Available online at www.sciencedirect.com
SCIENCE DIRECT®
Transportation Research Part B 40 (2006) 767–778
TRANSPORTATION RESEARCH PART B
www.elsevier.com/locate/trb

An analytical solution for the finite-horizon pavement resurfacing planning problem

Yanfeng Ouyang [a,*], Samer Madanat [b]

[a] *Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 1210 Newmark Civil Engineering Laboratory, Urbana, IL 61801, USA*
[b] *Institute of Transportation Studies and Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720, USA*

Received 19 November 2004; received in revised form 2 November 2005; accepted 3 November 2005

Abstract

This paper presents an analytical approach for planning highway pavement resurfacing activities in the case of continuous pavement state and continuous time. It solves for the optimal resurfacing frequency and intensity that minimize life-cycle costs in a finite horizon. Optimality conditions are derived analytically, and a simple algorithm is developed to solve for the exact optimal solution. The optimal resurfacing strategy is found to be consistent with findings previously obtained by using an approximate approach and the strategy obtained for the infinite-horizon problem.

Keywords: Pavement resurfacing; Optimization; Life-cycle costs; Calculus of variations; Threshold structure; Policies

1. Introduction

This paper addresses the problem of planning pavement resurfacing activities in continuous time and with a continuous pavement state. The goal is to minimize the life-cycle costs including agency and user cost over a finite planning horizon.

In earlier research, control theory was used to model continuous maintenance activities (Friesz and Fernandez, 1979; Fernandez and Friesz, 1981), under which pavement condition changes smoothly over time. Pavement rehabilitation activities (e.g., resurfacing), on the other hand, lead to discrete jumps in pavement condition. Markow and Balta (1985) studied the problem for a single rehabilitation activity over a finite horizon, but their solution procedure is not practical for multiple rehabilitations. Tsunokawa and Schofer (1994) introduced an approximate approach to multiple resurfacing activities in a finite horizon, where the problem of determining discrete resurfacing activities is replaced by one of solving for a continuous resurfacing rate. Their results, although approximate, suggest that pavement condition reaches a steady state after a

* Corresponding author. Tel.: +1 217 333 9858; fax: +1 217 333 1924.
E-mail addresses: yfouyang@uiuc.edu (Y. Ouyang), madanat@ce.berkeley.edu (S. Madanat).

0191-2615/$ - see front matter
doi:10.1016/j.trb.2005.11.001

few resurfacings. This is consistent with the findings by Li and Madanat (2002), for an infinite-horizon steady-state resurfacing problem. However, to the best knowledge of the authors, exact analytical results are not yet available for optimizing multiple rehabilitations over a finite horizon.

Discrete-time rehabilitation planning problems are often solved with Markov Decision Process (MDP) formulations that were originally proposed in Bellman (1955), Dreyfus (1960) and Klein (1962). Some analytical properties of MDP models (such as the threshold structure of optimal policies) are known, but most studies on pavement rehabilitation remain numerical (Golabi et al., 1982; Carnahan et al., 1987). Other researchers have solved finite-horizon rehabilitation problems numerically for a system of pavement facilities. They approached the problems by formulating mixed-integer mathematical programs in discrete time (Al-Subhi et al., 1990; Jacobs, 1992). These earlier works used unrealistic pavement performance models, such as linear deterioration curves, to avoid difficulties in the optimization. Later, Ouyang and Madanat (2004) used realistic performance models (e.g., Paterson, 1990) to solve a similar problem. They formulated and solved the problem as a mixed-integer nonlinear program. They also proposed a simple greedy heuristic that was shown to provide near-optimum solutions.

This paper focuses on the problem of optimizing multiple resurfacing activities in a finite horizon, with realistic pavement performance models very similar to those in the literature (Tsunokawa and Schofer, 1994; Li and Madanat, 2002; Ouyang and Madanat, 2004). The work aims to investigate the analytical side further with a method inspired by the calculus of variations. This method finds simple analytical optimality conditions that are then used to design algorithms for an exact optimal solution. The optimal resurfacing strategy found in this paper, which is exact, is consistent with earlier research findings under different settings (Tsunokawa and Schofer, 1994; Li and Madanat, 2002).

The paper is organized as follows: Section 2 introduces the problem formulation and pavement performance models. Section 3 derives the optimality conditions for this problem and proposes a simple solution procedure based on these conditions. This is followed by a numerical example in Section 4 and some conclusions in Section 5.

2. A resurfacing problem

2.1. General formulation

The net present value of life-cycle costs consists of agency investments and user costs over a time horizon of $[0, T]$. The highway agency selects the times, $\mathbf{t} = (t_1, t_2, \ldots, t_n)$, and intensities, $\mathbf{w} = (w_1, w_2, \ldots, w_n)$, of n sequential resurfacing activities, where n itself is a decision variable. Let $t_0 = 0$, and $t_{n+1} = T$, the problem can be formulated as

$$\min \quad J(\mathbf{t}, \mathbf{w}) = \sum_{i=0}^{n} \left\{ \int_{t_i}^{t_{i+1}} C(s(u)) \mathrm{e}^{-ru} \mathrm{d}u \right\} + \sum_{i=1}^{n} M(w_i) \mathrm{e}^{-rt_i}, \tag{1a}$$

$$\text{s.t.} \quad s(t_i^-) - s(t_i^+) = G(w_i, s(t_i^-)), \quad i = 1, 2, \ldots, n, \tag{1b}$$

$$s(t) = F(s(t_i^+), t - t_i) \quad \forall t \in (t_i, t_{i+1}), \quad i = 0, 1, 2, \ldots, n, \tag{1c}$$

$$0 \leqslant w_i \leqslant R_i, \quad i = 1, 2, \ldots, n, \tag{1d}$$

$$t_i - t_{i+1} + \tau \leqslant 0, \quad i = 1, 2, \ldots, n - 1, \tag{1e}$$

$$0 \leqslant t_i \leqslant T, \quad i = 1, 2, \ldots, n, \tag{1f}$$

$$s(0) = s_0, \quad s(T) \leqslant s_T. \tag{1g}$$

In the objective function (1a), $s(t)$ is a continuous pavement condition indicator (i.e., roughness) at time t, where a larger value of $s(t)$ indicates worse condition; $C(s)$ is the user cost per unit time as a function of roughness s; $M(w)$ is the agency cost for a resurfacing of intensity w (i.e., thickness of overlay); and the parameter r is the discount rate. Constraint (1b) defines the effectiveness of resurfacing in reducing roughness. The amount of reduction $(\Delta s)_i \equiv s(t_i^-) - s(t_i^+)$ depends on w_i and $s(t_i^-)$, where $s(t_i^-)$ and $s(t_i^+)$ denote the roughness before and after resurfacing i. Constraint (1c) describes how pavement deteriorates over time between two consecutive resurfacing activities. Constraint (1d) introduces a realistic feasibility bound $[0, R_i]$ for resurfacing intensity. Constraint (1e) simply requires any two consecutive resurfacing times to be separated by at least an interval τ, which may represent labor shift or equipment maintenance cycles. Finally, (1f) requires all resurfacing activities to be conducted within the planning horizon and (1g) gives roughness condition requirements at both ends of the planning horizon. Instead of incorporating a savage value at T, which is not well defined for highway pavements, we require $s(T)$ to be no worse than a certain condition s_T.

2.2. Performance models

In this section, we discuss how we specify the pavement performance models and cost functions, $C(s)$, $M(w)$, $F(s,t)$, and $G(w,s)$.

User costs consist of two major components: vehicle operating costs and delay costs. Delay cost is a function of traffic demand, trip purpose, traveler income, etc. Since these are relatively irresponsive to pavement roughness, the delay costs are largely constant. If the vehicle operating costs are approximately proportional to roughness, the user cost per unit time, $C(s)$, is a linear function:

$$C(s(t)) = c_1 \cdot s(t) + c_2 \quad \forall t, \tag{2}$$

where $c_1 \geqslant 0$, $c_2 \geqslant 0$ depend on traffic demand and other characteristics of the facility. We assume they are constant over time.

The agency costs for pavement resurfacing consist of a fixed set-up component (e.g., for machine rental, operation, labor, and possible traffic delay due to the resurfacing), and a variable component that is proportional to the intensity (e.g., material expenditures). Therefore, the un-discounted agency cost for a resurfacing of intensity (thickness) w_i may be represented by a linear function:

$$M(w_i) = m_1 \cdot w_i + m_2 \quad \forall i, \tag{3}$$

where $m_1 \geqslant 0$, $m_2 \geqslant 0$ are facility-specific parameters. Because we are planning resurfacing activities for a single facility, the budget constraint is not considered.

Based on the empirical data reported in Paterson (1990), the following formula is proposed in Ouyang and Madanat (2004) for resurfacing effectiveness:

$$G(w_i, s(t_i^-)) = g_1 s(t_i^-) \cdot \min\left\{1, \frac{w_i}{g_2 s(t_i^-) + g_3}\right\} \quad \forall i, \tag{4}$$

where $g_1 = 0.66$, $g_2 = 0.55$, $g_3 = 18.3$. In (4), $G(w,s)$ and s are in Quarter-Car Index (QI) and w_i in millimeter (mm). Due to the truncation, any resurfacing thickness $w_i > g_2 s(t_i^-) + g_3$ is suboptimal. Thus we rewrite (4) into:

$$G(w_i, s(t_i^-)) = g_1 s(t_i^-) \cdot \frac{w_i}{g_2 s(t_i^-) + g_3} = \frac{g_1 w_i}{g_2 + g_3/s(t_i^-)} \quad \forall i, \tag{5}$$

and

$$w_i \leqslant R_i \equiv g_2 s(t_i^-) + g_3. \tag{6}$$

In the rest of the paper, we will keep (6) as a separate constraint and use $G(w,s)$ to refer to the expression in (5).

Between two consecutive resurfacings, pavement roughness increases continuously. The deterioration process is modeled as an exponential function of time (see Tsunokawa and Schofer, 1994; Li and Madanat, 2002; Ouyang and Madanat, 2004), such that for $t > t_i\ \forall i$,

$$F(s(t_i^+), t - t_i) = s(t_i^+) \cdot e^{b(t-t_i)}, \tag{7}$$

where the parameter $b > 0$ (time^{-1}) depends on traffic demand and facility characteristics.

Finally, to keep the mathematical derivations relatively simple, we assume $\tau = 0$. This does not compromise the generality of the findings but simplifies the notation.

3. Solution procedure

3.1. Optimality conditions

Given a resurfacing plan, the development of pavement roughness over time can be plotted into a roughness trajectory. Such trajectories satisfy (1b) and (1c) and have a saw-tooth shape with alternating downward jumps and oblique deteriorating segments. Fig. 1 gives an illustration. The minimization of $J(\mathbf{t},\mathbf{w})$ is a multi-stage dynamic program that searches for optimal roughness trajectories.

It is hard to find the analytical form of the entire optimal trajectory directly. However, some ideas from the calculus of variations may help reduce the difficulty. Given any feasible trajectory, we can introduce infinitesimal perturbations to certain decision variables so as to construct alternative feasible trajectories. Then we can investigate the cost differences between the original trajectory and the alternative, and derive necessary conditions for the original trajectory to be optimal. As we will show, if these perturbations are carefully selected, the cost differences and the resulting optimality conditions may both be "local" (i.e., affecting only a small portion of the trajectory). These local conditions, involving only a few state and decision variables, may help design a simple algorithm that solves for an entire optimal trajectory.

Suppose curve [S–A–B–C–D] in Fig. 1 is part of an optimal trajectory. An alternative trajectory can be constructed by perturbing t_i to $(t_i+\varepsilon_i)$ and w_i to $(w_i+\varepsilon_i')$ simultaneously, while holding all other decision variables fixed. Here, subscript $i=1, 2,\ldots,n$ denotes a generic resurfacing activity, and ε_i, ε_i' are small quantities. Appendix A shows that if $\varepsilon_i' = g_2 b(g_2+g_3/s(t_i^-))^{-1} w_i\varepsilon_i$, the alternative trajectory will go through [S–A′–B′–C–D] instead but the rest of the two trajectories overlap. In so doing, the first-order derivative of J with respect to ε_i at $\varepsilon_i=0$ is

$$\left.\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i}\right|_{\varepsilon_i=0} = (C(s(t_i^-)) - C(s(t_i^+)))\mathrm{e}^{-rt_i} + M_w'(w_i)\frac{g_2 b w_i}{g_2+g_3/s(t_i^-)}\mathrm{e}^{-rt_i} - rM(w_i)\mathrm{e}^{-rt_i}. \tag{8}$$

To construct this alternative trajectory, t_i and w_i are perturbed in proportion. Constraint (1e) may restrict the construction of such alternative trajectories; e.g., t_i cannot increase by itself if $t_i = t_{i+1}$. The intensity restriction in (6), on the other hand, is shown to be non-binding for the perturbations described above; this is shown in Appendix A. Thus, a necessary condition for the original curve [S–A–B–C–D] to be optimal is

$$\left.\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i}\right|_{\varepsilon_i=0} \begin{cases} \geqslant 0, & \text{if } t_{i-1}=t_i \\ =0, & \text{if } t_{i-1}<t_i<t_{i+1} \\ \leqslant 0, & \text{if } t_i = t_{i+1}. \end{cases} \tag{9}$$

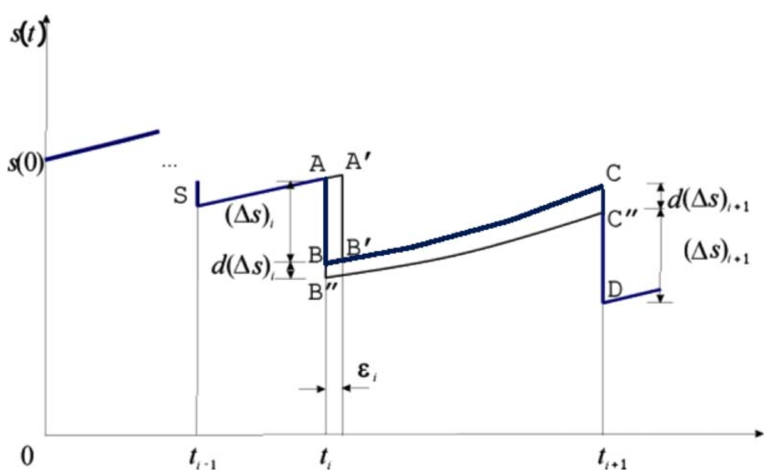


Fig. 1. Alternative roughness trajectories.

From now on, we will simply use $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i}$ to represent its value at $\varepsilon_i = 0$, $\left.\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i}\right|_{\varepsilon_i=0}$. Substituting (1b), (2), (3) and (5) into (8), it yields the following:

$$\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i} = \mathrm{e}^{-rt_i}\left(\frac{c_1g_1 + m_1g_2bw_i}{g_2 + g_3/s(t_i^-)} - rm_1w_i - rm_2\right). \tag{10}$$

Given t_i and $s(t_i^-)$, $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i}$ is a linear function of w_i, as plotted in Fig. 2(a). Note that w_i must also satisfy:

$$0 \leqslant w_i \leqslant R_i. \tag{11}$$

We have the following proposition.

Proposition 1. *On an optimal trajectory, if* $t_i < T$, *then* $\frac{c_1g_1+m_1g_2b}{g_2+g_3/s(t_i^-)} > rm_1\ \forall i$.

Proof. Suppose $\frac{c_1g_1+m_1g_2b}{g_2+g_3/s(t_i^-)} \leqslant rm_1$, then $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i} < 0$ regardless of $w_i \in [0, R_i]$. According to (9), $t_i = t_{i+1}$ on an optimal trajectory. The next resurfacing would start with a roughness of $s(t_{i+1}^-) = s(t_i^+) \leqslant s(t_i^-)$, implying that $\frac{c_1g_1+m_1g_2b}{g_2+g_3/s(t_{i+1}^-)} \leqslant rm_1$. Thus $t_i = t_{i+1} = t_{i+2}$. Repeating this we will get $t_i = t_{i+1} = \cdots = t_n$ and $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_n} < 0$. If $t_i = t_n < T$, perturbing t_n to the right will still reduce J. Therefore, this trajectory cannot be optimal. □

Let $w_i^*(s(t_i^-))$ denote the unique "optimal" resurfacing intensity. Setting $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i} = 0$, we obtain the following:

$$w_i^*(s(t_i^-)) = rm_2\left(\frac{c_1g_1 + m_1g_2b}{g_2 + g_3/s(t_i^-)} - rm_1\right)^{-1}. \tag{12}$$

Given t_i, therefore, $w_i^*(s(t_i^-))$ is positive and decreases with $s(t_i^-)$ from $+\infty$, as illustrated in Fig. 2(b).

Similarly, we can perturb w_i and w_{i+1} simultaneously to construct another trajectory; see for example [S–A–B″–C″–D] in Fig. 1. (When $i = n$, this local variation is achieved by varying w_i only, as will be seen later.) When $i < n$, if w_i varies by a small amount η_i, the roughness after resurfacing $s(t_i^+)$ will vary by $\mathrm{d}(\Delta s)_i$ and reach point B″. After deterioration, $s(t_{i+1}^-)$ will reach C″, deviating from C by $\mathrm{d}(\Delta s)_{i+1}$. Therefore, w_{i+1} should vary by η_i' for the rest of the trajectory to remain unchanged. Note that the roughness $\forall u \in (t_i^+, t_{i+1}^-)$ also changes by $\mathrm{d}s(u)$. According to (1b) and (1c), η_i, η_i', $\mathrm{d}(\Delta s)_i$, $\mathrm{d}(\Delta s)_{i+1}$ and $\mathrm{d}s(u)$ should satisfy the following relationships:

$$\frac{\mathrm{d}(\Delta s)_i}{\eta_i} = G_w'(w_i, s(t_i^-)),$$
$$\frac{\mathrm{d}(\Delta s)_{i+1}}{\mathrm{d}(\Delta s)_i} = -F_s'(s(t_i^+), t_{i+1} - t_i),$$
$$\frac{\mathrm{d}(\Delta s)_{i+1}}{\eta_i'} = G_w'(w_{i+1}, s(t_{i+1}^-)),$$
$$\frac{\mathrm{d}s(u)}{\mathrm{d}(\Delta s)_i} = F_s'(s(t_i^+), u - t_i) \quad \forall u \in (t_i^+, t_{i+1}^-).$$

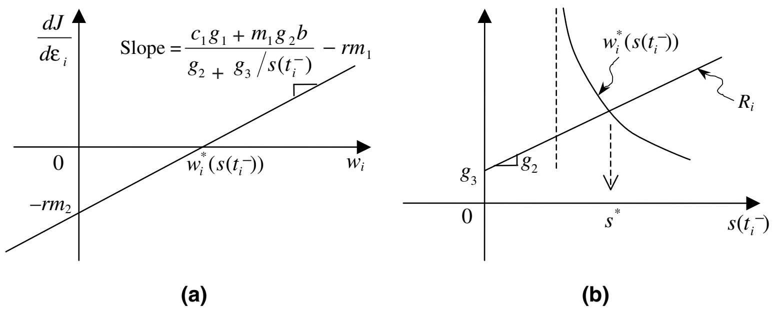


Fig. 2. (a) $\mathrm{d}J/\mathrm{d}\varepsilon_i$ as a linear function of w_i; (b) "optimal" resurfacing intensity $w_i^*(s(t_i^-))$ and maximum effective intensity R_i as functions of $s(t_i^-)$.

Thus, in constructing this alternative trajectory, $\eta_i' = F_s'(s(t_i^+), t_{i+1} - t_i)\frac{G_w'(w_i, s(t_i^-))}{G_w'(w_{i+1}, s(t_{i+1}^-))}\eta_i$, and $\mathrm{d}s(u) = F_s'(s(t_i^+), u - t_i)G_w'(w_i, s(t_i^-))\eta_i\ \forall u \in (t_i^+, t_{i+1}^-)$.

Then for $i = 1, 2, \ldots, n-1$, the first-order derivative of J with respect to η_i is

$$\left.\frac{\mathrm{d}J}{\mathrm{d}\eta_i}\right|_{\eta_i=0} = -\int_{t_i}^{t_{i+1}} [C_s'(s(u))F_s'(s(t_i^+), u - t_i)G_w'(w_i, s(t_i^-))\mathrm{e}^{-ru}]\mathrm{d}u + M_w'(w_i)\mathrm{e}^{-rt_i} - M_w'(w_{i+1})F_s'(s(t_i^+), t_{i+1} - t_i)\frac{G_w'(w_i, s(t_i^-))}{G_w'(w_{i+1}, s(t_{i+1}^-))}\mathrm{e}^{-rt_{i+1}}. \tag{13a}$$

Again, we will use $\frac{\mathrm{d}J}{\mathrm{d}\eta_i}$ to represent $\frac{\mathrm{d}J}{\mathrm{d}\eta_i}|_{\eta_i=0}$. Substituting (2), (3), (5) and (7) into (13a), we obtain after some algebraic manipulations that

$$\frac{\mathrm{d}J}{\mathrm{d}\eta_i} = [(P + m_1)\cdot y^{r/b} - y(P + Q(y))]\mathrm{e}^{-rt_{i+1}}, \quad i = 1, 2 \ldots n-1, \tag{13b}$$

where

$$y \equiv \mathrm{e}^{b(t_{i+1}-t_i)},$$

$$P = \frac{c_1 g_1}{(b - r)(g_2 + g_3/s(t_i^-))},$$

and

$$Q(y) = m_1\frac{g_2 + g_3/s(t_{i+1}^-)}{g_2 + g_3/s(t_i^-)} = \frac{m_1}{g_2 + g_3/s(t_i^-)}\left(g_2 + \frac{g_3}{s(t_i^+)y}\right).$$

Note that $y \geqslant 1$ because $b > 0$ and $t_{i+1} \geqslant t_i$. And $\frac{\mathrm{d}J}{\mathrm{d}\eta_i}$ has the following mathematical property.

Proposition 2. *$\frac{\mathrm{d}J}{\mathrm{d}\eta_i} \leqslant 0$ if $s(t_i^-) \geqslant 0$ and $y \in [1, \infty)$, $\forall i = 1, 2, \ldots, n-1$.*

Proof. See Appendix B. □

Therefore, whenever a resurfacing is conducted, J decreases with resurfacing intensity. Optimality is reached only at the boundary where increasing resurfacing intensity violates constraint (6). This leads to the following corollary.

Corollary 1. *On an optimal trajectory, $w_i = R_i$, $i = 1, 2, \ldots, n-1$.*

Note additionally that R_i as defined in (6) is a non-decreasing function of $s(t_i^-)$. It intersects the decreasing function $w_i^*(s(t_i^-))$ exactly once; see in Fig. 2(b). The roughness at this intersection, s^*, satisfies:

$$s^* = \frac{r(m_1 g_3 + m_2)}{c_1 g_1 + (b - r)m_1 g_2}. \tag{14}$$

From now on we call s^* the "trigger roughness". It appears in the following proposition.

Proposition 3. *For an optimal trajectory having a resurfacing at t_i: if $s(t_i^-) > s^*$, then $t_1 = t_2 = \cdots = t_i = 0$; else if $s(t_i^-) < s^*$, $t_i = t_{i+1} = \cdots = t_n = T$.*

Proof. If $s(t_i^-) > s^*$, then $w_i^*(s(t_i^-)) < R_i$ as shown in Fig. 2(b). For the optimal resurfacing at t_i with intensity $w_i = R_i > w_i^*(s(t_i^-))$, $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_i}$ in (10) is positive; see Fig. 2(a). Then according to (9), $t_i = t_{i-1}$. Note that $s(t_{i-1}^-) \geqslant s(t_{i-1}^+) = s(t_i^-) > s^*$, therefore $t_{i-1} = t_{i-2}$. Repeating this, we obtain $t_1 = t_2 = \cdots = t_i$ and $\frac{\mathrm{d}J}{\mathrm{d}\varepsilon_1} > 0$. If $t_1 \neq 0$, we can always perturb t_1 to the left to reduce J. Therefore, for the trajectory to be optimal, we must have $t_1 = t_2 = \cdots = t_i = 0$. This proves the first half of the proposition. The second half while $s(t_i^-) < s^*$ can be proven similarly. □

The above discussion about $\frac{\mathrm{d}J}{\mathrm{d}\eta_i}$ applies to $i \leqslant n-1$. When $i = n$, however, there is no next resurfacing at $t_{i+1} = T$. The last term in (13a) should be dropped:

$$\frac{\mathrm{d}J}{\mathrm{d}\eta_n} = -\int_{t_n}^{T} [C_s'(s(u))F_s'(s(t_n^+), u - t_n)G_w'(w_n, s(t_n^-))\mathrm{e}^{-ru}]\mathrm{d}u + M_w'(w_n)\mathrm{e}^{-rt_n}$$
$$= (P + m_1 - P\mathrm{e}^{(b-r)(T-t_n)})\mathrm{e}^{-rt_n}. \tag{15}$$

When $s(t_i^-)$ is known, Eq. (15) contains only one decision variable n.

The resurfacing intensity w_n should satisfy the boundary feasibility constraint (1g), such that

$$F(s(t_n^-) - G(w_n, s(t_n^-)), T - t_n) \leqslant s_T, \quad w_n \in (0, R_n]. \tag{16}$$

For optimality, w_n should be as large as possible if $\frac{\mathrm{d}J}{\mathrm{d}\eta_n} < 0$; or else as small as possible if $\frac{\mathrm{d}J}{\mathrm{d}\eta_n} \geqslant 0$. In summary, the conditions for optimality are

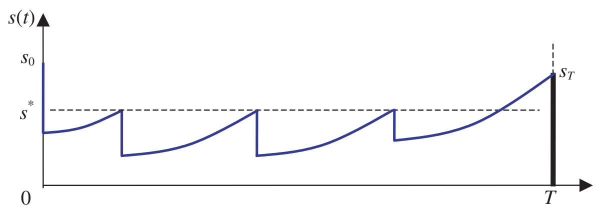
$$\begin{cases} \dfrac{\mathrm{d}J}{\mathrm{d}\eta_n} \geqslant 0, \quad 0 < w_n \leqslant R_n, \quad \text{and } s(T) = s_T; \quad \text{or} \\ \dfrac{\mathrm{d}J}{\mathrm{d}\eta_n} < 0, \quad w_n = R_n \quad \text{and } s(T) \leqslant s_T. \end{cases} \tag{17}$$

3.2. Algorithm

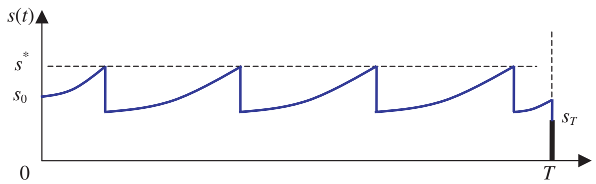
The previous section has derived necessary conditions for a trajectory to be optimal. According to Proposition 3, the trigger roughness s^* plays an important role in resurfacing planning. Whenever roughness is larger than or equal to s^*, resurfacing must be taken as soon as possible; else if roughness is smaller than s^*, resurfacing should be postponed (except at time T). Corollary 1 says that all resurfacing activities must be conducted with their maximal intensities, except for the last activity that should satisfy condition (17) instead.

Note that n itself is a decision variable and there may be multiple feasible trajectories that satisfy the optimality conditions described above and the boundary constraints. Fig. 3(a) and (b) shows some possible trajectories qualitatively. The algorithm proposed here enumerates all these roughness trajectories; the one with the least costs provides the global optimum.

To construct a trajectory that satisfies all necessary optimality conditions, we note that all these conditions, as stated in Corollary 1, Proposition 3, and (17), are local. They involve only a few state or decision variables that are relevant to only a small portion of the trajectory, and can be used to guide the design of a simple recursive algorithm. Starting from $t_0 = 0$ and $s(0) = s_0$, specifically, the algorithm constructs optimal trajectories by recursively determining the next optimal resurfacing time $t_{i+1} \leqslant T$ and intensity w_{i+1}. Every time t_{i+1} is determined, we test (17) to determine whether $n = i + 1$ can be true. If so, we record that trajectory as one local optimal solution. This continues until further actions are definitely non-optimal; i.e., when $t_{i+1} = T$ and $s(t_{i+1}^+) < s_T$.



(a) $s_0 > s^*$, s_T large, and $t_n < T$



(b) $s_0 < s^*$, s_T small, and $t_n = T$

Fig. 3. Local optimal roughness trajectories.

A possible implementation of this algorithm is given as follows:

1. Calculate s^* by (14); set $i \leftarrow 0$;
   Set $n \leftarrow 0$ with no resurfacing activity; if $\frac{dJ}{d\eta_n} \geqslant 0$ and (1g) holds, record $(\mathbf{t}, \mathbf{w})$;
2. If $s(t_i^+) \geqslant s^*$, let $t_{i+1} \leftarrow t_i = 0$, $s(t_{i+1}^-) \leftarrow s(t_i^+)$; else if $s(t_i^+) < s^*$, project $s(t)$ into $t > t_i$ by (7); calculate t_{i+1} such that $s(t_{i+1}^-) = s^*$; if $t_{i+1} \geqslant T$, set $t_{i+1} \leftarrow T$;
3. Let $n \leftarrow i + 1$; calculate $\frac{dJ}{d\eta_n}$ by (15);
   If $\frac{dJ}{d\eta_n} < 0$ then let $w_n \leftarrow R_n$; if (1g) holds, record $(\mathbf{t}, \mathbf{w})$;
   If $\frac{dJ}{d\eta_n} \geqslant 0$, then calculate the smallest w_n from (16); if solution exists, record $(\mathbf{t}, \mathbf{w})$;
4. Let $w_{i+1} \leftarrow R_{i+1}$; calculate $s(t_{i+1}^+)$ by (5); if $t_{i+1} = T$ and $s(t_{i+1}^+) < s_T$, go to Step 6;
5. Let $i \leftarrow i + 1$, repeat Step 2;
6. Pick from all recorded $(\mathbf{t}, \mathbf{w})$'s the one with the least cost J.

3.3. Remarks

As we can infer from the above algorithm, the optimal solution to this finite horizon pavement resurfacing problem has a "threshold" structure. This is consistent with known solutions to many other dynamic programming problems; see Bertsekas (1995). The new finding in this paper is that the threshold for this problem turns out to be constant over the entire horizon; the roughness trajectory reaches a "steady state" after a few resurfacings, and stays there until the last several actions, if needed, at the end of the horizon. In that "steady state", resurfacing is conducted whenever pavement roughness deteriorates to s^*, at a relatively steady pace. Obviously, the boundary conditions at both ends of the planning horizon affect only a few actions at both ends of the trajectory.

To the best knowledge of the authors, this is a new result for finite-horizon pavement resurfacing problems. It is consistent with previous findings that were obtained in a different setting or by using different methodologies. For example, Tsunokawa and Schofer (1994) used an approximation approach to a similar problem and found that the optimal policy reaches a steady state after just a few initial resurfacings. Li and Madanat (2002) later came up with similar arguments by analyzing an infinite-horizon problem.

Finally, Eq. (14) clearly shows how the threshold roughness, s^*, varies with the cost and performance model parameters. In general, it increases with agency cost parameters, and decreases with the user cost parameter. The role of deterioration parameter b is interesting: the faster the pavement deteriorates (larger b), the smaller the optimal threshold roughness.

4. Numerical example

In this example we plan resurfacing activities on a 1 km-long lane of a highway pavement over $T = 100$ years. We assume that the rate of variable user costs per one QI unit on this facility is 0.0002 \$/QI/veh, and the annual traffic volume in that lane is 5×10^6 veh/year. Then $c_1 = 1000$ \$/QI/year. Let $c_2 = 0$ \$/year for simplicity. (This assumption does not affect the optimization procedure.) The variable and fixed cost parameters for an overlay on that 1 km lane are assumed to be $m_1 = 11{,}000$ \$/mm, $m_2 = 150{,}000$ \$.

Table 1
Optimal resurfacing plan for the case study

Action i	0	1	2	3	4	5	6	$n = 7$
t_i (year)	0.00	0.00	0.00	19.31	40.86	62.41	83.96	100.00
$s(t_i^-)$ (QI)	150.00	150.00	51.03	45.62	45.62	45.62	45.62	34.64
R_i (mm)	–	100.80	46.37	43.39	43.39	43.39	43.39	37.35
w_i (mm)	–	100.80	46.37	43.39	43.39	43.39	43.39	7.58
$s(t_i^+)$ (QI)	150.00	51.03	17.37	15.53	15.53	15.53	15.53	30.00

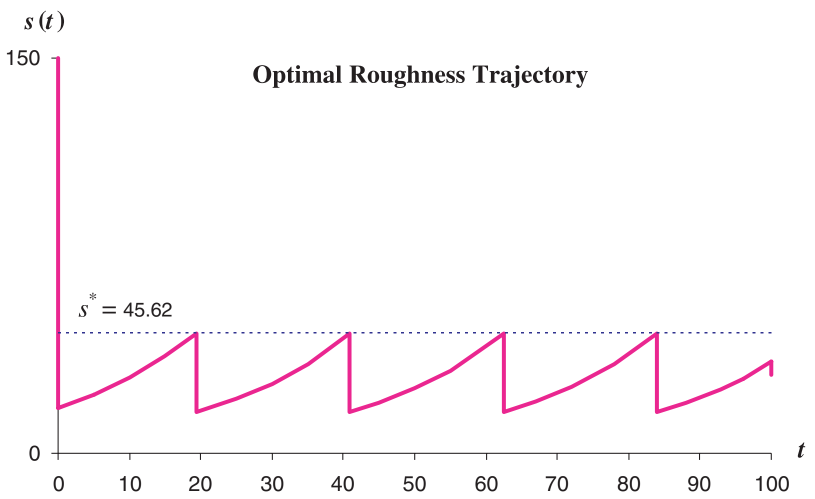


Fig. 4. Optimal roughness trajectory for the case study.

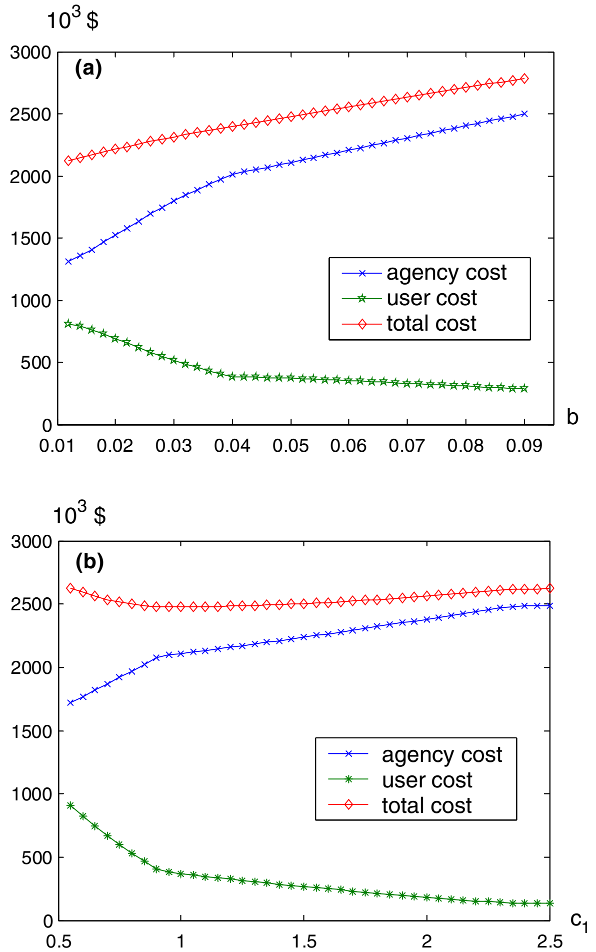


Fig. 5. Sensitivity of the optimal cost to some parameters.

We assume a deterioration parameter $b = 0.05$ year^{-1} so that a good pavement with a roughness of 25 QI will deteriorate to a poor condition of 100 QI in about 28 years. This value is used in the literature as well (Tsunokawa and Schofer, 1994; Li and Madanat, 2002). We assume the pavement begins with a very poor condition of $s_0 = 150$ QI, and it must reach $s_T = 30$ QI at the end of the planning horizon. Finally, the discount rate $r = 0.07$.

By using (14), the trigger roughness $s^* = 45.62$ QI. The optimal resurfacing plan is shown in Table 1, and the corresponding optimal roughness trajectory is plotted in Fig. 4. Note that two resurfacing activities are conducted at $t = 0$ before the system enters the "steady state"; one resurfacing activity is done at time $t = T$ to reach the final roughness requirement (1g). The minimal discounted total cost in this horizon is 2,478,500 $.

We can analyze the sensitivity of the optimal solution to specific parameters while holding the other parameters fixed. For example, Fig. 5(a) shows how the optimal cost and its components (agency cost and user cost) vary with parameter b (i.e., deterioration rate). When the pavement deteriorates faster (larger b), the trigger roughness s^* decreases, so the user cost decreases as well; the agency, however, has to spend more to maintain a lower roughness level. Overall, hence, the optimal system cost increases with b almost linearly. Fig. 5(b) shows how the optimal cost varies with the user cost parameter c_1 when all other parameters are fixed. When the user cost parameter increases, the trigger roughness s^* decreases and so does the user cost. The total cost, however, does not increase strictly and exhibits a flat and convex shape.

5. Conclusion

This paper has presented an analytical solution to the problem of optimizing pavement resurfacing frequency and intensity in a finite horizon. A simple algorithm was developed to solve for the exact optimal solution. The presented work focused on resurfacing activities only, but a similar approach can be applied to other rehabilitation activities as well.

The exact optimal resurfacing strategy found in this paper coincides with findings previously obtained by using an approximate approach or for the case of an infinite horizon (Tsunokawa and Schofer, 1994; Li and Madanat, 2002): calculate a trigger roughness and resurface with the maximum effective intensity so that roughness reduces below that trigger roughness as soon as possible. After that, resurface whenever the pavement deteriorates to that level. In general, the policy enters a steady state after just a few resurfacing activities, and stays there until near the end of the planning horizon. The initial roughness and the final roughness requirements affect only a few resurfacing activities.

The threshold structure of optimal solutions to our pavement resurfacing planning problem is also consistent with the structure of most optimal policies in the MDP framework. This also reminds us of familiar models in other operations research contexts; e.g., optimal Q–R policies for continuous review inventory systems, and the bang–bang policies in Friesz and Fernandez (1979). This connection then sheds light on the optimal resurfacing policy under uncertainty. Uncertainty arises, for example, from the performance models when deterioration and resurfacing effectiveness models are subject to random exogenous factors (e.g., traffic and precipitation). As the effect of these uncertainties is often additive, the threshold structure of the resurfacing policy will still hold.

Appendix A

In Fig. 1, when t_i varies by an infinitesimal time interval ε_i, the roughness before resurfacing equals

$$s(t_i^- + \varepsilon_i) = F(s(t_i^-), \varepsilon_i) \cong s(t_i^-)(1 + b\varepsilon_i), \tag{A.1}$$

and the trajectory will reach point A′. The new resurfacing at $(t_i + \varepsilon_i)$ should reach B′ in order for the rest of the trajectory to remain unchanged. Comparing the roughness paths [A–B–B′] with [A–A′–B′], the new resurfacing intensity $(w_i + \varepsilon_i')$ should satisfy

$$F((s(t_i^-) - G_w'(w_i, s(t_i^-))), \varepsilon_i) = s(t_i^-) + F(s(t_i^-), \varepsilon_i) - G(w_i + \varepsilon_i', s(t_i^-) + F(s(t_i^-), \varepsilon_i)). \tag{A.2}$$

Inserting (5) and (7) into (A.2), and ignoring higher order terms, we obtain that

$$\varepsilon_i' = g_2 b (g_2 + g_3/s(t_i^-))^{-1} w_i \varepsilon_i.$$

Next, we show that $w_i + \varepsilon_i' \leqslant R_i' \equiv g_2 s(t_i^- + \varepsilon_i) + g_3$ if $w_i \leqslant R_i = g_2 s(t_i^-) + g_3$, so that (6) is never violated by the perturbation. This ensures that (6) does not restrict the construction of alternative trajectories by perturbing t_i and w_i together as described.

$$\begin{aligned} w_i + \varepsilon_i' &= \lfloor 1 + g_2 b (g_2 + g_3/s(t_i^-))^{-1} \varepsilon_i \rfloor w_i \\ &\leqslant \left[1 + \frac{g_2 b \varepsilon_i}{g_2 + g_3/s(t_i^-)}\right](g_2 s(t_i^-) + g_3) \\ &= (g_2 s(t_i^-) + g_3 + g_2 b \varepsilon_i) = g_2 s(t_i^- + \varepsilon) + g_3. \end{aligned}$$

Appendix B

This section proves that $\frac{\mathrm{d}J}{\mathrm{d}\eta_i} \leqslant 0$ if $s(t_i^-) \geqslant 0$ and $y = \mathrm{e}^{b(t_{i+1}-t_i)} \in [1,\infty)$, $\forall i = 1,2,\ldots,n-1$, as stated in Proposition 2.

According to (13b), we only need to prove that

$$(P + m_1) \cdot y^{r/b} - y \cdot (P + Q(y)) \leqslant 0 \quad \text{if } y \in [1,\infty). \tag{B.1}$$

Or more specifically,

$$(P + m_1) \cdot y^{r/b} - \left[\left(P + \frac{m_1 g_2}{g_2 + g_3/s(t_i^-)}\right) \cdot y + \frac{m_1 g_3/s(t_i^+)}{g_2 + g_3/s(t_i^-)}\right] \leqslant 0 \quad \text{if } y \in [1,\infty). \tag{B.2}$$

Now we treat $s(t_i^-) \geqslant 0$ as a parameter. Because $s(t_i^+) \leqslant s(t_i^-)$, $Q(1) \geqslant m_1$. It is easy to verify that (B.2) is true at $y = 1$.

Note that if $t_i = T$, we must have $t_i = t_{i+1}$, thus $y = 1$ and (B.2) is true. Now consider the case when $t_i < T$. If the left hand side of (B.2) decreases with y when $y \geqslant 1$, then (B.2) is true. We only need to prove that the derivative of the left hand side with respect to y,

$$\frac{r}{b}(m_1 + P) y^{(r-b)/b} - \left(P + \frac{m_1 g_2}{g_2 + g_3/s(t_i^-)}\right),$$

is negative when $y \in [1,\infty)$ and $t_i < T$.

By Proposition 1, on an optimal trajectory, if $t_i < T$:

$$r m_1 < \frac{c_1 g_1 + m_1 g_2 b}{g_2 + g_3/s(t_i^-)} = (b - r)P + \frac{m_1 g_2}{g_2 + g_3/s(t_i^-)} b. \tag{B.3}$$

Eq. (B.3) immediately gives $\frac{r}{b}(m_1 + P) < P + \frac{m_1 g_2}{g_2 + g_3/s(t_i^-)}$. There are two possibilities:

(i) If $b - r < 0$. Then $y^{(r-b)/b} \geqslant 1$ on $[1,\infty)$. Since $m_1 > 0$, for (B.3) to hold we have $m_1 + P < 0$. Thus

$$\frac{r}{b}(m_1 + P) y^{(r-b)/b} < \frac{r}{b}(m_1 + P) < P + \frac{m_1 g_2}{g_2 + g_3/s(t_i^-)}; \tag{B.4}$$

(ii) If $b - r \geqslant 0$. Then $y^{(r-b)/b} \leqslant 1$ on $[1,\infty)$, and $m_1 + P > 0$. Eq. (B.4) also holds.

In both cases, (B.4) is true, and the derivative of the left hand side of (B.2) is negative when $y \in [1,\infty)$. Then (B.2) is true. This completes the proof. □

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.

Bellman, R.E., 1955. Equipment replacement policy. Journal of the Society of Industrial and Applied Mathematics 8 (3), 133–146.

Bertsekas, D.P., 1995. Dynamic Programming and Optimal Control, Vol. 1. Athena Scientific, Belmont, MA.
Carnahan, J.V., Davis, W.J., Shahin, M.Y., Kean, P.L., Wu, M.I., 1987. Optimal Maintenance Decisions for Pavement Management. Journal of Transportation Engineering 113 (5), 554–572.
Dreyfus, S., 1960. A generalized equipment replacement study. Journal of the Society of Industrial and Applied Mathematics 8 (3), 425–435.
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 Part B 13, 317–339.
Golabi, K., Kulkarni, R., Way, G., 1982. A statewide pavement management system. Interfaces 12 (6), 5–21.
Jacobs, T.L., 1992. Optimal long-term scheduling of bridge deck replacement and rehabilitation. Journal of Transportation Engineering 118, 312–322.
Klein, M., 1962. Inspection–maintenance–replacement schedules under Markovian deterioration. Management Science 9, 25–32.
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.
Ouyang, Y., Madanat, S., 2004. Optimal scheduling of rehabilitation activities for multiple pavement facilities: exact and approximate solutions. Transportation Research Part A 38, 347–365.
Paterson, W.D.O., 1990. Quantifying the effectiveness of pavement maintenance and rehabilitation. In: Proceedings at the 6th REAAA Conference, Kuala Lumpur, Malaysia.
Tsunokawa, K., Schofer, J.L., 1994. Trend curve optimal control model for highway pavement maintenance: case study and evaluation. Transportation Research Part A 28, 151–166.