (1994), assumed a simple linear deterioration rate $\frac{ds}{dt} = \beta s$ which is equivalent to an exponential deterioration function

$$s(t) = s(t_0) \cdot \exp[\beta(t - t_0)] \quad \text{for } t \geqslant t_0 \tag{2}$$

If time is discretized, Eq. (2) can be transformed into the recursive form:

$$s(t+1) = s(t) \cdot \exp(\beta), \quad t = 0, 1, 2, \ldots \tag{3}$$

This deterioration process has a Markovian property because the deterioration rate only depends on current pavement condition. The recursive formula is very suitable for mathematical programming. However, this formulation is not realistic for pavements in good conditions. Specifically, if $s(t^*) = 0$ at a certain t^*, then the pavement never deteriorates and $s(t) = 0$ for all $t > t^*$.

Paterson (1990a) gives an empirical relationship for flexible pavement deterioration, which does not suffer from the above problem. It is based on data collected in Brazil by the World Bank as part of the development of the HDM III model:

$$s(t) = [s(t_0) + f(t - t_0)] \cdot \exp[\beta(t - t_0)] \tag{4}$$

In Eq. (4), $f(t - t_0)$ is a function of the structural number of the pavement section as well as the cumulative traffic loading since the last rehabilitation. If constant traffic demand is assumed for a given pavement facility, $f(t-t_0)$ turns out to be a linear function of age (Paterson, 1987). This formulation says that pavement deterioration depends not only on its current condition, but also on its rehabilitation history. Including such non-Markovian relationships in a control model complicates the solution process.

In Paterson's formula, the term β in the exponent is a very small constant. The relationship can be closely approximated by a simpler one. In fact, if we select an appropriate constant $\bar{f}$ to substitute for $f(t-t_0)$ in Paterson's formula, the deterioration rate remains almost constant during short periods. The approximation yields

$$s(t) = \left(s(t_0) + \bar{f}\right) \cdot \exp[\beta(t-t_0)] - \bar{f}, \quad t \geqslant t_0 \tag{5}$$

If time is discrete, Eq. (5) can be transformed into the recursive form:

$$s(t+1) = \left(s(t) + \bar{f}^*\right) \cdot \exp(\beta), \quad t = 0, 1, 2, \ldots \tag{6}$$

The introduction of parameters $\bar{f}^*$ and $\bar{f}$ into the deterioration models can be viewed intuitively as including an average deterioration trend which is independent of the current pavement condition. The deterioration is now a combination of a fixed trend and a function of current condition. Obviously Eq. (6) is more realistic than Eq. (3). It can be easily shown that the parameters $\bar{f}^*$ and $\bar{f}$ satisfy the following relationship:

$$\bar{f}^* = \bar{f} \cdot (1 - e^{-\beta}) \tag{7}$$

3.1.2. Rehabilitation effectiveness

An important characteristic of pavement and many other transportation facilities is that their condition declines nonlinearly over time. Because of this characteristic, the timing of rehabilitation expenditures is very important. The marginal effectiveness of maintenance expenditures depends on the condition of the facility when the maintenance is performed. Paterson (1990a) developed models for pavement rehabilitation effectiveness under various conditions, which are functions of rehabilitation intensity and the roughness before action. With regard to the commonly used pavement overlay by automatic-leveling paver-finishers, an empirical relationship is described by Fig. 3.

It is shown in Fig. 3 that given the roughness before an overlay, there is a maximum possible effectiveness beyond which additional rehabilitation work does not have an effect. In Fig. 4, the information in Fig. 3 is translated by plotting the effectiveness as a function of the rehabilitation intensity, which gives a clearer view of pavement rehabilitation effectiveness. A new set of units (QI) is used rather than the International Roughness Index (IRI) so as to be consistent with previous research. The transformation formula is QI = 13 × IRI (Paterson, 1990b).

It can be seen in Fig. 4 that the empirical model yields a piecewise linear improvement curve. For a given pavement condition, the rehabilitation effect is linearly related to the intensity, until the intensity reaches a certain threshold. Beyond that point, the additional overlay thickness doesn't bring any significant improvement but causes the same variable cost. In a cost minimization problem, therefore, rehabilitation with intensity beyond the threshold will be sub-optimal. If we were to use the piecewise linear rehabilitation function in the optimization model, a set of additional integer variables would be necessary and add to the computational cost. By restricting

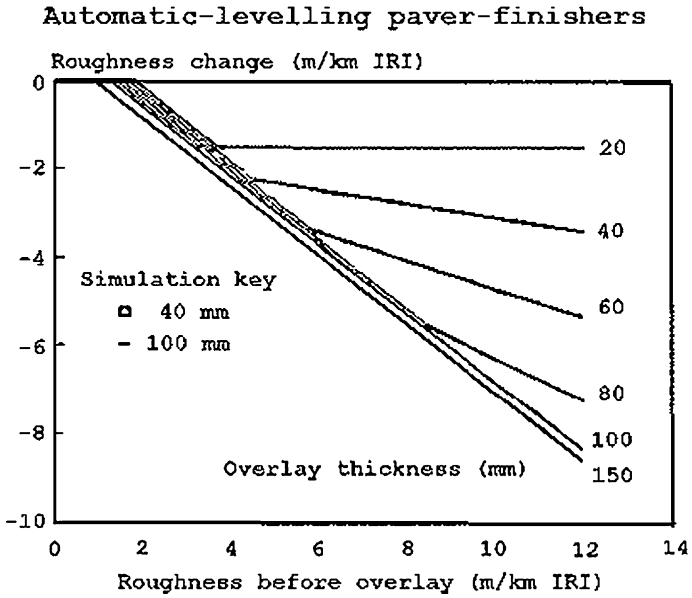

Fig. 3. Predicted effects of asphalt overlays on roughness (Paterson, 1990a).

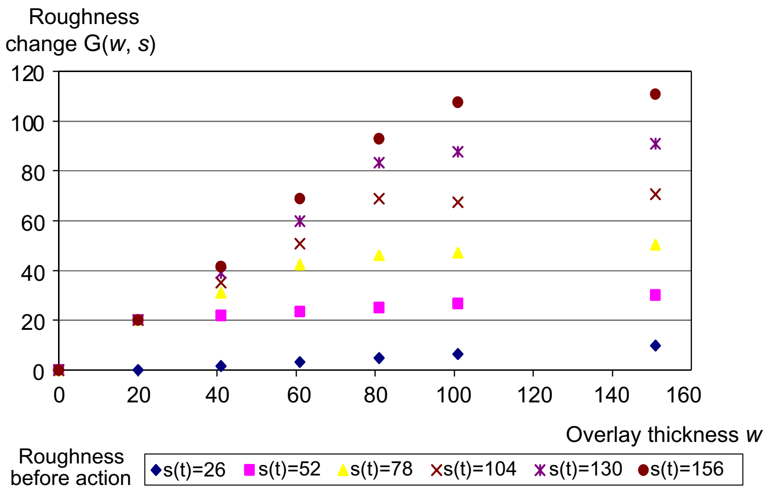

Fig. 4. Rehabilitation effectiveness from Paterson (1990a) in QI units.

the intensity to be less than the "effectiveness threshold", the problem can be simplified without changing the optimal solution.

Noting that both the effectiveness slopes and thresholds are functions of the pavement condition before rehabilitation, we specify the improvement function to be

$$G(s(t^-), w_t) = \frac{w_t}{w_{\max}(s(t^-))} \cdot G_{\max}(s(t^-)) \tag{8}$$

where

$$w_t \leqslant w_{\max}(s(t^-)) \tag{9}$$

Both the threshold value $w_{\max}$ and $G_{\max}$ are functions of the pavement condition $s(t^-)$. Using the data from Fig. 4, regressions of $w_{\max}$ and $G_{\max}$ on $s(t^-)$ provide very good linear relationships as shown in Fig. 5.

Substituting the regression results into the previous formula, we obtain the following empirical rehabilitation effectiveness for any facility in year t:

$$G(w_t, s(t^-)) = \frac{0.66s(t^-)}{0.55s(t^-) + 18.3} \cdot w_t \tag{10}$$

and

$$w_t \leqslant 0.55s(t^-) + 18.3 \tag{11}$$

where $s(t^-)$ is the roughness before action; w_t the overlay thickness in millimeters; G the pavement QI reduction.

With the above upper bound for rehabilitation intensity $w_{\max}$, the logical relationship between the continuous variable w_t and the binary variable X_t can be established. For w_t and X_t to be consistent, if one of them is zero, so should the other. In our formulation, the following constraint is used.

$$w_t \leqslant (0.55s(t^-) + 18.3) \cdot X_t \tag{12}$$

When $X_t = 0$, the above formula reduces to $w_t \leqslant 0$, which requires w_t to be zero since it is non-negative. If $X_t = 1$, the above formula gives the constraint $w_t \leqslant w_{\max}$. Conversely, the above inequality requires $X_t = 1$ when $w_t > 0$. This constraint does not have a restriction on X_t when $w_t = 0$. However, since X_t appears in the objective function with a positive coefficient, such unconstrained variables will always take the value of zero at optimality in a minimization problem.

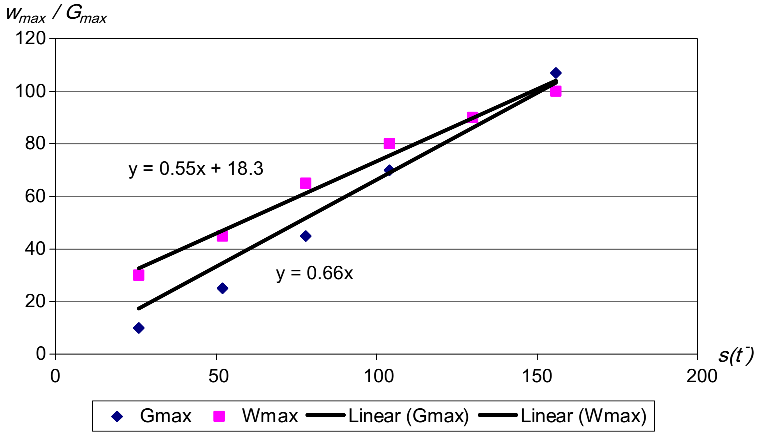

Fig. 5. $w_{\max}$ and $G_{\max}$ as functions of $s(t^-)$.

3.1.3. User and agency cost

User costs consist of two major components: vehicle operating costs and travel delays. Travel delay is a function of traffic demand, user trip purpose and income level. Assuming that travel demand is relatively constant and irresponsive to pavement roughness, the travel delays can be assumed constant. The vehicle operating costs are formulated as a linear function of roughness. Therefore, the user cost rate ($/time) is a linear function of roughness $C(s(t)) = c_1 \cdot s(t) + c_2$, where c_1 and c_2 are constant parameters. The constant c_2 can be omitted in our optimization without affecting the solution. Therefore, during a period of time $[t_1, t_2]$, the discounted total user cost incurred equals the integral of $C(s(t))$ over the period:

$$\int_{t_1}^{t_2} C(s(t))\mathrm{e}^{-rt}\,\mathrm{d}t = c_1 \cdot \int_{t_1}^{t_2} s(t)\mathrm{e}^{-rt}\,\mathrm{d}t \tag{13}$$

The agency cost of pavement rehabilitation consists two parts: fixed cost and variable cost. The fixed cost is the fixed machine rental/operation, labor wages and so on, which are incurred if a rehabilitation action is taken. The variable cost is the cost of overlay materials, etc, which are proportional to the thickness of the overlay. Therefore, we also use a linear function to approximate the agency cost when rehabilitation occurs:

$$M_t(w_t) = m_1 \cdot w_t + m_2 \tag{14}$$

In Eq. (14), m_2 is the fixed cost and m_1 is the variable cost related to the overlay thickness w_t.

3.1.4. Budget constraint and boundary conditions

The constraint that connects the individual facilities into a system is the budget constraint. Without a budget constraint, the agency could make rehabilitation decisions for each facility separately.

The boundary conditions in the model mainly have two parts. First, setting the initial roughness conditions at the beginning of the planning horizon; second, restricting the facility conditions at the end of the planning horizon to be better than a certain level. As mentioned in previous sections, the computation of salvage values has been problematic in the literature. In our model, we remove the salvage value term from the objective function but add a condition constraint as a lower bound on the final facility condition. The agency may decide on a case specific basis what condition constraint to impose. One reasonable option is to require the final condition to be no worse than the initial condition. This is the constraint adopted in our parametric study.

4. Algorithms

The model we formulated is a mixed-integer nonlinear programming (MINLP). In the operations research literature, problems of this type are known to be difficult: Generally applicable commercial solvers have not yet been developed for such problems. Several MINLP algorithms have been developed for such models, including branch-and-bound (B&B), generalized benders decomposition (GBD), outer approximation (OA), etc.

4.1. Branch-and-bound approach

For mixed-integer linear problems (MILP), the branch-and-bound (B&B) approach is well studied and widely available in commercial software. The general branch-and-bound algorithm deals with integer variables through linear programming (LP) relaxation and branching into a search-tree. The B&B algorithm is guaranteed to find the optimal solution, but its complexity in the worst case is as high as that of exhaustive search. In the simplest form of the algorithm, the search-tree is traversed in some order, and the score of the best leaf found so far is kept as a bound B. Whenever a node is reached whose score is worse than B, the tree is pruned at that node, i.e., its sub-tree will not be searched, since it is guaranteed not to contain a leaf with a score better than B.

The ideas of branch-and-bound can be adopted for our MINLP problem. If we perform LP relaxations to the binary integer variables in the model, the MINLP problem becomes an easier nonlinear programming (NLP) problem. If this NLP problem can be solved efficiently by available software, we only need to program for relaxation, branching and bounding control. For this purpose, the AMPL optimization software includes the solver MINOS that proves to be quite efficient for many NLP-type problems. More importantly, the AMPL package provides command language scripts that allow clients to build new algorithms by controlling the solver.

Our model is formulated by using AMPL scripts to control existing solvers. A branch-and-bound control is coded, using MINOS to solve the relaxed NLP sub-problems. In our B&B algorithm coding, the node selection rule is depth-first search plus backtracking, which is also known as the last in, first out strategy. In depth-first search, if the current node is not pruned, the next node considered is one of its two children. Backtracking means that when a node is pruned, we go back on the path from this node toward the root until the first node with an unconsidered child. This method is proven to be relatively economic in memory space and appropriate for personal computers.

Since the combinatorial problem is computationally expensive, many researchers in the operations research literature have proposed ways, such as parallel algorithms and approximate B&B algorithm, to reduce the computational cost of the B&B algorithm. The algorithm can also be improved by using other methods to come up with a relatively good candidate, which will yield a good bound before the search has even begun.

One additional concern for the B&B algorithm in solving a MINLP, compared to a MILP, is that it is based on NLP optimization. Such optimizations are highly sensitive to the model formulation and the values of the parameters. In fact, global convexity is not guaranteed in this model. Therefore, we need to try various starting values during the solution process.

4.2. A heuristic approach

The branch-and-bound approach gives a general solution to our mixed-integer programming problem. However, despite the efficiency-improving techniques that we used, the combinatorial nature of the problem makes the computational costs formidable when the planning horizon and the number of facilities are large, or when the budget constraint is tight. An easy-to-solve, close-to-optimum heuristic is thus appealing for practical applications.

Li and Madanat (2002) showed that for the problem of optimal pavement rehabilitation under steady state, the optimal total cost is insensitive to changes in the decision variables. Analogously, we may expect this conclusion to hold in general. Hence, it may be possible to find a heuristic that provides a close-to-optimum solution at a low computational cost. Depending on the quality of the solutions required, this approximate solution could be either the final answer or an input to our exact B&B algorithm. Using the upper bound provided by the heuristic solution can significantly reduce the computational cost of the branch-and-bound algorithm.

One of the most widely used heuristics is the greedy algorithm. The idea is to start from a simple solution, change relevant variables sequentially, each time selecting the variable that achieves the greatest immediate improvement in the objective function. Moreover, once the value of a variable is changed, it is kept constant in the remainder of the algorithm.

To apply this heuristic to our model, the objective function can be reformulated more directly so that it reveals more clearly the nature of the problem, as shown in Fig. 6. The user cost rate is mapped from the roughness through a linear transformation, as formulated earlier. Rehabilitation actions decrease pavement roughness, hence the "future" user costs (shaded area ΔC), at the expense of "current" investment ΔM. For our deterministic problem, maximizing total user cost reduction ($\sum \Delta C$) minus the total agency cost ($\sum \Delta M$) is equivalent to minimizing the total user cost plus the total agency cost. The difference between these two objectives is a constant for a given initial roughness. The problem is equivalent to maximizing the total shaded areas minus the

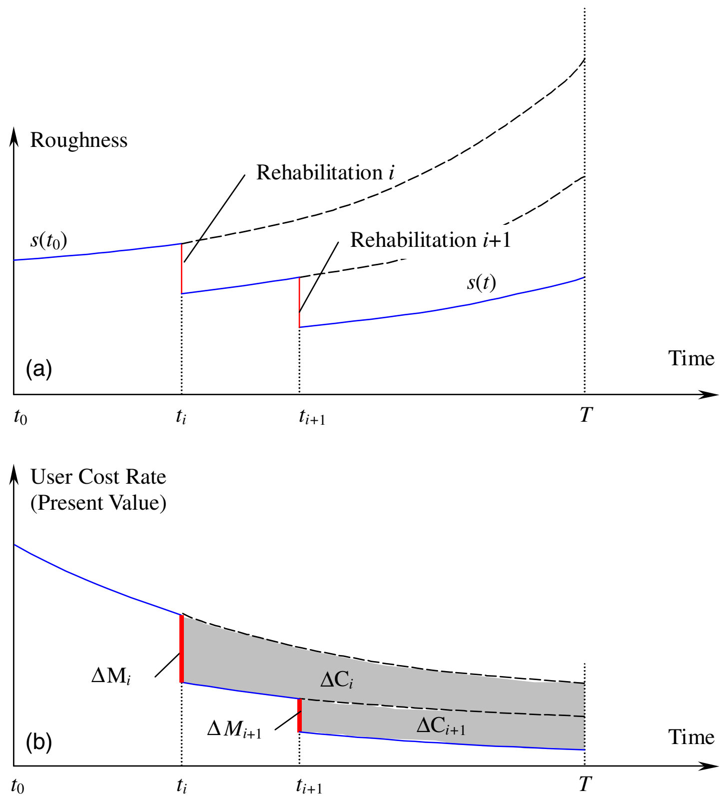

Fig. 6. Reformulated objective function and the effects of rehabilitation. (a) Roughness reduction under rehabilitation activities, (b) discounted user cost rate under rehabilitation activities.

total agency costs. Because all terms are calculated as present values, the total area $\sum \Delta C$ will converge to a finite number even when T goes to infinity, as long as the discounting effect is sufficient. This assumption is generally satisfied in practice as essential for the existence of a steady state.

For a single facility problem, the greedy algorithm is straightforward. Obviously, "doing-nothing" throughout the whole planning horizon is a possible starting point. Then we select rehabilitation activities in a sequential manner, one at a time so as to maximize the possible increase in objective function conditioned on previous actions. That is, we divide the total cost into multiple single-action problems, so that each gives an optimal solution under feasibility conditions. Since the deterioration process is deterministic, for a given $s(0)$, $s(t^-)$ is a monotonic function of the action time t. Let ΔC_i represent the reduced future cost, ΔM_i represent the agency cost for action i, then the decision variable t_i and w_i form the optimal solution to the simpler problem

$$\text{Max} \quad [\Delta C_i(t_i, w_i) - \Delta M_i(t_i, w_i)] \tag{15.1}$$

$$\text{s.t.} \quad t_{i-1} \leqslant t_i \leqslant T \tag{15.2}$$

$$0 \leqslant w_i \leqslant w_{\max} \tag{15.3}$$

$$\Delta M_i(t_i, w_i) \leqslant B(t) \tag{15.4}$$

where $i = 1, 2, \ldots$ is the index of rehabilitation activities that occur between $(0, T)$.

In this partial model, the costs ΔC_i, benefits ΔM_i, and budget $B(t)$ are already discounted into present values. Constraint (15.2) is based on the sequential feature of this heuristic; while constraints (15.3) and (15.4) are analogous to constraints (1.4) and (1.5). Note that if $\Delta C_i(t_i, w_i) \geqslant \Delta M_i(t_i, w_i)$ then the activity adopted at time t_i will improve the objective function. This one-action problem has been thoroughly studied in the literature (Friesz and Fernandez, 1979; Markow and Balta, 1985).

With this heuristic, we solve a series of one-action problems with two bounded decision variables. For example, starting from t_0 (beginning of planning horizon), we find the first action time $t_1 > t_0$ and intensity $w(t_1)$ that improve the objective function the most. Then, taking $s(t_1^+)$ as the starting roughness, we find $t_2 > t_1$ and $w(t_2)$ that increase the objective function the most. We repeat the above steps until the end of the planning horizon is reached and no additional action can increase the objective function. To maintain feasibility, however, we need to check the boundary condition (1.6) in the end. If it is not satisfied, additional rehabilitation actions with minimal additional agency costs are necessary although they decrease the objective function.

The above single-facility heuristic can be easily extended to the multiple-facility case, in which decisions are made in three dimensions: action time, intensity and the facility to be rehabilitated. Action decisions are made sequentially under all constraints. Starting from the beginning of the planning horizon, the best combination of (time, intensity, facility) is chosen. Once a facility is selected for an optimal action at a certain time, the budget constraint of that year and this facility's condition for the remainder of the planning horizon are updated. Repeating these steps in a sequential manner, actions are selected until any additional action decreases the objective function and boundary condition (1.6) is satisfied for every facility. The sequence of actions selected above is the heuristic solution.

The above heuristic approach can also be coded into AMPL scripts. The steps to apply this heuristic to a system of N facilities are given below:

1. Initialization: $i = 1$; facilities $n = 1, 2, \ldots, N$.
2. Solve the partial model (15.1)–(15.4) for every facility in the system. Find $(t_i^*, w_i^*, n_i^*) = \arg\max_{(t_i, w_i, n)} \{[\Delta C_i(t_i, w_i) - \Delta M_i(t_i, w_i)]_n\}$, $n = 1, 2, \ldots, N$.
3. If $[\Delta C_i(t_i^*, w_i^*) - \Delta M_i(t_i^*, w_i^*)]_{n_i^*} \geqslant 0$, go to step 5.
4. Calculate roughness at the end of planning horizon, $s_n(T)$, $n = 1, 2, \ldots, N$; if $s_n(T) \leqslant s_{Tn}$ for $\forall\, n$, go to step 6.
5. Let $(t_i^*, w_i^*, n_i^*) = \arg\max_{(t_i, w_i, n), s_n(T) \geqslant s_{Tn}} \{[\Delta C_i(t_i, w_i) - \Delta M_i(t_i, w_i)]_n\}$.
6. Record action (t_i^*, w_i^*, n_i^*), update the budget available in year t_i^* and the planning horizon for facility n_i^* which now starts from t_i^*; $i = i + 1$; go to step 2.
7. Output the recorded actions, (t_i^*, w_i^*, n_i^*), $i = 1, 2, \ldots$; terminate the algorithm.

As we will numerically show later, this algorithm produces close-to-optimal solutions. In fact, when the planning horizon is short such that the optimal solution has at most one action, the greedy algorithm is exact. At the other extreme, the reduced user cost ΔC_i, evaluated at the action time t_i, will converge to a finite value for every i in the steady state ($T \to \infty$). The steady-state solution consists of similar rehabilitation actions repeated with constant intervals and intensities. For planning horizons of intermediate lengths, the greedy algorithm is sub-optimal. Even in this case, the greedy algorithm gives good results due to the effect of discounting.

Generally, the computational costs of the branch-and-bound approach increase exponentially with the scale of the problem (number of facilities and length of planning horizon). On the other hand, the heuristic approach which solves the problem in a sequential manner has only pseudo-polynomial computational cost. This reduction in computational complexity is very attractive for large-scale problems.

5. Parametric analyses

In the present section, we will apply our MINLP model to solve for the optimal rehabilitation plan for both a system of facilities and a single facility. There are several reasons for doing so. First, we want to compare the solutions from the two algorithms (the B&B with NLP relaxation and the heuristic) and, therefore, test the applicability of the newly developed heuristic. Second, we want to perform parametric analyses to obtain approximate relationships between life-cycle cost and various influencing factors.

5.1. System of facilities: algorithm comparison and influence of budget

For illustration purposes, we first consider the rehabilitation scheduling for a system of three pavement facilities. In our model we measure roughness in QI units, time in years, overlay thickness in millimeters, and cost in thousand dollars. Based on the aforementioned empirical formulae (Paterson, 1990a, etc.), we choose the following parameters:

Table 1
Parameters of a system of three facilities

	β	$\bar{f}^*$	$s(0)$	c_1	m_1	m_2	r	T	B
Facility 1	0.0153	2.0	40	1.2	3.0	170	0.07	60	900
Facility 2	0.0153	1.5	50	1.0	2.5	150	0.07	60	900
Facility 3	0.0153	1.6	60	1.1	2.5	150	0.07	60	900

In Table 1, β is a constant obtained from Paterson's empirical study. Therefore, the parameter $\bar{f}^*$ alone determines the deterioration rate. The rate is dependent on various factors such as pavement design, traffic loading and environmental conditions. In our model, the order of their values are taken so that a pavement with a roughness of 25 QI will deteriorate to a roughness of 100 QI in approximately 30 years if no rehabilitation is conducted. This deterioration rate is reasonable. The discount rate is taken to be 7% over the planning horizon. The budget constraint B could be time-dependent, but for simplicity we set it to a constant of \$900k each year. The length of planning horizon is taken to be 60 years, which is neither an infinite horizon case nor a single-rehabilitation case. Therefore, this case is in some ways the "worst case" for our heuristic, and thus worthy of study.

Since the budget constraint is the only constraint that connects the individual facilities into a system level problem, we are interested in its influence on the total life-cycle cost. Actually, if the budget constraint is nonbinding, the system of facilities can be decomposed into multiple single facility problems, each of which can be solved separately.

The optimization model for this system of three facilities includes 540 variables and 420 constraints after being preprocessed by AMPL's pre-solver. The computational time of the exact branch and bound approach varies with the value of budget B. Although slightly dependent on the choice of starting values, the average computational time on the Intel Pentium IV-1.7 GHz CPU ranges from about 64 s ($B = 1000$k) to about 110 s ($B = 300$k). This is quite large compared to the heuristic approach whose computational time is normally less than 1 second for these problems.

Fig. 7 shows the roughness progressions under the optimal rehabilitation strategy in the system with sufficient budget. We observe that each pavement segment, no matter what the initial roughness is, enters a close-to-steady state after the first few rehabilitation actions. Also, the rehabilitation intervals are almost constant in the following years, except for the last few years where the boundary condition requirements take effect.

Fig. 8 gives the relationship between the optimal total cost (user cost + agency cost) and the annual budget constraint.

In Fig. 8, the lower curve shows the minimal cost obtained by using the B&B algorithm, while the upper curve shows its counterpart obtained by using the heuristic. We observe that the two curves are very close to each other. The largest difference is less than 3% and it approaches zero very fast when the budget decreases. Noting the constant part of the user costs that is omitted in the objective function, the actual difference is even smaller. We also see that the cost curves remain almost flat when the budget constraint is above \$300k, which is approximately the threshold where exactly one facility can be optimally rehabilitated each year. For a budget constraint above \$300k, the effect on the value of the total objective function is very small. The only small improvements occur at budgets around \$550k and \$870k, which correspond respectively to the

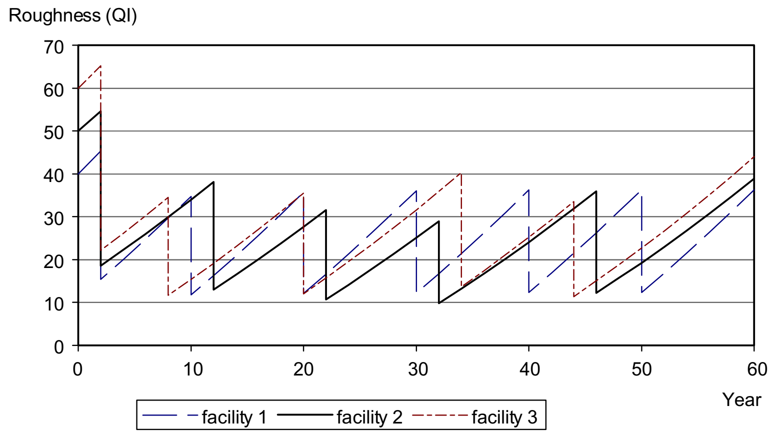

Fig. 7. Roughness progressions under the optimal rehabilitation plan ($B = 900$).

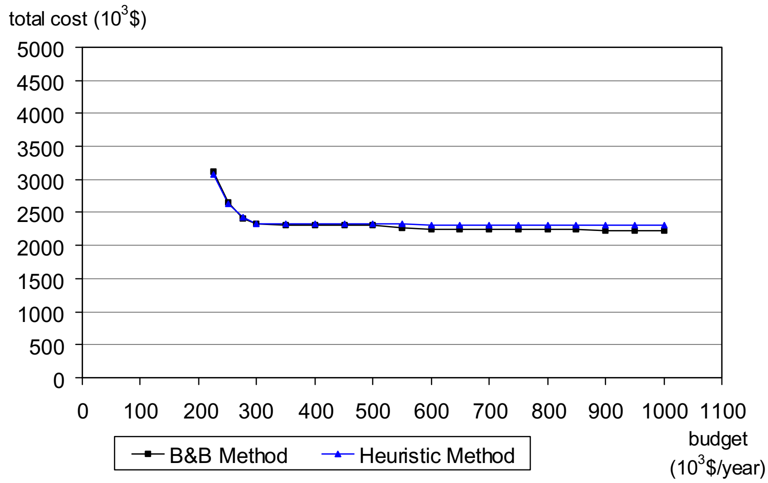

Fig. 8. Effect of budget constraint on total system costs.

budget thresholds that allow two and three facilities to be rehabilitated simultaneously. The budget constraint over \$870k is entirely non-binding. When the budget decreases below \$300k, the minimal cost increases very rapidly until the problem becomes infeasible. This dramatic increase in total cost comes from the fact that even a single facility cannot be optimally rehabilitated in a given year with a budget below \$300k. The problem becomes infeasible when no effective rehabilitation can be carried out and the boundary constraint on the roughness at the end of the horizon cannot be satisfied.

5.2. Single facility: deterioration and cost dominance

We have discussed the influence of the budget constraint on a system of three facilities. In the model there are other parameters for deterioration and cost that are determined empirically.

Below, we carry out a sensitivity analysis to understand how such parameters influence the optimal solution. To exclude all irrelevant factors, we dropped the budget constraint and select "Facility 2" to study the influence of the deterioration and cost parameters.

5.2.1. Deterioration parameters

Intuitively the deterioration rate of a pavement has an effect on the minimal life-cycle cost and the optimal rehabilitation frequency. The faster the pavement deteriorates, the higher the user costs, thus increasing the optimal rehabilitation frequency.

In our deterioration model, β is fixed and $\bar{f}$ determines the deterioration curve. A higher $\bar{f}^*$ value represents faster deterioration. Figs. 9 and 10 show, respectively, the total life-cycle cost and rehabilitation frequency (in years) when $\bar{f}^*$ takes values from 1.2 to 9. The time it takes a pavement of QI = 25 to deteriorate to QI = 100 is about 35 years when $\bar{f}^* = 1.2$, and 7.5 years when $\bar{f}^* = 9$.

In Fig. 9 we observe that the total life-cycle cost increases almost linearly with $\bar{f}^*$. But the slope shows that the effect of deterioration rate on total cost is small. Note that when $\bar{f}^*$ changes from

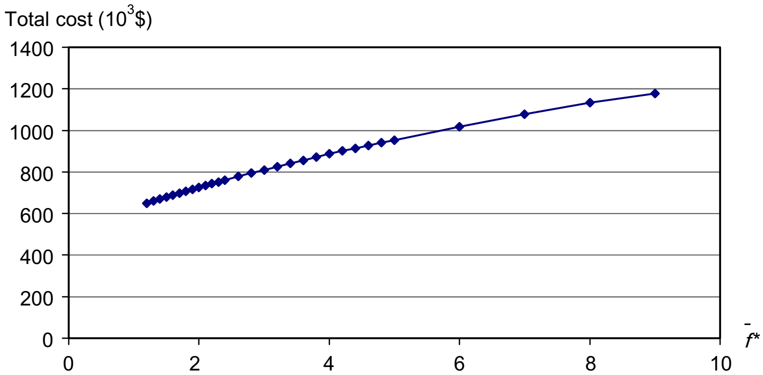

Fig. 9. Effect of the deterioration rate on life-cycle cost.

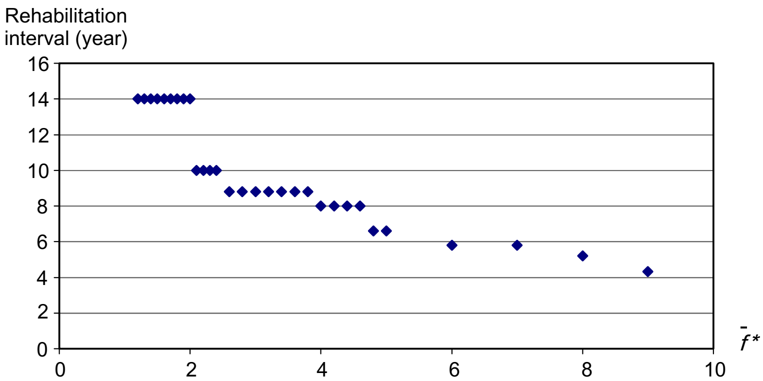

Fig. 10. Effect of the deterioration rate on the rehabilitation interval.

1.2 to 9.0, the deterioration rate is five times as large, but the total life-cycle cost increases by a factor of < 2. Considering the constant part of the user cost that is omitted in our objective function, this result shows that the total life-cycle cost is not very sensitive to the deterioration rate. Fig. 10 shows the optimal rehabilitation frequency as a function of $\bar{f}^*$. We can see that the frequency of rehabilitation actions is significantly influenced by the deterioration rate.

5.2.2. Dominance of user costs

Another issue of interest is the effect of the relative importance of agency and user costs on the optimal solution. In our formulation, the total life-cycle cost function is a homogenous function of c_1, m_1 and m_2 of degree one. The optimal strategy is influenced by the relative importance of the user and agency costs rather than the absolute value of each. For simplification, we hold the agency cost parameters m_1 and m_2 constant and observe the influence of the user cost parameter c_1.

The increase in the importance of user cost should have a similar effect as the increase in the rate of deterioration. Therefore, we expect to see similar curves as those of Figs. 9 and 10. The results are shown in Figs. 11 and 12.

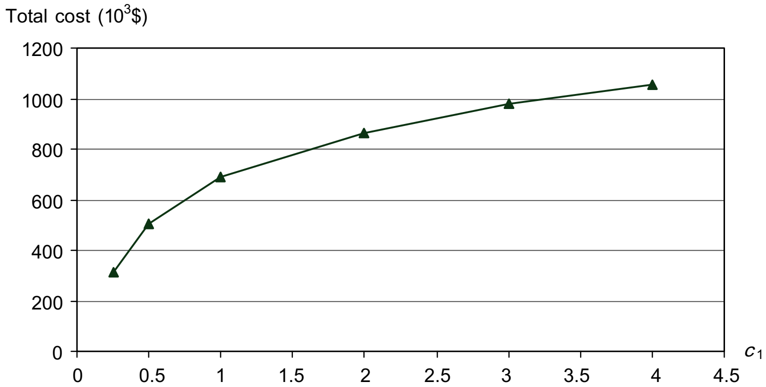

Fig. 11. Effect of the user cost parameter c_1 on life-cycle cost.

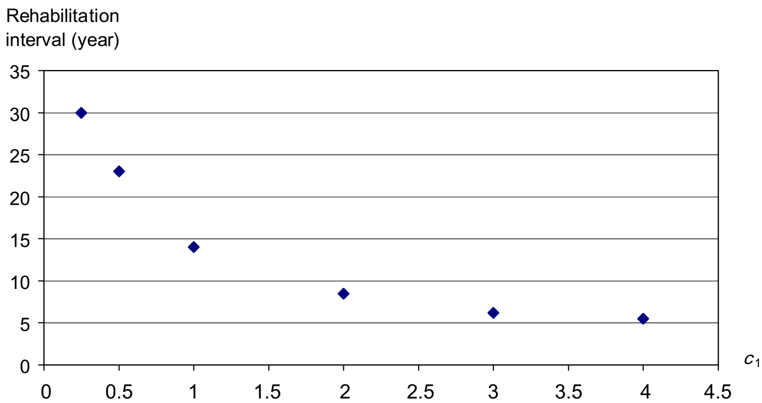

Fig. 12. Effect of the user cost parameter c_1 on rehabilitation interval.

In Figs. 11 and 12, the parameter c_1 varies over the range 0.25–4.0. The benchmark is $c_1 = 1.0$. We observe that the total cost increases with c_1 but at a decreasing speed. Mean while, the rehabilitation interval decreases (Fig. 12). This implies that when the user cost parameter increases, it is optimal to undertake more frequent rehabilitation actions to compensate for the increase in user cost.

6. Conclusions

In this paper, we have presented a mixed-integer nonlinear programming (MINLP) model to schedule multiple rehabilitation actions in a system of pavement facilities under budget constraints. The objective function of this model is to minimize in a finite horizon the discounted total life-cycle cost, including both user cost and agency cost. The facility condition at the end of planning horizon is constrained by a minimum serviceability requirement. This model is based on the theory of optimal control for discrete multi-stage systems. Both deterioration and rehabilitation effectiveness models are nonlinear and the fixed cost of rehabilitation introduces integer decision variables. The deterioration and rehabilitation effectiveness models are obtained from previous empirical studies.

MINLP problems are known in the operations research literature as some of the most difficult. A branch and bound algorithm is used in conjunction with a nonlinear programming solver to come up with a solution. The influence of various factors such as the budget constraint, deterioration rate, and user/agency costs were investigated through parametric analyses. One drawback of the algorithm is that the MINLP problem belongs to the category of NP problems, and therefore incurs prohibitive computational cost when the problem scale increases.

To circumvent this problem, a heuristic approach was proposed and tested. Numerical analyses showed that the heuristic gives close-to-optimal solutions. The heuristic approach has pseudo-polynomial computational cost, therefore is very attractive for large-scale problems in practice.

Acknowledgements

Partial funding for this research was provided by a research grant from the University of California Transportation Center.

References

Al-Subhi, K.M., Johnston, D.W., Farid, F., 1990. A resource constrained capital budgeting model for bridge maintenance, rehabilitation and replacement. Transportation Research Record 1268, 110–117.

Bryson, A.E., Ho, Y., 1975. Applied Optimal Control: Optimization, Estimation and Control. Hemisphere, New York, NY.

Fernandez, J.E., Friesz, T.L., 1981. Influence of demand quality interrelationships in optimal policies for stage construction of transportation facilities. Transportation Science 15, 16–31.

Friesz, T.L., Fernandez, J.E., 1979. A model of optimal transport maintenance with demand responsiveness. Transportation Research B 13, 317–339.

Jacobs, T.L., 1992. Optimal long-term scheduling of bridge deck replacement and rehabilitation. Journal of Transportation Engineering, ASCE 118 (2), 312–322.
Li, Y., Madanat, S., 2002. A steady-state solution for the optimal pavement resurfacing problem. Transportation Research Part A 36, 525–535.
Markow, M., Balta, W., 1985. Optimal rehabilitation frequencies for highway pavements. Transportation Research Record 1035, 31–43.
Murakami, K., Turnquist, M.A., 1985. A dynamic model for scheduling maintenance of transportation facilities, Presented at Transportation Research Board 64th Annual Meeting, Washington, DC.
Paterson, W.D.O., 1987. Road Deterioration and Maintenance Effects: Models for Planning and Management. Johns Hopkins University Press, Baltimore, MD.
Paterson, W.D.O., 1990a. Quantifying the effectiveness of pavement maintenance and rehabilitation. In: Proceedings at the 6th REAAA Conference, Kuala Lumpur, Malaysia.
Paterson, W.D.O., 1990b. International roughness index: relationship to other measures of roughness and riding quality. Transportation Research Record 1084, 49–59.
Tsunokawa, K., Schofer, J.L., 1994. Trend curve optimal control model for highway pavement maintenance: case study and evaluation. Transportation Research A 28, 151–166.