Optimal Route Planning under Uncertainty

Evdokia Nikolova
MIT Computer Science & AI Lab
Cambridge MA 02139, USA
enikolova@csail.mit.edu

Matthew Brand
Mitsubishi Electric Research Labs
Cambridge, MA 02139, USA
brand@merl.com

David R. Karger
MIT Computer Science & AI Lab
Cambridge MA 02139, USA
karger@csail.mit.edu

Abstract

We present new complexity results and efficient algorithms for optimal route planning in the presence of uncertainty. We employ a decision theoretic framework for defining the optimal route: for a given source S and destination T in the graph, we seek an ST-path of lowest expected cost where the edge travel times are random variables and the cost is a nonlinear function of total travel time. Although this is a natural model for route-planning on real-world road networks, results are sparse due to the analytic difficulty of finding closed form expressions for the expected cost (Fan, Kalaba & Moore), as well as the computational/combinatorial difficulty of efficiently finding an optimal path which minimizes the expected cost. We identify a family of appropriate cost models and travel time distributions that are closed under convolution and physically valid. We obtain hardness results for routing problems with a given start time and cost functions with a global minimum, in a variety of deterministic and stochastic settings. In general the global cost is not separable into edge costs, precluding classic shortest-path approaches. However, using partial minimization techniques, we exhibit an efficient solution via dynamic programming with low polynomial complexity.

Keywords: route planning under uncertainty, non-linear objective, stochastic shortest path, complexity, algorithms.

Introduction

In this paper, we present new complexity results and efficient algorithms for path planning under uncertainty. The motivation for the problem comes from route planning in road networks. Current navigation systems use information about road lengths and speed limits to compute deterministic shortest or fastest paths. When realized (driven), these paths often turn out to be quite suboptimal, for the simple reason that the deterministic solution ignores the inherent stochasticity of traffic as well as changing traffic conditions. The statistics of traffic flows are now estimable in real time from road sensor networks, thus we ask how effectively and efficiently such information can be exploited.

The *static stochastic* route planning problem asks for optimal routes on a graph where travel times on the edges are random variables with fixed distributions. In this setting, one must optimize an objective that makes some trade-off between the speediness (expected travel time) and reliability (variance) of a route. Optimizing one or the other, though quite tractable, makes little sense. For example, finding the route with the lowest expected travel time has little value because a driver can only sample a single realization of that drive in current traffic conditions; with variance unoptimized, that realization could be quite far from the mean. Optimizing a linear combination of the mean and variance is another possibility, though it seems ad-hoc and not clearly motivated, interestingly it turns out to be a special case of our formulation.

Decision theory, the standard framework for making optimal plans and policies under uncertainty, expresses the trade-off between speediness and reliability through a utility or cost function $C : \mathcal{R} \rightarrow \mathcal{R}^+$. In our setting $C(t)$ assesses a reward or penalty for arriving at time t relative to a deadline. For example, a linear $C(t)$ minimizes expected travel time; quadratic $C(t)$ minimizes variance; the minimizer of their weighted sum takes a surprising form related to the cumulant generating function of the travel time distributions (see last Section), however it cannot tell us when to set out.

We will consider a variety of stochastic route planning problems, with an emphasis on cost functions that value timeliness without time-wasting. *E.g.* "What is the optimal start time and route for a given deadline?" and "Now that I am on the road, what is the optimal route for that deadline?" Surprisingly, for some cost functions of interest, the former question is tractable while the latter is NP-hard.

This highlights the dependence of stochastic solutions on time. For example, imagine that we have a choice of two routes and only care to arrive at the destination before a given deadline. Maximizing the probability of doing so implies that $C(t)$ is a step function. If we set out close to the deadline, a slower and highly *variable* route will actually be preferable to a faster and highly reliable route, because the less predictable route offers a greater chance of arriving on time (see Figure 1). Note that this function is monotone increasing and as such cannot be used to plan an optimal departure time: it would imply that the best time to set out is the dawn of time.

Copyright © 2006, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

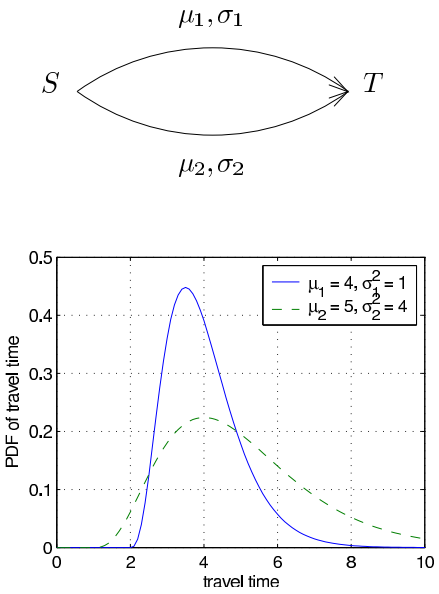


Figure 1: The optimal route from S to T depends on the start time. Route one is on average faster ($\mu_1 < \mu_2$) and more reliable ($\sigma_1 < \sigma_2$), but if one starts less than 3 minutes before the deadline, route two offers a higher (albeit small) probability of avoiding lateness.

Related Work

Traditionally, the work on path planning in stochastic networks has focused on the notion of shortest paths in expectation (Papadimitriou & Yannakakis 1991), (Bertsekas & Tsitsiklis 1991). Some models have added costs on the edges in addition to travel times where the costs depend on the realized travel times and in this way can capture a measure of uncertainty (Chabini 2002), (Miller-Hooks & Mahmassani 2000). However there has been little work on decision-theoretic models which directly incorporate uncertainty and output an optimal path on the basis of a comprehensive measure of user utility and all available distributional information of the stochastic edge weights.

In particular, two lines of work most closely resemble our setting. The first (Loui 1983) considers a similar decision theoretic framework for optimal paths under uncertainty, however the author only studies monotone increasing costs. These are arguably easier since they admit exact efficient solutions for a number of special cases, including linear and exponential objectives (Loui 1983), as well as arbitrary costs with identically distributed edge weights (Nikolova 2005). Mirchandani and Soroush (1985) extend Loui's work to a quadratic cost function of the travel time, however their algorithm is essentially an exhaustive search over all potentially optimal paths, and thus exponential in the worst case.

The second line of work (Fan, Kalaba & Moore) considers a special monotone increasing cost (the probability of arriving late) and suggests that the Gamma distribution is natural for modelling stochastic edge travel times.

There are also two related problems, one easier and one harder. First, Markov decision theory most naturally leads to the construction of on-line policies, thus the stochastic route planning problem has been considered mainly in the context of adaptive algorithms that compute the optimal next edge in light of travel times already realized en route to the current node (Fan, Kalaba & Moore), (Gao & Chabini 2002), (Boyan & Mitzenmacher 2001). Some of the results presented below can be adapted to compute these policies in closed form. Second, approximations for expected shortest paths in stochastic networks with nonstationary (time-varying) distributions have also been proposed, *e.g.,* (Miller-Hooks & Mahmassani 2000), (Fu & Rilett 1998), (Gao & Chabini 2002), (Hall 1986). However most of the approximations are based on heuristics with unknown approximation ratios. This is not surprising in light of a recent result that the problem with time-varying distributions is in general #P-hard (Nikolova 2005).

Our results

We give a variety of hardness results and algorithms for a natural decision-theoretic framework for route planning under uncertainty. A major obstacle for studying this framework has been the difficulty of finding closed-form expressions for the expected cost function, as well as the non-separability of the cost function into the edges, precluding standard dynamic programming techniques.

We identify a family of appropriate cost models for drivers and uncertainty models for road networks. In a departure from the stochastic path-planning literature, these are continuous and closed under convolution, so that the expected cost of any one path can be computed analytically. We survey a range of stochastic route-planning problems, finding that some can be converted into classic deterministic shortest-path. We prove hardness of approximability for simple paths (which do not contain loops) and NP-hardness for general paths for a very broad class of cost functions and fixed start times. Our hardness results extend in particular to stochastic (*e.g.,* Gamma-distributed) travel times. This is not generally implied by the hardness proofs for deterministic travel times since there are known instances of problems which are NP-hard in a deterministic setting, yet become polynomially solvable in a stochastic setting (Bruno *et al.* 1981).

We consider a richer decision-theoretic framework than (Loui 1983), by defining the objective both as a function of the path and the departure time at the source. This allows us to distinguish between two problems, finding the optimal path for a fixed departure time, as well as planning an optimal departure time. We show that for some cost functions the latter problem (which optimizes over two variables, path and departure time) reduces to deterministic shortest path while the former (optimizing only over path) is NP-hard. Focusing on the NP-hard instances, we exhibit pseudopolynomial algorithms which have low polynomial complexity in the size of the graph and the largest mean travel time of an edge. The average travel times are almost certainly polynomially bounded in real world roadway networks. With this, our algorithm offers the first practical solution, which

is much more efficient than the previous exponential algorithms based on exhaustive search.

In the last section, we show that our model admits as a special case a standard objective in mean-risk analysis, which aims to optimize a linear combination of the mean and variance of the random variables.

Problem Statement & Preliminaries

Let $G = \{V, E\}$ be a directed graph with a source node S and destination node T. Assume that the time to traverse an edge $e \in E$ in the graph follows a distribution with probability density function $f_e(.)$ and the travel times on different edges are independent. Suppose a driver needs to reach the destination by a given deadline, denoted as time 0. The penalty for arriving at time t is denoted by $C(t)$; t is positive for late arrivals and negative for early arrivals.

Let e be the last edge on a path to the destination. Then the expected cost $EC(t)$ of starting to traverse this edge at time t is given by the convolution $EC(t) = \int_0^\infty f_e(y)C(t+y)dy$. By independence of the edge travel times, the expected cost of traversing a path $P = \{e_1, \cdots, e_r\}$ departing at time t is

$$EC_P(t) = \int_0^\infty ... \int_0^\infty \Big[f_{e_1}(y_1)...f_{e_r}(y_r) \; C(t + y_1 + ... + y_r) \Big] dy_1...dy_r. \quad (1)$$

We now distinguish two different problems:

1. Find the optimal path P and optimal start time t:
$$\min_{P,t} EC_P(t). \quad (2)$$
2. Find the optimal path for a given start time.

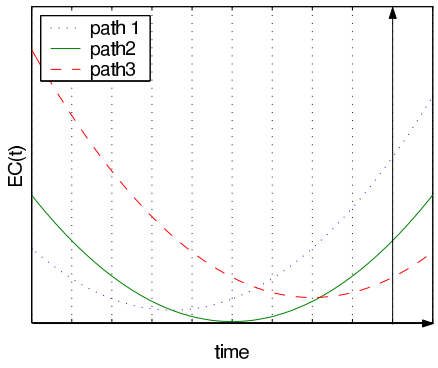


Figure 2: Each path has an associated expected penalty function $EC(t)$ which takes as argument the start time t. If we depart at the time marked by the vertical arrow, path 3 is optimal, however the globally optimal start time is located at the minimum of path 2.

If we graph the expected cost of each path as a function of start time, we obtain a family of curves, cartooned in Figure 2. The best path for a given start time is indicated by the lowest curve at that point. Note that each path may be optimal over a different range of start times. The global miminum of the lower envelope of all such curves indexes the optimal time to start out.

Calculating the Cost of a Single Path

In general, the expected cost expression in equation (1) may be impossible to compute in a closed form. We will therefore focus on two families of cost functions for which the integral can be computed exactly, and which are a sensible model of user utility: polynomials and exponentials. We assume that the driver values her time and does not want to set out too early or arrive too early, thus the cost function should be expressive enough to (asymmetrically) penalize both lateness and earliness[1]. Although many of the results of subsequent sections apply to general polynomial functions (and our hardness results hold for arbitrary functions with global minima), we will consider here quadratic and quadratic+exponential cost functions for illustrative purposes.

Quadratic Cost Suppose the cost of reaching the destination at time t is $C(t) = t^2$. Suppose the path from the source to the destination consists of a single edge with random travel time Y of density $f(.)$, mean μ and variance σ^2. Then the expected cost of departing the source at time t is

$$\begin{aligned} EC(t) &= \int_0^\infty f(y)(t+y)^2 dy \\ &= t^2 + 2tE[Y] + E[Y^2] = (t+\mu)^2 + \sigma^2 \end{aligned}$$

If instead the path consists of r edges with random travel times Y_i having density $f_i(.)$, mean μ_i and variance σ_i^2 for $i = 1, ..., r$, then iterating the above calculation r times gives

$$\begin{aligned} EC(t) &= \int_0^\infty ... \int_0^\infty \Big[f_1(y_1)...f_r(y_r) \; C(t + y_1 + ... + y_r) \Big] dy_1...dy_r \\ &= \Big(t + \sum_{i=1}^r \mu_i \Big)^2 + \sum_{i=1}^r \sigma_i^2. \end{aligned}$$

Therefore, the cost of a path is minimized at start time $t = -\sum_{i=1}^r \mu_i$, the (negative) average travel time for that path. At this optimum, the expected cost value is the variance of the path, $EC_{min} = \sum_{i=1}^r \sigma_i^2$.

The quadratic cost function might not be regarded as realistic since it assigns the same penalty to being equally early and late. As we saw above, this leads to preferring the most certain route, without any care for the average travel time. On the other hand, linear costs, which favor on average faster paths, do not have any effect when added to the quadratic cost other than shifting the effective deadline. Thus, we augment the quadratic cost function with an exponential term which gives a higher penalty to being late.

Quadratic+Exponential Cost Consider cost function $C(t) = t^2 + \lambda e^{kt}$, where again t is the time of arrival with

[1]It may make sense to penalize early arrival even when the driver is on the road and requesting a new path, because, as in the case of transportation depots, the destination may not have the capacity to accept early arrivals.

respect to deadline at time 0 and $\lambda \geq 0$ and k are parameters which determine the strength of the penalty for being late. The sign of k can be negative if one is more averse to arriving early than late.

In this case, we can still get a closed form expression for the expected cost of a path. For a path of one link with density $f(Y)$, mean μ and variance σ^2, we get

$$\begin{aligned} EC(t) &= \int_0^\infty f(y)\big[(t+y)^2 + \lambda e^{k(t+y)}\big]dy \\ &= (t+\mu)^2 + \sigma^2 + \lambda e^{kt} E[e^{kY}] \end{aligned}$$

where $E[e^{kY}]$ is the moment-generating function of the density of Y. When the path consists of r links with densities $f_i(Y_i)$, means μ_i and variances σ_i^2, for $i = 1, .., r$, the expected cost of departing at time t is

$$EC(t) = \Big(t + \sum_{i=1}^r \mu_i\Big)^2 + \sum_{i=1}^r \sigma_i^2 + \lambda e^{kt} \prod_{i=1}^r E[e^{kY_i}].$$

Choice of Travel Time Distributions

Traditionally, the travel times on the edges have been modeled by normal distributions. For a normally distributed random variable $Y \sim N(\mu, \sigma^2)$, we have $E[e^{kY}] = exp(k\mu + k^2\sigma^2/2)$. Thus, the expected cost of a path with r independent normally distributed edge travel times is given by

$$EC(t) = \Big(t + \sum_{i=1}^r \mu_i\Big)^2 + \sum_{i=1}^r \sigma_i^2 + \lambda e^{kt} e^{k\sum_i (\mu_i + k\sigma_i^2/2)}. \quad (3)$$

However, the normal distribution is unrealistic as it assigns positive probability to negative travel times. A more physically appropriate distribution is obtained by observing that in ordinary (collision-free) traffic, car arrivals at any particular landmark can be viewed as a Poisson process, which implies that travel times are Gamma distributed; Gamma distributions are also proposed in (Fan, Kalaba & Moore). We write $t_e \sim \gamma(a_e, b_e)$ to indicate that the time to traverse an edge e in the graph is Gamma distributed with shape parameter a_e and width parameter b_e. The mean of this distribution is given by $\mu_e = a_e b_e$ and the variance is $\sigma_e^2 = a_e b_e^2$. The density of the gamma distribution is given by

$$\gamma(a, b, y) = \frac{y^{a-1} e^{-y/b}}{b^a \Gamma(a)},$$

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt$ is the gamma function. The Gamma distribution has strictly nonnegative support and we can additionally specify a minimum travel time by shifting y. To keep notation uncluttered, we will use unshifted distributions in this paper; the generalization to shifts is straightforward.

For a gamma random variable Y, $E[e^{kY}] = (1 - kb)^{-a}$ so the expected cost of a path with independent gamma distributed edge travel times is given by

$$\Big(t + \sum_{i=1}^r a_i b_i\Big)^2 + \sum_{i=1}^r a_i b_i^2 + \lambda e^{kt} \Big[\prod_{i=1}^r (1 - kb_i)^{-a_i}\Big], \quad (4)$$

which no longer has a simple analytic expression for the minimum.

Optimal Routing and Optimal Start Time

In this section we consider the subproblem of jointly optimizing for the path and start time. We show that the quadratic cost function with general edge distributions, as well as the quadratic+exponential cost with Gaussian distributions result in selecting the lowest variance path, and thus admit a standard shortest path solution. On the other hand, the quadratic+exponential cost with Gamma travel distributions does not satisfy the sub-path optimality property needed for a dynamic programming approach, and remains an open problem.

Recall that when $C(t) = t^2$, the expected cost of a single path is minimized at start time $t = -\sum_{i=1}^r \mu_i$, the negative average travel time for the path, and at this optimum, the expected penalty is the sum of the variances over the individual links, $EC_{min} = \sum_{i=1}^r \sigma_i^2$. Therefore, we can find the optimal path—the one of smallest total variance, with a simple application of Dijkstra's shortest path algorithm, where each edge e is labelled with its variance σ_e^2. Consequently, the optimal departure time would be given by the mean travel time of that path. Thus, the optimal path and optimal departure time problem turns out easy in the case of quadratic cost. The second problem of finding the optimal path for a given departure time does not benefit from the simple form of the expected cost function, we show in the following section that it is NP-hard.

If we add an exponential penalty for being late by taking $C(t) = t^2 + \lambda e^{kt}$, the expected cost for a path under Gamma distributions still has a simple closed form, given by Equation (4). However, we lose the separability property of the quadratic cost functions, which allows for a dynamic programming solution.

Theorem 1. *Finding the optimal path and optimal start time*

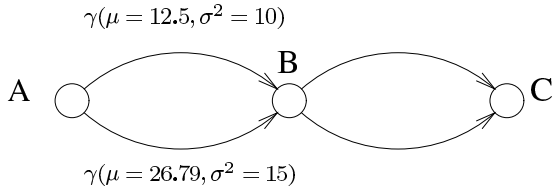
(i) under quadratic cost and general distributions can be solved exactly with a deterministic shortest path algorithm.

(ii) under quadratic+exponential cost and normal distributions can be solved exactly with a deterministic shortest path algorithm.

(iii) under quadratic+exponential cost and general distributions, may not satisfy subpath optimality for any subpath and thus precludes standard dynamic programming techniques.

Proof: Part (i) follows from the discussion above. We show part (iii) via a counterexample in which suboptimality does not hold for any subpath. Consider the graph with two parallel pairs of edges in Table 1. All edges are Gamma-distributed with mean-variance pairs $(12.5, 10)$ on the top and $(26.8, 15)$ on the bottom. The optimal path from A to C consists of the lower two edges, with optimal start time 74.79 units before the deadline (hence the negative sign in the table) and minimum cost 522.65. However, the best paths from A to B and from B to C both consist of the top edges, with departure time 22.18 before the deadline and minimum cost 123.14. Thus, no subpath of the optimal path from A to C is optimal.

Curiously, the same cost function with normally distributed edge travel times admits dynamic programming. In



Path	$A \to B$	$A \to C$
Top edges	$(-22.2, 123.1)$	$(-46.5, 526.7)$
Bottom & Top	-	$(-60.7, 524.6)$
Bottom edges	$(-36.3, 125.0)$	$(-74.8, 522.7)$

Table 1: In the network above, the top edges are identical and Gamma distributed, and so are the bottom edges. When the cost of arriving at time t is $C(t) = t^2 + e^t$, the optimal path from A to C uses the bottom edges while the optimal paths from A to B and from B to C use the top edges. The table entries give the values of the optimal start time and expected cost at the minimum for each path.

part (ii), the cost of leaving path P at time t given by Equation (3), can be written as

$$EC(\tilde{t}) = \tilde{t}^2 + s + e^{k\tilde{t}} e^{k^2 s/2}, \tag{5}$$

after the change of variables $\tilde{t} = t + \sum_{e \in P} \mu_e$ and $s = \sum_{e \in P} \sigma_e^2$. In particular, a path with a higher total variance will have an expected cost function strictly above that of a path with a lower variance, because for $s_1 < s_2$, $k \neq 0$ and for any fixed $\tilde{t}$, $e^{k\tilde{t}} e^{k^2 s_1/2} < e^{k\tilde{t}} e^{k^2 s_2/2}$. Hence the path of lowest variance will have the lowest minimum expected cost, and we can find it via any shortest path algorithm with edge weights equal to the variances. Thus, when travel times are normally distributed, both the quadratic and quadratic plus exponential cost functions will choose the same optimal path, although the optimal start time would naturally be earlier under the second family of cost functions. □

The problem of finding the optimal path at a given departure time is again NP-hard, as in the quadratic cost case. However, we shall see that dynamic programming there is more promising when combined with partial minimization.

Optimal Routing with a Given Start Time

In this section, we show NP-hardness and hardness of approximation results for arbitrary cost functions with global minima. We then give pseudopolynomial algorithms for the quadratic and quadratic+exponential cost functions, which generalize to polynomial (plus exponential) cost functions.

We may be interested in the optimal route and optimal departure time to a destination, while planning ahead of time. Once we start our journey, it is natural to ask for an update given that current traffic conditions may have changed. Now, we are really posing a new problem: to find the path of lowest expected cost, $EC(t_{start})$, for a given departure time t_{start}. This may sound like a simpler question than the one of finding optimal route and optimal start time though it turns out to be NP-hard for a very broad class of cost functions.

Complexity of Costs with Global Minimum

Let $C(t)$, the penalty for arriving at the destination at time t, be any function with a global minimum at t_{min}. In case of several global minima, let t_{min} be the smallest one. Denote the number of nodes in the graph by n.

Theorem 2. *The problem of finding a lowest-cost simple ST-path is NP-hard.*

Proof: Suppose all edges have deterministic unit edge lengths. Then the cost of departing at time t along a path with total length L is simply $C(t + L)$.

Consider departure time $t = t_{min} - (n - n^\epsilon)$. If there exists a path of length $n - n^\epsilon$, it would be optimal since its cost would be

$$C(t + n - n^\epsilon) = C(t_{min}) \leq C(t + L) \tag{6}$$

for all other paths of any length L. In particular, since t_{min} is the leftmost global minimum, we have a strict inequality for paths of length $L < n - n^\epsilon$. Now suppose the optimal path is of length L^*. We have three possibilities:

1. $L^* < n - n^\epsilon$. Then by above, there is no path of length $n - n^\epsilon$.
2. $L^* = n - n^\epsilon$. Then we have found a path of length $n - n^\epsilon$.
3. $L^* > n - n^\epsilon$. Then by removing edges, we can obtain a path of length exactly $n - n^\epsilon$.

Therefore, the problem of finding an optimal path reduces to the problem of finding a path of length $n - n^\epsilon$ where $\epsilon < 1$. Since the latter problem is NP-complete (Karger, Motwani & Ramkumar 1997), our problem is NP-hard. □

Intuitively, if we incur a higher cost for earlier arrivals and depart early enough, the problem of finding an optimal path becomes equivalent to the problem of finding the longest path. Further, if the cost function is not too flat on the left of its minimum, we can see that an approximation of the min-cost path automatically gives a corresponding approximation on the longest path, hence a corollary to the above is that the optimal path is hard to approximate.

Corollary 1. *For any cost function which is strictly decreasing and positive with slope of absolute value at least $\lambda > 0$ on an interval $[-\infty, t_{min}]$, there does not exist a polynomial constant factor approximation algorithm for finding a simple path of lowest expected cost at a given departure time prior to t_{min}, unless $P = NP$.*

Proof. Suppose the contrary, namely that we can find a path of cost $C = (1 + \alpha) C_{opt}$ where C_{opt} is the cost of the optimal path, and $\alpha > 0$ is a constant.

Assume as in the theorem above that we have an n vertex graph with unit length edges and consider departure time $t = t_{min} - (n - 1)$ at the source. Let L_{opt} be the length of the optimal path and let L be the length of the path that we find. Then $L \leq L_{opt}$ so L_{opt} is the longest path between the source and destination and $\frac{C - C_{opt}}{L_{opt} - L} \geq \lambda$,

otherwise there would be a point in $[-\infty, t_{min}]$ of absolute slope less than λ. Hence $L_{opt} - L \leq \lambda\alpha C_{opt}$ and so $L_{opt}/L \leq 1 + (\lambda\alpha C_{opt}/L) \leq 1 + \lambda\alpha C_{opt}$, where $C_{opt} = C(t_{min})$ is constant, so this would give a polynomial constant factor approximation algorithm for the longest path problem, which does not exist unless $P = NP$ (Karger, Motwani & Ramkumar 1997). □

Remark 1. *We can obtain a stronger inapproximability result, based on the fact that finding a path of length $n - n^{\epsilon}$ is NP-complete, for any $\epsilon < 1$ (Karger, Motwani & Ramkumar 1997). However, our goal is simply to show the connection between the inapproximability of our problem to that of the longest path problem and show the need to settle for non-simple paths in the algorithms of the following section.*

The NP-hardness result in Theorem 2 crucially relies on simple paths, and it makes optimal paths equivalent to longest paths (due to a very early departure time), which is not usually the case. We can show that finding optimal paths is NP-hard even for more reasonable start times and non-simple paths, via a reduction from the subset-sum problem, for any cost function with a unique global minimum.

Theorem 3. *Suppose we have a cost function $C(\tilde{t})$ with a unique minimum at t_{min}, which gives the penalty of arriving at the destination at time $\tilde{t}$. Then for a start time t at the source, it is NP-complete to determine if there is a path P to the destination of expected cost $EC_P(t) \leq K$, where*

$$EC_P(t) = \int_0^{\infty} f_P(y)C(t+y)dy,$$

and $f_P(y)$ is the travel time distribution on path P.

Proof. The problem is in NP since there is a short certificate for it, given by the actual path if it exists.

To show NP-hardness, we reduce from the Subset Sum problem, namely given a set of integers $\{x_1, ..., x_n\}$ and a target integer t, is there a subset which sums exactly to t? The subset sum problem, which is a special case of the knapsack problem, is NP-complete (Chvatal 1980). Set

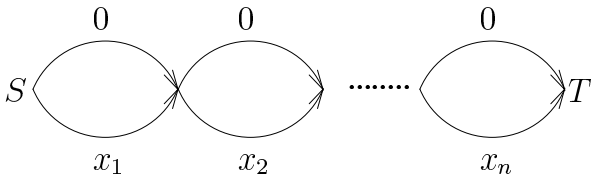


Figure 3: If we can solve for the optimal path with a given start time t in this graph, then we can solve the subset sum problem with integers $x_1, ..., x_n$ and target t.

$K = C(t_{min})$. Consider the graph in Figure 3 with deterministic edge travel times $x_1, ..., x_n$ on the bottom and 0 on the top. Any path P from S to T in this graph has travel time $\sum_{i \in P} x_i$ and cost $C(t' + \sum_{i \in P} x_i)$ if we leave the source at time $t' = t_{min} - t$. Since the cost function $C(t)$ has a unique global minimum at t_{min}, there is a path of cost at most $K = C(t_{min})$ if and only if there is a path with travel time satisfying $t' + \sum_{i \in P} x_i = t_{min}$, *i.e.,* if and only if there is a subset of the x_i's summing exactly to t. □

Remark 2. *Note that Theorem 2 only shows that it is NP-hard to find a simple optimal path. Theorem 3 on the other hand applies to non-simple paths as well since the subset sum problem is NP-complete even if it allows for repetitions (Chvatal 1980).*

Complexity of Stochastic Travel Times

The theorems in the preceding section show that the stochastic routing problem contains instances with deterministic subgraphs that make routing NP-hard, though we do not know whether the class of purely stochastic routing problems (with non-zero variances) is also NP-hard with general cost objectives and travel time distributions. Indeed, there are known problems in scheduling where the scenario with deterministic processing times is NP-hard, while its variant with stochastic, exponentially distributed processing times can be solved in polynomial time (Bruno *et al.* 1981).

It may be difficult to extend Theorems 2 and 3 to the non-zero variance case partly because the integral defining the expected cost will likely not have a closed form for most cost functions. However, we can prove NP-hardness similarly to Theorem 3 for stochastic instances and the function classes we considered earlier, for which we know the form of the expected costs functions.

Recall that under quadratic cost, the expected cost of departing at time t, along a path P with general edge travel time distributions, is $EC_P(t) = \left(t + \sum_{e \in P} \mu_e\right)^2 + \sum_{e \in P} \sigma_e^2$, where μ_e and σ_e are the mean and variance of edge e. Define `Stochastic Cost Routing` to be the problem of deciding whether, for a fixed departure time t, there is a path of expected cost less than K for some constant K.

Corollary 2. `Stochastic Cost Routing` *is NP-hard for quadratic cost with general edge distributions and quadratic+exponential cost with Gamma distributions.*

Proof Sketch: In both cases, the proof reduces to that of Theorem 3, by choosing the means and variances on the top and bottom edges carefully so that the parts of the expected cost function which contain the variances become equal for each path.

For the quadratic cost case, this is straightforward by choosing the same variance σ_i^2 to each pair $i = 1, ..., n$ of top and bottom edges in Figure 3. The quadratic+exponential cost with Gamma travel times takes a little more work since the means and variances are not so well separated in the exponential term. The full proof for this case is given in the appendix. □

Algorithms for the Optimal Path with Fixed Start Time

We have shown that finding simple optimal paths with a given departure time is hard to approximate within a constant factor, as it is similar to finding the longest path. When we remove the restriction to simple paths, we can give a

```
// Initialize paths out of source S with mean 1
path variance Φ(S,0) := 0; predecessor node π(S,0) := S
for each vertex v
   if v is a neighbor of S and μ_sv = 1
      Φ(v,0) := σ_sv; π(v,0) := S
   else
      Φ(v,0) :=null ; π(v,0) :=null

// Fill in the rest of the table
for m = 1 to M
   for each vertex v with neighbors v'
      Φ(v,m) := min_{v'~v} [Φ(v', m − μ_v'v) + σ²_v'v]
      π(v,m) := arg min_{v'~v} [Φ(v', m − μ_v'v) + σ²_v'v]

// Find the lowest cost path from S to T at departure time t
m_opt = argmin_{m∈{0,...,M}} {(t+m)² + Φ(T,m)}
π_opt = π(T, m_opt)
EC_min(t) = (t + m_opt)² + Φ(T, m_opt)
```

Figure 4: Pseudopolynomial Algorithm for Quadratic cost and fixed departure time. M is the upper bound on the mean travel time of a path.

pseudopolynomial dynamic programming algorithm, which depends polynomially on the largest mean travel time of a path M (or equivalently on the maximum expected travel time of a link). For real world applications such as car navigation, M will likely be polynomial in the size of the network, hence this algorithm would be efficient in practice, and significantly better than the previous algorithms based on exhaustive search (Mirchandani & Soroush 1985).

For clarity, we present in detail algorithms for the quadratic and quadratic+exponential cost function families under general travel time distributions, for which we derived simple closed form expressions, however the algorithms can readily extend to general polynomials and polynomials+exponentials.

Quadratic Costs

First, consider the case of quadratic cost, where the expected cost of a path P is $EC_P(t) = \left(t + \sum_{e\in P} \mu_e\right)^2 + \sum_{e\in P} \sigma_e^2$. We are interested in finding the path of smallest $EC_P(t)$, for a fixed t. Denote by $\pi(v,m)$ the predecessor of node v on the path from the source S to v of mean m and smallest variance. Denote by $\Phi(v,m)$ the variance of this path. Then we can find $\pi(v,m)$ and $\Phi(v,m)$ for all v and all $m = 0,1,...,M$ by considering the neighbors v' of v (denoted $v' \sim v$) and choosing the predecessor leading to smallest variance of the path from s to v:

$$\Phi(v,m) = \min_{v'\sim v} \left[\Phi(v', m - \mu_{v'v}) + \sigma^2_{v'v}\right],$$

$$\pi(v,m) = arg\min_{v'\sim v} \left[\Phi(v', m - \mu_{v'v}) + \sigma^2_{v'v}\right],$$

where $\mu_{v'v}$ and $\sigma^2_{v'v}$ denote the mean and variance of the travel time on the link (v', v). Note that by our assumption of independence of the travel times, the variance of each path is given by the sum of variances of the edges, hence suboptimality holds—the path of smallest variance from S to v through v' must be using a path of smallest variance from S to v'.

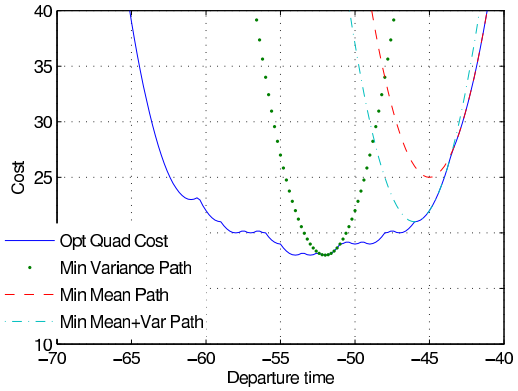
Suppose the maximum degree in the graph is d. Then computing the paths of smallest variance above for each vertex and each possible expected travel time from 0 to M can be done in time $O(Mdn)$. Finally, we find the path from S to destination T by taking the minimum of $(t+m)^2+\Phi(T,m)$ over all $m = 0,1,...,M$, so that the total running time of the algorithm is $O(Mdn)$. If instead of integer, the travel time means are discrete with discrete step ϵ, the running time would be $O(Mdn/\epsilon)$ for small degrees d, or $O(Mn^2/\epsilon)$ in general. The algorithm is summarized in Figure 4.

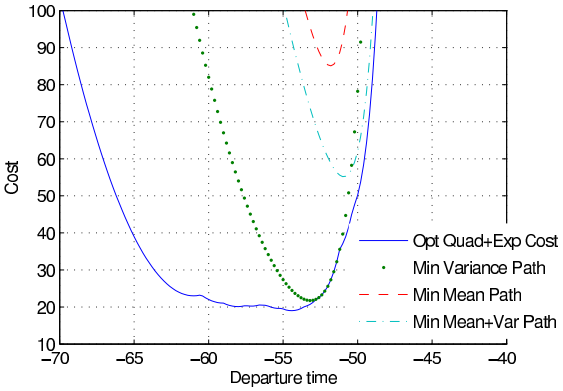
Quadratic+Exponential Costs

We can solve the case of quadratic+exponential penalty similarly, only this time our dynamic programming table would have an extra dimension for possible values of the variance of a path and the table entries would contain the path with smallest exponential term $\prod_{e\in P} E[e^{Y_e}]$. Denote by $\pi(v,m,\sigma^2)$ the predecessor of node v on the path from the source s to v of total mean travel time m, total variance σ^2 and smallest value of $\prod_{e\in P} E[e^{Y_e}]$. Further denote by $\Phi(v,m,\sigma^2)$ the value of $\prod_{e\in P} E[e^{Y_e}]$ on this path. Then as before, once we have computed $\Phi(v,\mu,\sigma^2)$ and $\pi(v,\mu,\sigma^2)$ for all nodes v, path means $\mu \le m-1$ and variances $\sigma^2 = 0,1,...,M$, we can compute $\Phi(v,m,\sigma^2)$ and $\pi(v,m,\sigma^2)$ for all v and $\sigma^2 = 0,1,...,M$ by setting

$$\Phi(v,m,\sigma^2) = \min_{v'\sim v} \left[\Phi(v', m-\mu_{v'v}, \sigma^2-\sigma^2_{v'v}) * E[e^{Y_{v'v}}]\right],$$

$$\pi(v,m,\sigma^2) = arg\min_{v'\sim v} \left[\Phi(v', m-\mu_{v'v}, \sigma^2-\sigma^2_{v'v}) * E[e^{Y_{v'v}}]\right],$$





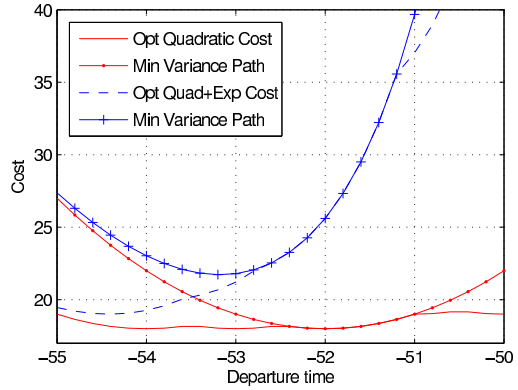


Figure 5: Minimum-cost envelopes for the same network under quadratic (top left) and quadratic+exponential (top right) cost functions. The envelopes are superimposed in the bottom graph in a neighborhood of their global optima.

where $Y_{v'v}$ is the random variable representing the travel time on link (v', v) and we assume that the variance of a path is upper bounded by its expected travel time, so it is at most M. Correctness follows as above by noting that the subpath-optimality property holds for $\prod_{e \in P} E[e^{Y_e}]$. Similarly, we find the path of lowest expected cost from the source to the destination T by taking the minimum over $m = 0, 1, ..., M$ and $\sigma^2 = 0, 1, ..., M$ of $(t+m)^2 + \sigma^2 + \lambda e^{-t}\Phi(T, m, \sigma^2)$. The running time is now $O((M/\epsilon)^2 dn)$ for discrete travel times with discrete step ϵ.

The standard technique of scaling, which turns a pseudopolynomial algorithm into a fully polynomial approximation scheme such as in the knapsack problem (Vazirani 2001) would work here if the ratio of the maximum mean of an edge to the cost of the optimal solution is polynomially bounded and it would fail otherwise. If we do not have a bound on this ratio, we cannot achieve a polynomial approximation scheme, either. Note that the ratio can be arbitrarily large if the optimal path has arbitrarily small variance, say under a quadratic cost function. Even if we lift the cost function by a constant so as to avoid zero values as part of its definition, we may still have a constant optimum cost compared to large mean travel times of edges so we cannot eliminate the dependance of the running times above on the maximum path mean M.

General Polynomial plus Exponential Costs

The above dynamic programming algorithms extend to the case when the expected cost is a general polynomial (plus exponential) with a constant number of terms. Since it is not clear how the various terms trade-off, we would have to keep track of each term individually in a separate dimension of the dynamic programming table, and the running time would scale as M to the power of the number of terms. Scenarios which would fall in this category include general polynomial (plus exponential) cost functions and additive edge distributions, such as Gaussian, Gamma with a fixed width parameter, etc. Under these distributions, the expected cost of path travel time Y would depend only on the distribution of Y as opposed to that of each individual link on the path, and would therefore have a constant number of terms. For example, when the cost $C(Y)$ is a polynomial of degree l, the expected cost $E[C(Y)]$ is a linear combination of the random variable Y's first l moments, as noted by Loui (Loui 1983), and in this case the dynamic programming algorithm would have running time proportional to M^l if each moment is bounded by M.

Experimental Evaluation

We ran the pseudopolynomial algorithms on grid graphs with up to 1600 nodes for the quadratic objective and up to 100 nodes for quadratic+exponential objective. The former graph instances can be viewed as the problem of navigating from one corner of Manhattan to another; the latter as finding a path around a city through a highway network. Runtimes were typically a few seconds, while memory turned out to be limiting factor: In the case of quadratic objective the dynamic programming table is two-dimensional, and in the quadratic+exponential objective the table is three-dimensional, *i.e.,* cubic in the size of the graph. Given the memory constraint we had the largest edge mean set to 10 on the graphs with 1600 nodes and 4 on the graphs with 100 nodes. The edge means and variances were generated uniformly at random. The memory usage of the algorithms was not optimized; it could be made an order of magnitude smaller (linear for $C(t) = t^2$, quadratic for

$C(t) = t^2 + \lambda e^{kt}$) if one only wanted to compute the objective function values without outputting the actual paths.

For the sake of comparison, Figure 5 shows the resulting optimal cost for the same graph with 100 nodes under both quadratic and quadratic+exponential objectives with Gamma distributed travel times. Recall that the optimal cost envelope for a graph under a given objective is the infimum of the cost functions of each path in the graph. Each plot on the top shows the min-cost envelope and the expected cost function of three paths—those having smallest variance, smallest mean, and smallest mean+variance. As predicted, the path with smallest variance yields the globally optimal cost in the quadratic case but this is not necessarily true in the quadratic+exponential case.

We note that the basic quadratic+exponential objective

$$\Big(t + \sum_{i=1}^{r} a_i b_i\Big)^2 + \sum_{i=1}^{r} a_i b_i^2 + e^t \Big[\prod_{i=1}^{r} (1 - b_i)^{-a_i}\Big],$$

tends to be dominated by the quadratic term near the function's minimum so that its plot is almost identical to the plot of the quadratic objective case for the interval of departure times around the global minimum. However, a bigger positive coefficient in front of the exponential term (featured at the top right of Figure 5) balances the quadratic and exponential influence and illustrates a situation where the smallest variance path is not globally optimal and the smallest mean or mean+variance paths are not even locally optimal, *i.e.,* do not participate in the min-cost envelope.

The third plot in the Figure superimposes the two different objectives and zooms into their global minima. The plot demonstrates clearly the qualitative difference between the two objective costs, not only in the fact that the global optimum is attained on different paths but also in that different paths may be locally optimal at the same fixed departure time. For example, at departure 51 minutes (units) before the deadline a quadratic objective would recommend the smallest variance path while the quadratic+exponential objective would recommend some other path. We see similar differences of recommendation at time a little over 52.5 minutes before the deadline. Naturally, since the exponential term assigns a more severe penalty for being late, the quadratic+exponential objective recommends an earlier globally optimal departure time.

Monotone Increasing Costs

The optimal path problem becomes significantly easier if we consider some natural monotone increasing costs, such as linear and exponential, for which the global cost is separable into edge costs. As noted above linear cost with a given start time translates to minimizing the expected travel time, a basically uninteresting quantity in this stochastic setting. An exponential cost $C(t) = e^{kt}$ on the other hand, *is* interesting because it gives rise to the expected path cost

$$EC_P(t) = e^{kt} \prod_{e \in P} E[e^{kY_e}].$$

where Y_e is the travel-time random variable at edge e. We make this separable by moving into the log domain, where finding the path of lowest expected cost starting from the source at time t turns out to be equivalent to finding the shortest path on the same graph with edge weights set to the cumulant-generating function $K(k) = \log\big(E[e^{kY_e}]\big)$. The cumulant-generating function is a series sum over cumulants that is dominated by the lowest central moments of the distribution; for many distributions it is effectively a weighted sum of mean and variance, a common objective in portfolio selection and mean-risk analysis in general. For gamma-distributed travel times, $K(k) = a \log(1/(1-kb))$; compare with the more familiar $K(k) = k\mu + k^2\sigma^2/2$ for normally distributed variables.

Discussion

We have obtained complexity results and algorithms for two problems of route planning under uncertainty in a decision-theoretic framework: planning an optimal path for a given departure time as well as planning an optimal departure time. Path and start time are jointly optimizable because there are penalties for both late and early arrivals. Surprisingly, some instances of the joint optimization are reducible to classic shortest path algorithms, while the seemingly easier problem of finding an optimal path for a given start time is NP-hard, and it is hard even to approximate for simple paths. For fixed start times and non-simple paths, we have presented pseudopolynomial algorithms, which are polynomial in the number of nodes in the graph and the largest mean travel time of an edge. These will likely perform very well in practice since the average travel times on edges in real world networks are relatively small. We would like to extend the full analysis to the case of dynamic networks, where both topology and distributions change in time. Other open questions include comparing the optimal non-adaptive to the optimal adaptive solution, under both static and dynamic travel time distributions, and modeling correlations between the edge travel times.

Acknowledgment

We thank MohammadTaghi Hajiaghayi, Daniel Nikovski, Rahul Sami and Andreas Schultz for valuable suggestions. The first and third authors are supported in part by NSF grants ANI-0225660 and ITR-0219018. The first author is also grateful for support from Mitsubishi Electric Research Labs.

References

Bertsekas, D. P. and Tsitsiklis, J. N. 1991. An analysis of stochastic shortest path problems. *Math. Oper. Res.* **16** (3): 580-595.

Boyan, J. and Mitzenmacher, M. 2001. Improved Results for Route Planning in Stochastic Transportation Networks. *ACM-SIAM Symposium on Discrete Algorithms.*

Bruno, J.L.; Downey, P.J.; and Fredrickson, G.N. 1981. Sequencing tasks with exponential service time to minimize the expected flowtime or makespan. *Journal of the ACM*, **28**: 100-113.

Chabini, I. 2002. Algorithms for k-Shortest Paths and Other Routing Problems in Time-Dependent Networks. *Transportation Research* Part B: Methodological.

Chvatal, V. 1980. Hard Knapsack Problems. *Operations Research*, **28**: 1402-1411.

Fan, Y.; Kalaba, R.; and Moore, II, J. E. Arriving on Time. *Journal of Optimization Theory and Applications*, forthcoming.

Fu, L. and Rilett, L. R. 1998. Expected Shortest Paths in Dynamic and Stochastic Traffic Networks. *Transportation Research*, Part B, **32**: 499-516.

Gao, S. and Chabini, I. 2002. Optimal Routing Policy Problems in Stochastic Time-Dependent Networks. *Proceedings of the IEEE 5th International Conference on Intelligent, Transportation Systems*, Singapore, pp. 549-559.

Hall, R. W. 1986. The Fastest Path Through a Network with Random Time-Dependent Travel Times, *Transportation Science*, **20**: 182-188.

Karger, D.; Motwani, R.; and Ramkumar, G. D. S. 1997. On approximating the longest path in a graph. *Algorithmica*, **18**: 82-98.

Loui, R. P. 1983. Optimal Paths in Graphs with Stochastic or Multidimentional Weights. *Communications of the ACM*, **26**: 670-676.

Miller-Hooks, E. D. and Mahmassani, H. S. 2000. Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks. *Transportation Science*, **34**: 198-215.

Mirchandani, P. and Soroush, H. 1985. Optimal paths in probabilistic networks: a case with temporary preferences. *Computers and Operations Research,* **12** (4): 365-381.

Nikolova, E. 2005. Stochastic Shortest Paths. Manuscript.

Papadimitriou, C.H. and Yannakakis, M. 1991. Shortest paths without a map. *Theoretical Computer Science* **84**: 127-150.

Papadimitriou, C. H. and Steiglitz, K. 1998. *Combinatorial optimization: Algorithms and complexity.* Dover Publications, Inc., N.Y.

Vazirani, V. V. 2001. *Approximation Algorithms.* Springer-Verlag Berlin Heidelberg.

Appendix

Corollary. 2. `Stochastic Cost Routing` *is NP-hard for quadratic+exponential cost with Gamma distributions.*

Proof: Recall that when the penalty of arriving at the destination t minutes before the deadline is $C(t) = t^2 + e^t$, the expected cost of a path at start time t before the deadline is given by

$$\Big(t + \sum_{i=1}^{n} a_i b_i\Big)^2 + \sum_{i=1}^{n} a_i b_i^2 + \Big[\prod_{i=1}^{n} (1-b_i)^{-a_i}\Big] e^t.$$

Consider the chain graph in Figure 3 with distributions $\Gamma(a_i, b_i)$ on the bottom edges and $\Gamma(a_i', b_i')$ on the top ones. It suffices to show that there exist positive a_i, b_i, a_i', b_i' such that

$$\begin{aligned} a_i b_i &= z_i + q_i \\ a_i' b_i' &= q_i \\ a_i b_i^2 &= a_i' b_i'^2 \\ (1-b_i)^{-a_i} &= (1-b_i')^{-a_i'} \end{aligned}$$

for all $i = 1, ..., n$. Note also because of the last equation, we need to restrict the domain of b_i and b_i' to $(0, 1)$. From the first two equations, $a_i = \frac{z_i+q_i}{b_i}$ and $a_i' = \frac{q_i}{b_i'}$. If $z_i + q_i, q_i, b_i, b_i'$ are all positive, then a_i and a_i' will be positive as well. Recall also that q_i is a positive constant of our choice such that $z_i + q_i > 0$. Henceforth, we drop the indexes to ease notation. Then, it suffices to show that there exist $b, b' \in (0, 1)$ such that

$$\begin{aligned} b(z+q) &= b'q \\ (1-b)^{(z+q)/b} &= (1-b')^{q/b'}. \end{aligned}$$

From the first equation in (7), $b' = b\frac{z+q}{q} = b(1+\frac{z}{q}) = bk$ and recall that q is a constant of our choice which we can set as high as we want. Substituting in the second equation in (7), we get

$$(1-b)^{1/b} = (1-kb)^{1/(k^2 b)}. \tag{7}$$

If $z > 0$, we can pick for example $k = 1/0.9 > 1$ and this immediately yields a solution $b = .676948$. With this, $b' = b/0.9 = .752165$ and we can find the corresponding a and a'. Note that this b and b' will work for all pairs of links for which $z > 0$. Similarly, if $z < 0$, we have $k < 1$ in Equation (7) and picking $k = 0.9$ gives the same values for b and b' but reversed: $b = .752165, b' = .676948$. □