all but a finite number converge to local maximum points in $\mathcal{S} \cap \partial A_{x_0}$. The finite number of exceptions converge to saddle points in $\mathcal{S} \cap \partial A_{x_0}$, exactly one to each saddle point.

The following discussion is purely formal, since our purpose here is simply to suggest a numerical scheme. According to the discussion in the previous paragraph, if we start a particle off from $x_0$ in a random direction selected according to the uniform distribution and let it proceed along a steepest ascent path, then it has probability 1 of ending up at a local maximum point. Suppose we could compute the probability distribution $\mu(dy)$ on $\partial \mathcal{S}$ that describes the distribution of the particle in the limit as $t \to \infty$. If $x_i$, $i = 1, 2, \ldots, n$, denote the local maxima in $\partial A_{x_0}$, then $\mu(\{x_1, \ldots, x_n\}) = 1$, and obviously $\mu(\partial A_{x_0} \setminus \{x_1, \ldots, x_n\}) = 0$. In particular, for at least one $i \in$
\{1, \ldots, n\} we must have \(\mu(\{x_i\}) > 0\). Thus, the distribution \(\mu\) allows the identification of at least one local maximum point in \(\partial A_{x_0}\).

A scheme for computing an approximation to \(\mu\) on \(\mathcal{G}^h\) can be developed that is similar in spirit to our reconstruction algorithms. However, it is much simpler. Let \(V^h(\cdot)\) be the reconstruction obtained by the algorithm using only the value \(z(x_0)\) as the initial boundary datum. Let \(\xi^h_i\) be a Markov chain defined on \(\mathcal{G}^h\) such that \(\xi^h_0 = x_0\). The transition probabilities are defined so that the discrete time evolution of \(\xi^h_i\) approximates via \(V^h(\cdot)\) a steepest ascent curve on \(z(\cdot)\). A simple choice for these probabilities is the following. At \(x_0\), let \(p^h(x_0, x_0 \pm h(1,0)) = p^h(x_0, x_0 \pm h(0,1)) = 1/4\), giving an equal probability of emerging from \(x_0\) in any of the compass directions. We can also allow diagonal transitions to give a better approximation to the uniform distribution. Now consider \(y \neq x_0\) and define

\[
\begin{align*}
  s_1 &= \text{sign}[V^h(y + h(1,0)) - V^h(y - h(1,0))] , \\
  s_2 &= \text{sign}[V^h(y + h(0,1)) - V^h(y - h(0,1))] .
\end{align*}
\]

Also, let \(v_1\) and \(v_2\) be the largest values from the sets

\[
\{V^h(y + h(1,0)), V^h(y - h(1,0))\}
\]

and

\[
\{V^h(y + h(0,1)), V^h(y - h(0,1))\},
\]

respectively, and

\[
d_{y,i} = \max(0, h^{-1}(v_i - V^h(y))),
\]

for \(i = 1, 2\). The \(d_{y,i}\) are essentially forward or backward derivatives, depending on which direction gives the steepest ascent. For \(d_{y,1} = d_{y,2} = 0\), define \(p^h(y, y) = 1\). Otherwise, define

\[
p^h(y, y + s_1 h(1,0)) = \frac{d_{y,1}}{d_{y,1} + d_{y,2}} ,
\]

\[
p^h(y, y + s_2 h(0,1)) = \frac{d_{y,2}}{d_{y,1} + d_{y,2}},
\]

with all other probabilities zero.

Let \(\{\xi^h_i, i \in \mathbb{N}\}\) be the Markov chain that is defined by the construction outlined previously. For \(y \in \mathcal{G}^h\) set

\[
P^h_n(y) = P[\xi^h_n = y].
\]

The Markov property gives the recursion

\[
P^h_{n+1}(y) = \sum_x p^h(x, y) P^h_n(x).
\]

We would like to use \(\lim_{n \to \infty} P^h_n(y)\) as the approximation to \(\mu\). However, because the reconstruction \(V^h(\cdot)\) is computed with \(g(y) < B\) only at \(y = x_0\), the infimal expected cost \(V^h(y)\) for \(y \in \mathcal{G} \setminus A_{x_0}\) will approximately satisfy
$V^h(y) \geq z(y)$ and also $V^h(y) \geq \inf_{y' \in \partial A_{x_0}} V^h(y')$. Since the evolution of the chain approximates steepest ascent on $V^h(\cdot)$ and because of the discretization, it is likely that $\xi^h_i$ will evolve outside a neighborhood of the set $A_{x_0}$ and that the support of $P^h_n(y)$ as defined will eventually extend beyond $A_{x_0}$. A direct use of this distribution therefore gives a poor approximation to $\mu$ as $n \to \infty$, and may result in spurious concentrations of probability at singular points in $\mathcal{F} \setminus A_{x_0}$, leading to the false identification of these points as local maxima.

We avoid this problem by modifying the transition probabilities. Since the steepest ascent paths converge on the local maxima points in $\partial A_{x_0}$, $\xi^h_i$ should pass near a point in the set $\{x_1, \ldots, x_n\}$ with probability that approaches 1 as $h \to 0$. Let $m > 0$ be such that $m < (1/2)\min\{\|y - y'\|: y, y' \in \mathcal{F}\}$. We modify the probabilities so that the chain is stopped as soon as it enters the set $\{y': \|y' - y\| < m, y \in \mathcal{F} \setminus x_0\}$, that is, all points $x \in \mathcal{F}^h \cap \{y': \|y' - y\| < m, y \in \mathcal{F} \setminus x_0\}$ are taken to be absorbing, so that $p^h(x, x) = 1$.

With this modification, the strength of $P^h_n(y)$ at singular points outside $\partial A_{x_0}$ should be sharply reduced. Let $P^h(y) = \lim_{n \to \infty} P^h_n(y)$. This limit should be well defined at all points and is obviously well defined because of monotonicity for $x \in \mathcal{F}^h \cap \{y': \|y' - y\| < m, y \in \mathcal{F} \setminus x_0\}$. Since the structure of the recursion (4.1) is similar to that of the reconstruction algorithm itself, the number of iterations required for convergence to $P^h$ is expected to be similar to the number of iterations used in the convergence to $V^h$. Once an approximation to $P^h(y)$ has been computed, those singular points near which this approximation has a large value can be identified as local maxima.

In our actual experiments, a variation of this procedure was used. Rather than an approximation to the probability distribution $\mu$, we consider the sum

$$Q^h_n(y) = \sum_{i=0}^n P^h_i(y).$$

The Markov chain $\xi^h_i$ is the same as before, except that at $y$ such that $d_{y,1} = d_{y,2} = 0$, we define the transition probability to be $p^h(y, \mathcal{F}) = 1$, where $\mathcal{F}$ is a "fictitious" absorbing point adjoined to $\mathcal{F}^h$. We also define $p^h(y, \mathcal{F}) = 1$ for $y \not\in \mathcal{F}^h$. At points $y \in \mathcal{F}^h \setminus \{x_0\}$, $Q^h_n(y)$ satisfies the same recursion (4.1) as before, while $Q^h_n(x_0) = 1$ for all $n$. The advantage of working with the summed distribution is that it is monotonically nondecreasing in $n$ for all $y$. Define $Q^h(y) = \lim_{n \to \infty} Q^h_n(y)$. As before, this limit should exist at all points in $\mathcal{F}^h$ because of the new definition of the chain with an absorbing state and since, apart from the case $d_{y,1} = d_{y,2} = 0$, transitions are uphill with respect to $V^h(\cdot)$. Due to the monotonicity, $Q^h(y)$ can be calculated using a Gauss–Seidel iteration similar to that of the reconstruction algorithm, where the ordering of the states is changed after each iteration. As before, the number of iterations necessary for convergence is expected to be the same as for the original reconstruction of $V^h(\cdot)$. Also, it is again necessary to modify the transition probabilities so as to reduce the probability of the process exiting a neighborhood of the set $A_{x_0}$. In this approach, local maxima in $z(\cdot)$
are identified by looking for a singular point near a local maximum of \( \overline{Q}^h(y) \). More precisely, parameters \( \theta_1 > 0 \) and \( \theta_2 > 0 \) are chosen and we identify a singular point as a local maximum if there is a local maximum of \( Q^h(y) \) with value greater than \( \theta_1 \) within \( \theta_2 \) of the singular point.

The algorithm as outlined still requires that at least one singular point be identified as a local maximum or minimum point. Since the height \( z(\cdot) \) is ambiguous up to the addition of an overall constant, the height of this point is not needed to start the procedure. However, for any singular point \( x_0 \), it is often possible to determine a priori whether it is a local maximum, minimum or saddle point, as we now discuss.

Let \( I'(\cdot) \) be an arbitrary function in some region around \( x_0 \) such that \( I'(x_0) = 1 \) and \( I'(y) \in (0, 1) \), \( y \neq x_0 \). Suppose \( g(x_0) = z(x_0) \) and \( g(x) = B \) for all \( x \neq x_0 \), where \( B \) is an upper bound for \( z(\cdot) \) on \( \mathcal{F} \). The corresponding height function \( V(x) \), given by the control representation of (2.9), will in general not be \( C^1 \). In the typical case, \( V_\lambda(\cdot) \) will have "line" discontinuities, and the \( V^h(\cdot) \) reconstructed by our algorithm will approximately reproduce these discontinuities. For \( I(\cdot) \), our experiments show in general that if \( V^h(\cdot) \) is computed using a singular point \( x_0 \) under the incorrect assumption that it is a local minimum (or maximum), then the discrete "derivatives," for example, \( \Delta V^h_{x_1}(y) \equiv V^h(y + h(0, 1)) - V^h(y) \), will display abrupt changes that can be interpreted as "discontinuities." These can be easily detected. We have used the occurrence of these "discontinuities" both to determine an initial local minimum point and to check the assignments of singular points as local minima, maxima or saddle points, as determined by our iterative procedure.

We have applied this procedure to the surface of Figure 4, under the oblique lighting used previously. In order to reduce the effects of discretization, the surface was first scaled by a factor 0.5, so that the height range was approximately 25 units compared to a range for \( x_1 \) and \( x_2 \) of 128 units. Using no boundary data other than \( I(\cdot) \), the surface shown in Figure 14 was reconstructed by this procedure. The reconstruction took four cycles: that is, a surface was first reconstructed using a single local minimum, then again

Fig. 14.
using the recovered local maxima, again using local minima and the result of reconstructing a final time using the final recovered local maxima is displayed in the figure. The reconstruction took less than 30 seconds of CPU time on a DEC 5000 workstation.

Figure 15 illustrates the reconstruction error, the magnitude of the difference between the original surface height and that of the reconstruction. The reconstruction is good except near the edges of the image. This is due in part to the fact that, as in the previous subsection, the surface does not satisfy Assumption 2.1 for the entire set \( \mathcal{D} \) (or its analog for reconstruction based on local maxima). The average reconstruction error in the interior of the image with \( 20 \leq x_2 \leq 105 \) is 0.5 units, or about 2\% of the height range. Figures 16 and 17 show the surface and its reconstruction, respectively, over this region, illustrating the accuracy of the reconstruction there.

A second surface (illuminated as before) and its reconstruction are displayed in Figures 18 and 19, respectively. The reconstruction error is shown in Figure 20. Only three cycles, again starting from a local minimum singular point, were enough to give this reconstruction. The surface also was reconstructed starting from a different local minimum singular point, with comparably good results. The average height error in the interior of the image is 1 unit, in comparison to the overall height range for this surface of 44 units. As before, the accuracy of reconstruction is on the order of 2\%.

Figure 21 depicts the summed probability distribution \( Q^h \) generated from Figure 18 after one cycle, together with the singular points in the image of this surface (shown as dark isolated clusters). Larger magnitudes for \( Q^h \) are indicated by increasing darkness. The distribution \( Q^h \) achieves its maximum at the central local minimum singular point used to generate this distribution. The figure clearly shows the dominant evolution of the chain toward
four other singular points, which in fact correspond to local maxima of the height function. There are local maxima of $Q^h$ near each of the four singular points, as indicated by the shaded patches near these points. A fifth point has a shaded patch nearby, but is not identified as a maximum since it does not satisfy the threshold rule.

We have also studied the effect of adding noise to the image of Figure 18. The noise added had a uniform distribution on the interval $[-0.1, 0.1]$, and was independent for different lattice points. Since the maximum image intensity is only $I = 1$, this is a large noise of $\pm 10\%$. The reconstruction based on the image with added noise is shown in Figure 22. The surface shown was generated using the computed local minima, and required three cycles. Although there are large errors in some parts of the image, the reconstruction is still good over much of the image. The error in the height is displayed in Figure 23, where saturated white represents a height error of 3. The error is less than 3 units over most of the image. In the region of the image with $127 > x_{1,2} > 40$, the mean height error is just 1.6. This represents a surprising immunity to the large image noise.
In Figure 24 the error is shown for a different reconstruction from the same noisy image with the same scale as before. In this case, the surface was generated from the local maxima after just two cycles. As expected, near the boundary of the image, the region of accurate reconstruction for the maxima-based method is complementary to that of the minima-based method. Since the image boundary does not respect Assumption 2.1 (for either method), the maximum-based method does better at those points near the boundary where the steepest descent direction is outward, while the minima-based method does better where this direction is inward. Together, the two methods give reconstruction with error less than 3 units over most of the image.

5. Proofs of the theorems and propositions. In this section we give the proofs of the main theoretical results stated in this paper, including the
representation and convergence theorems, Theorems 2.1 and 2.3. We begin
with the representation theorem, which states that in certain regions the
minimal cost of the control problem specified in Section 2 equals the unknown
function \( f \).

**Proof of Theorem 2.1.** We first show that \( V(x) \geq f(x) \). Let \( u(\cdot) \) be any
admissible control and define

\[
\phi(t) = x + \int_0^t u(s) \, ds, \quad \tau = \inf\{ t : \phi(t) \in \partial D \cap A \}.
\]

Since \( L \) is the Legendre transform of \( H \) and since \( H(x, f_x(x)) = 0 \) for \( x \in \partial D \),
\[
0 \geq -\langle f_x(x), \beta \rangle - L(x, \beta)
\]
for all \( \beta \in \mathbb{R}^2 \), and in particular
\[
-\langle f_x(\phi(t)), u(t) \rangle \leq L(\phi(t), u(t))
\]
for \( t \in [0, \rho \wedge \tau] \). (Recall that \( \rho \) is the controlled stopping time.) This implies that
\[
-f(\phi(\rho \wedge \tau)) + f(x) = -\int_0^{\rho \wedge \tau} \langle f_x(\phi(t)), u(t) \rangle \, dt
\]
\[
\leq \int_0^{\rho \wedge \tau} L(\phi(t), u(t)) \, dt,
\]
and thus for any admissible control and any \( \rho \in [0, \infty) \),
\[
\int_0^{\rho \wedge \tau} L(\phi(t), u(t)) \, dt + f(\phi(\rho \wedge \tau)) \geq f(x).
\]

Since \( g(\phi(\rho \wedge \tau)) \geq f(\phi(\rho \wedge \tau)) \), we obtain \( V(x) \geq f(x) \).

Next we show \( V(x) \leq f(x) \). In order to do so we will verify that for each
\( \varepsilon > 0 \) there exists a control \( u(\cdot) \) such that for \( \phi \) and \( \tau \) defined by (5.1) we have \( \tau < \infty \), and

\[
\int_0^\tau L(\phi(t), u(t)) \, dt + g(\phi(\tau)) \leq f(x) + \varepsilon.
\]
Recall that \( \mathcal{Z}(x) = \{ (u_1, u_2) : |u_1|^2 + |u_2 + \gamma_2|^2 \leq 1 \} \) and \( L(x, 0) = 0 \) for \( x \in \mathcal{S} \). Let \( \mathcal{S}_c \) be a maximal smoothly connected component of \( \mathcal{S} \). We first show that any two points near \( \mathcal{S}_c \) can be connected by a piecewise continuous control with low cost. For any two points \( x, y \) in \( \mathcal{S}_c \), there exists a uniformly bounded control \( u(t) \) and a time \( t^* < \infty \) such that if \( \phi(t) = x + \int_0^t u(s) \, ds \), then \( \phi(t) \in \mathcal{S}_c \) for \( t \in [0, t^*] \) and \( y = \phi(t^*) \). Define a new control \( u_\lambda(t) = \lambda u(t\lambda) \), where \( \lambda > 0 \), and let \( \phi_\lambda(t) = \phi(t\lambda) \) be the corresponding path. Since

\[
\frac{L(x, u)}{\|u\|} \to 0 \quad \text{as} \quad \|u\| \to 0
\]
for $x$ such that $I(x) = 1$, we can choose $\lambda$ such that

$$
\int_0^{t^*} L(\phi_\lambda(t), u_\lambda(t)) \, dt = \int_0^{t^*} \frac{L(\phi(t), \lambda u(t))}{\lambda} \, dt \leq \varepsilon/3.
$$

Since $L(\cdot, \cdot)$ is continuous on an open neighborhood of $\mathcal{S}_C \times \{0\}$, we can also assume $u_\lambda$ is continuous. Further, since $|\gamma_2| < 1$, $U(x)$ contains an open neighborhood of the origin for all $x \in \mathcal{S}_C$. Hence there exists $a > 0$ such that for any component $\mathcal{S}_C$ as above and $x$ such that $d(x, \mathcal{S}_C) \leq a$, we have the following. Let $y$ be the point in $\mathcal{S}_C$ closest to $x$. Then there exists a time $t_a \in [0, \infty)$, a constant control $u(t) = (y - x)/t_a$ and corresponding path $\phi(t) = x + \int_0^t u(s) \, ds$, such that $\phi(t_a) = y$ and $\int_0^{t_a} L(\phi(t), u(t)) \, dt \leq \varepsilon/3$. Finally, this shows that for any $\mathcal{S}_C$, and $x, y$ such that $d(x, \mathcal{S}_C) \leq a$ and $d(y, \mathcal{S}_C) \leq a$, 

---

**FIG. 24.**
there exists a piecewise continuous control $\bar{u}_{xy}(t)$ and time $\sigma_{xy} \in [0, \infty)$ such that for the corresponding path $\phi_{xy}(t)$ we have

$$\phi_{xy}(0) = x, \quad \phi_{xy}(\sigma_{xy}) = y$$

and

$$\int_0^{\sigma_{xy}} L(\phi_{xy}(t), \bar{u}_{xy}(t))\, dt \leq \varepsilon.$$ 

Since $f$ is constant on $\mathcal{S}_C$, we can assume (by choosing $\alpha > 0$ smaller if need be) that $|f(x) - f(y)| \leq \varepsilon$.

We now construct the control that satisfies (5.2). If $x \in \mathcal{S}_C$ and $\mathcal{S}_C \subset \mathcal{A}$, then we simply take $t = 0$ and are done. There are then two remaining cases: (1) $x$ is contained in some $\mathcal{S}_C$ with $\mathcal{S}_C \cap \mathcal{A} = \emptyset$ or (2) $x \notin \mathcal{S}$. If case (1) holds, then $\mathcal{S}_C$ is either a set of local maxima or saddle points, which implies the existence of a point $y$ such that $f(y) < f(x)$ and $d(y, \mathcal{S}_C) \leq \alpha$. Since Assumption 2.1 implies $\mathcal{S} \subset \mathcal{G}^0$, we can assume that $y \in \mathcal{G}$. In this case we will set $u(t) = \bar{u}_{xy}(t)$ for $t \in [0, \sigma_{xy})$.

Next consider the definition of the control for $t \geq \sigma_{xy}$. For $c > 0$ let $b = \inf\{L(x, u) : x \in \mathcal{G}, d(x, \mathcal{S}) > c, u \in \mathbb{R}^2\}$. The continuity of $I(\cdot)$ and the fact that $I(x) < 1$ for $x \notin \mathcal{S}$ implies $b > 0$. Consider any solution (there may be more than one) to

$$\dot{\phi}(t) = \bar{u}(\phi(t)), \quad \phi(0) = y. \tag{5.3}$$

According to (2.7), for any $t$ such that $\phi(t) \notin \mathcal{S}$ and $d(\phi(t), \mathcal{S}) > c$,

$$\frac{d}{dt} f(\phi(t)) = \langle f_x(\phi(t)), \bar{u}(\phi(t)) \rangle$$

$$= -L(\phi(t), \bar{u}(\phi(t)))$$

$$\leq -b. \tag{5.4}$$

Assumption 2.1 implies $\phi(t)$ cannot exit $\mathcal{G}$. Thus, since $f(x)$ is bounded on $\mathcal{G}$, (5.4) implies that $\phi(t)$ must enter the set $\{x : d(x, \mathcal{S}) \leq c\}$ in finite time, for any $c > 0$. If $\phi(t) \in \mathcal{S}$ for some $t < \infty$ we define $\eta_y = \inf\{t : \phi(t) \in \mathcal{S}\}$ and $w = \phi(\eta_y)$. Otherwise, let $t_i$ be any sequence tending to $\infty$ as $i \to \infty$. Since $\mathcal{S}$ is compact, we can extract a subsequence (again labeled by $i$) such that $\phi(t_i) \to v$ for some $v \in \mathcal{S}$. Let $i^*$ be large enough that $\|\phi(t_i) - v\| \leq \alpha$. Since $f(\phi(t_i)) - f(v)$, we have $f(\phi(t_i^*)) > f(v)$. For this case we define $\eta_y = t_i^*$ and $w = \phi(\eta_y)$.

Integrating (5.4) gives

$$f(y) - f(w) = \int_0^{\eta_y} L(\phi(t), \bar{u}(\phi(t)))\, dt.$$ 

We then define the control $u(t)$ to be used for $t \in [\sigma_{xy}, \sigma_{xy} + \eta_y)$ to be $\bar{u}(\phi(t - \sigma_{xy}))$.

We now consider the point $w$. We first examine the case in which the solution to (5.3) does not enter $\mathcal{S}$ in finite time. Since $\|w - \bar{v}\| \leq \alpha$, $\bar{u}_{wv}(t)$ gives a control such that the application of this control moves $\phi(\cdot)$ from $w$ to
\[ v \text{ with accumulated running cost less than or equal to } \varepsilon. \text{ We define } u(t) = \tilde{u}_{wv}(t - (\eta_y + \sigma_{xy})), \text{ for } t \in [\sigma_{xy} + \eta_y, \sigma_{xy} + \eta_y + \sigma_{wv}). \text{ If the solution to (5.3) reached } \mathcal{S} \text{ in finite time, we define } w = v \text{ and } \sigma_{wv} = 0. \text{ Let } \sigma = \sigma_{xy} + \eta_y + \sigma_{wv}.

Let us summarize the results of this construction. Given any point \( x \in \mathcal{S} \) that is not a local minimum, we have constructed a piecewise continuous control \( u(\cdot) \) and \( \sigma < \infty \) such that if \( \phi(t) = x + \int_0^t u(s) \, ds \), then

\[
\begin{align*}
    f(x) - f(\phi(\sigma)) &= f(x) - f(y) + f(y) - f(w) + f(w) - f(v) \\
    &\geq \int_{\sigma_{xy}}^{\sigma_{xy} + \eta_y} L(\phi(t), u(t)) \, dt \\
    &\geq -2\varepsilon + \int_0^\sigma L(\phi(t), u(t)) \, dt.
\end{align*}
\]

We have also shown that \( f(x) > f(v) = f(\phi(\sigma)) \), \( \phi(\sigma) \in \mathcal{S} \). Thus, either the component \( \mathcal{S}_C \) containing \( \phi(\sigma) \) satisfies \( \mathcal{S}_C \cap \mathcal{M} \neq \emptyset \) and we are done, or we are back into case 1 and can repeat the procedure. Let \( K \) be the number of disjoint compact connected sets that comprise \( \mathcal{S} \). Then the strict inequality \( f(x) > f(\phi(\sigma)) \) and the fact that \( f(\cdot) \) is constant on each \( \mathcal{S}_C \) imply the procedure can be repeated no more than \( K \) times before reaching some \( \mathcal{S}_C \) contained in \( \mathcal{M} \). If case 2 holds, we can use the same procedure, save that the very first step is omitted. Thus, in general, we have exhibited a control \( u(\cdot) \) such that

\[
\int_0^\tau L(\phi(t), u(t)) \, dt + g(\phi(\tau)) \leq f(x) + (2K + 1)\varepsilon.
\]

Since \( \varepsilon > 0 \) is arbitrary, the theorem is proved. \( \square \)

The following lemma is used in the proof of Theorem 2.3.

**Lemma 5.1.** Let Assumption 2.1 hold, and let \( L(x, \beta) \) and \( \mathcal{V}(x) \) be defined as in Section 2. Assume that \( I \in (0, 1] \) on \( \mathcal{S} \), let \( u \) be the piecewise continuous control constructed in the proof of Theorem 2.1 that gives the inequality in (5.2) and let \( \phi \) and \( \tau \) be as defined as in (5.1). Then the following conclusions hold.

1. The function \( L(x, \cdot) \) satisfies the following property: Given \( B < \infty \), there exists \( \delta > 0 \) such that \( \|\nabla_u L(x, u)\| \leq B \) implies \( d(u, \mathcal{V}(x)^c) \geq \delta \), where the superscript \( c \) denotes complement.
2. The path \( \phi(t) \) remains in \( \mathcal{S} \) for all \( t \in [0, \tau] \) and there exists \( \delta > 0 \) such that \( d(u(t), \mathcal{V}(\phi(t))^c) \geq \delta \) for all \( t \in [0, \tau] \).

The assumption \( I(x) > 0 \) implies that \( \mathcal{V}(x) \) has a nonempty interior and is clearly needed for part 2 of the lemma. Although we must also assume this condition in the convergence theorem (Theorem 2.3), by working with a sequence of domains (see the remark after the statement of Theorem 2.1), we can extend the convergence theorem to cover the case \( I(x) \in [0, 1] \).
PROOF. If part 1 were not true, then there would exist a sequence $u_i \to u \in \partial \mathcal{Z}(x)$ for which $\|\nabla_u L(x, u_i)\| \leq B$ (i.e., $L(x, \cdot)$ would not be essentially smooth in the terminology of Rockafellar [26]). This can be contradicted either via a direct calculation or the general result given in Theorem 26.3 of [26].

Since $f(x)\in\mathfrak{B}$, (2.6), (2.7) and part 1 of the lemma imply that for each $x \in \mathcal{F}$ there exists $\delta > 0$ such that $d(\tilde{u}(x), \mathcal{Z}(x)^\circ) \geq \delta$. Since $\tilde{u}(x)$ and $\mathcal{Z}(x)$ are continuous on $\mathcal{F}$, we can assume that $\delta$ is independent of $x$ for $x \in \mathcal{F}$.

It is also straightforward to show that we can take $\delta > 0$ small enough that $d(\tilde{u}_x(t), \mathcal{Z}(\phi(t))) \geq \delta$ for all $t \in [0, \sigma_{xy}]$ whenever both $d(x, \mathcal{S}_c)$ and $d(y, \mathcal{S}_c)$ are sufficiently small, where $\tilde{u}_x$ and $\sigma_{xy}$ are as defined in the proof of Theorem 2.1 and $\phi$ is the associated controlled path. This implies part 2 of the lemma. □

We next give the proof of the convergence theorem.

PROOF OF THEOREM 2.3. The basis for the proof of the theorem will be the representation for $V^h(x)$ as the minimal cost for a stochastic control problem, as given in Section 2. Thus we have

$$V^h(x) = \inf E_x\left[ \sum_{i=0}^{(N^h \wedge M^h) - 1} L\left(\xi_i^h, u_i^h\right) \Delta t^h(u_i^h) + g\left(\xi_{N^h \wedge M^h}^h\right) \right],$$

where the infimum is over all admissible control sequences and controlled stopping times $M^h$, and $N^h$ is the first time of exit from $\mathcal{F}$ or entrance into $\mathcal{K}$. The transition probabilities and interpolation times are those of Example 2.2, and $L(x, u)$ is the running cost used in Section 2 and given in (2.5). General references for terminology, basic results from weak convergence theory and the properties of martingales that are used are references 5 and 14. Let $\{y^n, n \in \mathbb{N}\}$ and $y$ be random variables that take values in some metric space $S$. We follow the standard practice of saying that the random variables $y^n$ converge weakly to $y$ (denoted $y^n \Rightarrow y$) if the measures induced by $y^n$ on $S$ converge weakly to the measure induced by $y$ on $S$.

In order to study $V^h(x)$, we must first put the representation into a form suitable to taking limits. Let $\xi_i^h$ and $u_i^h$ be the state of the controlled chain and the control applied at time $i$, respectively. Define $t_i^h = \sum_{n=0}^{i-1} \Delta t^h(u_i^h)$. Thus $t_n^h$ is the interpolated time up until discrete time $n$. Define $\tau^h = t_{N^h}^h$ and $\rho^h = t_{M^h}^h$, and also the piecewise constant interpolations

$$\xi^h(t) = \xi_i^h, \quad u^h(t) = u_i^h \quad \text{for} \quad t \in [t_i^h, t_{i+1}^h).$$

We consider the processes $\xi^h(\cdot)$ as taking values in $D([0, \infty); \mathbb{R}^2)$, the metric space of $\mathbb{R}^2$-valued functions that are continuous from the right and have limits from the left. [Our interest in the control actually stops when $i = N^h \wedge M^h$. However, to simplify the notation we may assume the control is defined for all $i \in \mathbb{N}$. Specifically, its value will be defined as $(0, -\gamma_2)$ when
$i \geq N^h \wedge M^h$, so that the running costs per unit time after $N^h \wedge M^h$ are automatically bounded. For points $x \in \mathcal{D}^h$ we can let $L(x, \beta)$ be $L(y, \beta)$, where $y$ is any point in $\mathcal{S}$. In order to take limits of the sequences of controls it will be convenient to use an alternative representation for the control processes. This alternative representation will allow the use of weak convergence methods and also provide a topology on the space of controls. For all $t \geq 0$ and $h > 0$ define

\begin{equation}
(5.7) \quad m^h_t(d\alpha) = \delta_{u^h(t)}(d\alpha) \quad \text{and} \quad m^h(A \times B) = \int_B m^h_t(A) \, dt,
\end{equation}

where $\delta_u(d\alpha)$ is the probability measure that puts a unit mass at $u$. The $m^h$ are random and take values in the space of measures on $\mathbb{R}^2 \times [0, \infty)$. We consider this space as endowed with the following topology: A sequence $l_n$ of measures on $\mathbb{R}^2 \times [0, \infty)$ converges to $l$ if

\begin{equation}
\int_{\mathbb{R}^2 \times [0, \infty)} s(u, t)l_n(du \times dt) \to \int_{\mathbb{R}^2 \times [0, \infty)} s(u, t)l(du \times dt)
\end{equation}

for every $s \in C(\mathbb{R}^2 \times [0, \infty))$ with compact support. The random measures $\{m^h, h > 0\}$ will be called tight if the corresponding restrictions of the measures to $\mathbb{R}^2 \times [0, T]$ are tight in the usual sense for each $T < \infty$. The representation $m^h$ for the control process $u^h(\cdot)$ is known as the relaxed control representation. A measure $l^r$ on $\mathbb{R}^2 \times [0, \infty)$ will be called a relaxed control process if $l^r(\mathbb{R}^2 \times [0, t]) = t$ for all $t \in [0, \infty)$. For such a measure, it follows from the existence of regular conditional probability measures that $l^r(\cdot)$ has a derivative in the following sense: For each $t \in [0, \infty)$ there exists a probability measure $l^r_t(\cdot)$ on $\mathbb{R}^2$ such that $l_t(A \times B) = \int_B l^r_t(A) \, dt$ for all Borel sets $B \subset [0, \infty)$ and $A \subset \mathbb{R}^2$ (see [5], page 502).

With these definitions, we can write

\begin{equation}
(5.8) \quad V^h(x) = \inf E_x \left[ \int_0^T L(\xi^h(t), u) m^h(du \times dt) + g(\xi^h(\rho^h \wedge \tau^h)) \right].
\end{equation}

In the next lemma we derive some properties of the controlled processes under an arbitrary admissible control with bounded running cost. In the lemma's statement, the expectation operator actually depends on the admissible control strategy that is used. In order to simplify the notation, this dependence is not denoted explicitly. The function $L$ in the statement of the lemma is of course the same as that used in (5.8) and is the running cost defined in Section 2.

**Lemma 5.2.** Consider any sequence of initial conditions $\xi^h_0 \in \mathcal{D}$ and admissible controls for which

\begin{equation}
(5.9) \quad \limsup_{h \to 0} E_{\xi^h_0} \int_0^T L(\xi^h(t), u) m^h(du \times dt) < \infty
\end{equation}

for all $T < \infty$. Then the random measures $\{m^h(\cdot), h > 0\}$ are tight. Suppose that a subsequence (again indexed by $h$) is given such that $m^h(\cdot)$ converges
weakly to a limit $m(\cdot)$ and that $\xi^h_0$ converges to $x_0$. Then $m(\cdot)$ is a relaxed control process (w.p.1), and hence can be written $m(du \times dt) = m_t(du) dt$. Furthermore, the sequence \{(\xi^h, m^h), h > 0\} converges weakly to a limit \((x, m)\) that satisfies

\[(5.10) \quad x(t) - x_0 = \int_0^t \int_{\mathbb{R}^2} um(du \times ds) = \int_0^t \int_{\mathbb{R}^2} um_s(du) ds\]

for all $t \in [0, \infty)$, w.p.1.

PROOF. If (5.9) holds, then the fact that $L(x, u) = \infty$ whenever $\|u\| > 2$ implies for all $h > 0$ that the restrictions of $m^h(\cdot)$ to $\mathbb{R}^2 \times [0, T]$ are supported on the compact set $\{u: \|u\| \leq 2\} \times [0, T]$ (w.p.1). Hence the tightness of \{\{m^h, h > 0\}\} is automatic. Assume that $m^h$ converges weakly to $m$. Since $m^h([\mathbb{R}^2 \times [0, t]) = t$ for all $\omega$ and $t \in [0, \infty)$, $m(\mathbb{R}^2 \times [0, t]) = t$ for $t \in [0, \infty)$ (w.p.1). Thus $m(\cdot)$ is a relaxed control process (w.p.1). Define the random processes

\[x^h(t) = \int_0^t \int_{\mathbb{R}^2} um^h(du \times ds) + \xi^h_0 \quad \text{and} \quad x(t) = \int_0^t \int_{\mathbb{R}^2} um(du \times ds) + x_0.\]

We view these processes as taking values in $C([0, \infty); \mathbb{R}^2)$, the space of continuous functions from $[0, \infty)$ to $\mathbb{R}^2$. This space is equipped with the metric inherited as a subset of $D([0, \infty); \mathbb{R}^2)$, under which convergence is equivalent to uniform convergence on each compact subset of $[0, \infty)$. Then the fact that the second marginal of $m$ is Lebesgue measure together with the uniform (in $t$, $h$ and $\omega$) boundedness of the supports of the $m^h$ implies $x^h \Rightarrow x$.

We next prove the weak convergence $\xi^h \Rightarrow x$. If we use the fact that $x^h \Rightarrow x$, then to prove $\xi^h \Rightarrow x$ it suffices to show for each $T < \infty$ that

\[(5.11) \quad \sup_{t \in [0, T]} \|\xi^h(t) - x^h(t)\| \to 0\]

is probability as $h \to 0$. Let $N^h(T) = \inf\{n: t^h_n \geq T\}$. By (2.10),

\[E\left[\xi^h_{i+1} - \xi^h_i - u^h_i \Delta t^h(u^h_i)\| (\xi^h_j, u^h_j), j \leq i\right] = 0.\]

This implies that

\[(5.12) \quad \xi^h_i - \xi^h_0 - \int_0^{t^h_i} \int_{\mathbb{R}^2} um^h(du \times ds)\]

is a martingale in $i$. Using (2.11), we compute that

\[E_{\xi^h_0}\left[\xi^h_{N^h(T)} - \xi^h_0 - \int_0^{t^h_{N^h(T)}} \int_{\mathbb{R}^2} um^h(du \times ds)\right]^2 \to 0\]

as $h \to 0$. Hence by the submartingale inequality ([5], Chapter 2),

\[\sup_{0 \leq i \leq N^h(T)} \left\|\xi^h_i - \xi^h_0 - \int_0^{t^h_i} \int_{\mathbb{R}^2} um^h(du \times ds)\right\| \to 0\]

in probability as $h \to 0$. This clearly implies (5.11), and completes the proof of the lemma. \(\Box\)
Proof of the upper bound. We first prove \( \lim \sup_{h \to 0} V^h(x) \leq V(x) \). We begin by describing a "controllability" property of the chain \( \{ \xi^h_i, i \in \mathbb{N} \} \). An elementary calculation shows \( L(x, u) \leq 2 \) for all \( x \in \mathcal{D} \) and \( u \in \mathcal{A}(x) \). Let \( y \) be any point in \( \mathcal{D} \). The running cost at such a point is \( \gamma_3^2 - \gamma_2 u_2 - \gamma_3(1 - |u_1|^2 - |u_2 + \gamma_2|^2)^{1/2} \) for all \( u \) such that \( |u_1|^2 + |u_2 + \gamma_2|^2 \leq 1 \), and \( \infty \) otherwise. Since \( |\gamma_2| < 1 \), there exists \( c > 0 \) such that \( L(y, u) \leq 2 \) for all \( u \) satisfying \( ||u|| \leq c \). Owing to the continuity of \( I(\cdot) \) and the form of \( L(x, u) \), we can in fact assume \( L(x, u) \leq 2 \) whenever \( ||x - y|| \leq c \) and \( ||u|| \leq c \).

Fix \( x \) such that \( ||x - y|| \leq c \) and define \( j_i = |x_i - y_i|/h \), \( i = 1, 2 \). Suppose the control \( u = ([-\text{sign}(x_1 - y_1)]c, 0) \) is applied for exactly \( j_1 \) steps, and that after this the control \((0, [-\text{sign}(x_2 - y_2)]c)\) is applied for \( j_2 \) steps. It is easy to see that this control is admissible. The "deterministic" form of the transition probabilities under these controls guarantees that the controlled chain arrives at \( y \) at discrete time \( j = j_1 + j_2 \). Since for all steps \( \Delta t^h(u) = h/c \), the running cost accumulated during this time will be bounded above by

\[
(5.13) \quad 2 \left[ \frac{1}{h} (|x_1 - y_1| + |x_2 - y_2|) \right] \left[ \frac{h}{c} \right] = \frac{2}{c} (|x_1 - y_1| + |x_2 - y_2|).
\]

We now give the proof of the upper bound. Let \( \mathcal{G} \) be any set that satisfies Assumption 2.1. According to Lemma 5.1, for each \( x \in \mathcal{G} \) and \( \varepsilon > 0 \) there exists \( \delta > 0 \), \( \tau < \infty \) and piecewise continuous \( u: [0, \tau] \to \mathbb{R}^2 \) such that if \( x(t) = x + \int_0^t u(s) \, ds \), then:

1. \( \int_0^\tau L(x(t), u(t)) \, dt + g(x(\tau)) \leq V(x) + \varepsilon \).
2. \( x(\tau) \in \mathcal{M}, x(t) \in \mathcal{M} \) for all \( t \in [0, \tau) \).
3. \( d(u(t), \mathcal{A}\varepsilon(x(t))) \geq \delta \) for all \( t \in [0, \tau] \).

Since \( \mathcal{A}(x) \) is continuous in \( x \), we can assume \( \delta > 0 \) has been chosen small enough that

\[
(5.14) \quad d(u(t), \mathcal{A}\varepsilon(x)) \geq \delta
\]

for all \( x \) such that \( |x - x(t)| \leq \delta \) and a.e. \( t \in [0, \tau] \).

Since \( \mathcal{G} \subset \mathcal{D} \) is compact and \( \mathcal{D} \) is open, we can also assume \( d(x, \mathcal{D}\varepsilon) \geq \delta \) for all \( x \in \mathcal{G} \).

Let \( y = x(\tau) \in \mathcal{M} \). Fix \( c > 0 \) so that the "controllability" property and the bound \((5.13)\) on the running cost will hold. We now define an admissible control scheme for the Markov chain in terms of \( u(\cdot) \). To apply \( u(\cdot) \) to the chain \( \{ \xi^h_i, i < \infty \} \) we recursively define the control applied at discrete time \( i \) by \( u^h_i = u(t^h_i) \) and \( t^h_{i+1} = t^h_i + \Delta t^h(u^h_i) \). This defines a control until \( i \) such that \( t^h_{i+1} \geq \tau \). Let \( \{ \xi^h_i, i < \infty \} \) be the chain that starts at \( x \) and uses this control.

Define

\[
S^h = \inf \{ i: t^h_i \geq \tau \text{ or } \xi^h_i \in \mathcal{M} \text{ or } \xi^h_i \notin \mathcal{D} \text{ or } \|\xi^h_i - x(t^h_i)\| \geq \delta \},
\]
and let $\sigma^h = t^h_{\delta h}$. By construction the restrictions of the measures $m^h(du \times ds)$ to $\mathbb{R}^2 \times [0, \tau]$ converge weakly to the measure $m(du \times ds) = \delta_{\mu}(du) ds$. Hence by Lemma 5.2, we have $\sup_{0 \leq t \leq \sigma^h} \| \xi^h(t) - x(t)\| \to 0$ in probability, and for each $\theta > 0$, $P_x(\| \xi^h_{\delta h} - x(\tau)\| \geq \theta) \to 0$ as $h \to 0$.

Assume $\theta \in (0, c)$. If $\| \xi^h_{\delta h} - x(\tau)\| \geq \theta$, then we stop the process at discrete time $S^h$ and pay the stopping cost $g(\xi^h_{\delta h}) \leq B$. On the set where $\| \xi^h_{\delta h} - x(\tau)\| < \theta$, we extend the definition of the control sequence for discrete times larger than $S^h$ according to the discussion at the beginning of the proof. The control applied after $S^h$ will drive $\xi^h_i$ to $y$ in fewer than $2\theta/c$ steps with a running cost that is bounded above by $4\theta/c$.

The total cost for the control scheme and stopping time defined in this way is bounded above by

$$E_x\int_0^{\sigma^h} \int_{\mathbb{R}^2} L(\xi^h(s), u) m^h(du \times ds) + 4\theta/c$$

$$+ P_x(\| \xi^h_{\delta h} - x(\tau)\| \geq \theta)B$$

$$+ P_x(\| \xi^h_{\delta h} - x(\tau)\| < \theta)\sup\{ g(x) : \| x - y\| \leq \theta, x \in \mathcal{M} \}.$$ 

According to the Skorokhod representation theorem ([5], Theorem 3.1.8), we can assume that the convergence $(\xi^h(\cdot), m^h(\cdot)) \to (x(\cdot), m(\cdot))$ is w.p.1 for the purposes of evaluating the limits of the expectations above, and that $x$ and $m$ satisfy (5.10). Owing to the definition of $S^h$ and (5.14), we can apply the dominated convergence theorem to obtain

$$\limsup_{h \to 0} V^h(x) \leq \int_0^\tau \int_{\mathbb{R}^2} L(\xi(s), u) m(du \times ds)$$

$$+ \sup\{ g(z) : \| z - y\| \leq \theta, z \in \mathcal{M} \} + 4\theta/c$$

$$= \int_0^\tau L(x(s), u(s)) \, ds + \sup\{ g(z) : \| z - y\| \leq \theta, z \in \mathcal{M} \}$$

$$+ 4\theta/c.$$ 

Since $\varepsilon > 0$ and $\theta \in (0, c)$ are arbitrary and $y = x(\tau)$,

$$\limsup_{h \to 0} V^h(x) \leq V(x).$$

**Proof of the lower bound.** We now prove $\liminf_{h \to 0} V^h(x) \geq V(x)$. Fix $x \in \mathcal{S} - \mathcal{M}$ and $\varepsilon > 0$. Owing to the definition of $V^h(x)$, there is a controlled Markov chain $(\xi^h_i, i < \infty)$ with admissible control sequence $(u^h_i, i < \infty)$ that satisfies $\xi^h_0 = x$, and a stopping time $\tau^h$ such that

$$V^h(x) \geq E_x \sum_{j=0}^{(N^h \wedge M^h) - 1} L(\xi^h_j, u^h_j) \Delta t^h(u^h_j) + E_x g(\xi^h_{N^h \wedge M^h}) - \varepsilon,$$

where $N^h$ is the time of first exit from $\mathcal{D}$ or entrance into the set $\mathcal{M}$. Let $\xi^h(\cdot)$ and $u^h(\cdot)$ be the continuous parameter interpolations of $(\xi^h_i, i < \infty)$ and $(u^h_i, i < \infty)$ as defined by (5.6), and define $\rho^h = t^h_{M^h}, \tau^h = t^h_{N^h}$. We can then
rewrite (5.15) as

\begin{equation}
V^h(x) \geq E_x \int_0^{\rho_h \wedge \tau_h} \int_{\mathbb{R}^2} L(\xi^h(s), u) m^h(du \times ds) \\
+ E_x g(\xi^h(\rho_h \wedge \tau_h)) - \varepsilon,
\end{equation}

where \(m^h(\cdot)\) is the relaxed control representation of the ordinary control \(u^h(\cdot)\).

Let \(\mathcal{S}_q, q = 1, \ldots, Q\), be disjoint compact connected sets such that \(\mathcal{S} = \bigcup_{q=1}^Q \mathcal{S}_q\). The existence of such a decomposition has been assumed in the statement of Theorem 2.3. Now \(V(x)\) is constant on each \(\mathcal{S}_q\), so there exists \(\theta > 0\) such that

\begin{equation}
x \in \mathcal{S}_q, \quad y \in N_\theta(\mathcal{S}_q) \Rightarrow |V(x) - V(y)| \leq \varepsilon
\end{equation}

and such that the sets \(N_\theta(\mathcal{S}_q)\) are separated by a distance greater than \(\theta\) for distinct \(q\). Because the reflected light intensity \(I(\cdot)\) is continuous on the closure of \(\mathcal{D}\), there is \(c > 0\) such that

\begin{equation}
L(x, u) \geq c \quad \text{for all } u \in \mathbb{R}^2, \quad x \in \mathcal{D} - \bigcup_{q=1}^Q N_{\theta/2}(\mathcal{S}_q).
\end{equation}

For simplicity, we will consider the proof of the lower bound for the case when the initial condition satisfies \(x \in N_{\theta/2}(\mathcal{S}_{q^*})\) for some \(q^*\). The general case follows easily using the same arguments. We define a sequence of stopping times by

\(\tau^h_0 = 0\),

\(\sigma^h_j = \inf\left\{ t \geq \tau^h_j : \xi^h(t) \notin \bigcup_{q=1}^Q N_\theta(\mathcal{S}_q) \right\},\)

\(\tau^h_j = \inf\left\{ t \geq \sigma^h_{j-1} : \xi^h(t) \notin \bigcup_{q=1}^Q N_{\theta/2}(\mathcal{S}_q) \text{ or } \xi^h(t) \notin \mathcal{D} \right\}.

Consider the processes

\(\Xi^h(\cdot) = (\xi^h_0(\cdot), \xi^h_1(\cdot), \ldots), \quad M^h(\cdot) = (m^h_0(\cdot), m^h_1(\cdot), \ldots),\)

where \(\xi^h_j(\cdot) = \xi^h(\cdot + \sigma^h_j)\) and where \(m^h_j(\cdot)\) is the relaxed control representation of the ordinary control \(u^h(\cdot + \sigma^h_j)\). We consider \((\Xi^h(\cdot), M^h(\cdot))\) as taking values in the product space endowed with the usual product space topology, that is, a sequence converges if and only if each finite subset of components converges.

Consider any subsequence along which \(V^h(x)\) converges to a point in \((-\infty, \infty]\). Then it is enough to prove that the limit of this subsequence is no less than \(V(x)\). We can eliminate the case where the limit is \(\infty\), since in this case the lower bound is automatic. For the rest of the proof we shall assume that we are working with a subsequence (again labeled by \(h\)) along which
$V^h(x)$ converges to a bounded limit. Thus the sequence $V^h(x)$ will be uniformly bounded from above.

Lemma 5.2 shows that given any subsequence of $\{(\Xi^h(\cdot), M^h(\cdot)), h > 0\}$, we can extract a further subsequence that converges weakly, and that any limit point

$$(X(\cdot), M(\cdot)) = ((x_0(\cdot), x_1(\cdot), \ldots), (m_0(\cdot), m_1(\cdot), \ldots))$$

of such a convergent subsequence satisfies

$$x_j(t) - x_j(0) = \int_0^t \int_{\mathbb{R}^2} u m_j(du \times ds),$$

where each $m_j(\cdot)$ is a relaxed control process. In addition, the definition of the stopping times $\{\sigma^h\}$ guarantees that $x_0(0) \in \partial N_0(S_{q_1})$, and that for all $j > 0$ either $x_j(0) \in \partial N_0(S_{q_2})$ for some $q$ or $x_j(0) \notin S$.

Let $J^h = \min(j: \tau^h_j \geq \tau^h)$, where $\tau^h$ has been defined to be the interpolated time at which $\xi^h(\cdot)$ first left $S$ or entered $\mathcal{M}$. By construction, if $\xi^h(\tau^h) \in S_q$ for some $q$ (i.e., $\xi^h$ enters $\mathcal{M}$ before it leaves $S$), then $\xi^h(\tau^h_{j-1}) \in N_{\theta/2}(S_{q_2})$ for the same $q$. It follows from $\lim_{h \to 0} V^h(x) < \infty$ and the uniform bound from below given in (5.18) that

$$\limsup_{h \to 0} \sum_{0 \leq j < J^h} (\tau^h_{j+1} - \sigma^h_j) < \infty. \tag{5.19}$$

Define $s^h_j = \tau^h_{j-1} - \sigma^h_j$ and $S^h = (s^h_0, s^h_1, \ldots)$. It also follows from (5.18) that there exists $\bar{c} > 0$ such that for all $q_1$ and $q_2$,

$$\inf \left\{ \int_0^T L(\phi, \dot{\phi}) \, ds: \phi(0) \in \partial N_0(S_{q_1}), \phi(T) \in N_{\theta/2}(S_{q_2}), \phi(t) \in \overline{S}, \, t \in [0, T], \, T > 0 \right\} \geq \bar{c}. \tag{5.20}$$

and $\bar{c} > 0$ that is independent of $\theta$ for all small $\theta > 0$ such that for all $q_1 \neq q_2$,

$$\inf \left\{ \int_0^T L(\phi, \dot{\phi}) \, ds: \phi(0) \in \partial N_0(S_{q_1}), \phi(T) \in N_{\theta/2}(S_{q_2}), \phi(t) \in \overline{S}, \, t \in [0, T], \, T > 0 \right\} \geq \bar{c}. \tag{5.21}$$

We now prove the lower bound $\lim_{h \to 0} V^h(x) \geq V(x)$. Extract a subsequence along which

$$(\xi^h(\cdot), m^h(\cdot), \rho^h, \tau^h, \Xi^h(\cdot), M^h(\cdot), J^h, S^h)$$

converges weakly to a limit

$$(x(\cdot), m(\cdot), \rho, \tau, \Xi(\cdot), M(\cdot), J, S).$$

Note that the bounds (5.19) and (5.20) imply that $J$ and the $s_j$, $j \in \{0, \ldots, J - 1\}$, are finite w.p.1. The $\rho^h$ and $\tau^h$ are regarded as taking values in the
compacted space $[0, \infty]$ in order to guarantee tightness, and hence $\tau$ and $\rho$ may take the value $\infty$. We assume via the Skorokhod representation [5] that the convergence is w.p.1, and consider any $\omega$ for which there is convergence. If $\rho < \tau$, that is, if we choose to stop before entering $\mathcal{M}$ or leaving $\mathcal{D}$, we have

$$\liminf_{h \to 0} g(\xi^h(\rho^h \wedge \tau^h)) = B \geq V(x).$$

Next assume $\tau \leq \rho$. The nonnegativity and lower semicontinuity of $L(x,u)$ in $(x,u)$ then imply

$$\liminf_{h \to 0} \int_0^{\rho^h \wedge \tau^h} \int_{\mathbb{R}^2} L(\xi^h(s),u) m^h(du \times ds)$$

$$\geq \liminf_{h \to 0} \sum_{0 \leq j < J^h} \int_{\sigma_j^h} \int_{\mathbb{R}^2} L(\xi^h(s),u) m^h(du \times ds)$$

$$\geq \sum_{0 \leq j < J} \int_{\mathbb{R}^2} L(x_j(s),u) m_j(du \times ds).$$ (5.22)

Let $j_k$, $k = 1, \ldots, K$, index those values of $j \in \{0, \ldots, J - 1\}$ such that $x_j(0) \in N_q(\mathcal{S}_q)$ and $x_j(s_j) \not\in N_{q/r}(\mathcal{S}_q)$, that is, the $j_k$ label those trajectories that actually leave the neighborhood of one of the $\mathcal{S}_q$ and either enter the neighborhood of some $\mathcal{S}_q'$, $q' \neq q$, or else end up at $\partial \mathcal{D}$. Assume that $j_{k_1} < j_{k_2}$ whenever $k_1 < k_2$.

For any $j \in \{0, \ldots, J - 1\}$ and $s \in [0, s_j]$ define $u_j(s) = \int_{\mathbb{R}^2} m_{j,s}(du)$, where $m_{j,s}(\cdot)$ is the derivative of $m_j(\cdot)$. Then $x_j(t) - x_j(0) = \int_0^t u_j(s) ds$ for $t \in [0, s_j]$ and by Jensen’s inequality,

$$\int_0^{s_j} \int_{\mathbb{R}^2} L(x_j(s),u) m_j(du \times ds) \geq \int_0^{s_j} L(x_j(s),u_j(s)) ds.$$

From the definition of $V(\cdot)$ and an elementary dynamic programming argument,

$$V(x_j(0)) \leq \int_0^{s_j} L(x_j(s),u_j(s)) ds + V(x_j(s_j)).$$

Assembling these inequalities gives that for each $j \in \{0, \ldots, J - 1\},$

$$\int_0^{s_j} \int_{\mathbb{R}^2} L(x_j(s),u) m_j(du \times ds) \geq V(x_j(0)) - V(x_j(s_j)).$$ (5.23)

According to the definitions of the $\tau_i^h$, $\sigma_i^h$ and the indices $j_k$, if $x_{j_k}(s_{j_k}) \in \partial N_{q/r}(\mathcal{S}_q)$, then $x_{j_k}(0) \in N_q(\mathcal{S}_q)$ for all $k \in \{1, \ldots, K - 1\}$. Thus $|V(x_{j_k}(s_{j_k})) - V(x_{j_{k+1}}(0))| \leq 2\varepsilon$ for all such $k$. Together with the fact that
\( x_j(0) \in \mathcal{N}_\theta(\mathcal{G}_q') \) (recall that \( x \in \mathcal{N}_{\theta/2}(\mathcal{G}_q') \), the last sentence together with (5.22) and (5.23) implies

\[
\liminf_{h \to 0} \int_0^{\rho^h + \tau^h} \int_{\mathbb{R}^2} L\left( \xi^h(s), u \right) m^h(du \times ds) \\
\geq \sum_{k \in \{1, \ldots, K\}} \int_0^{s_{j_k}} \int_{\mathbb{R}^2} L\left( x_{j_k}(s), u \right) m_{j_k}(du \times ds) \\
\geq \sum_{k \in \{1, \ldots, K\}} [V(x_{j_k}(0)) - V(x_{j_k}(s_{j_k}))]
\]
\[
\geq V(x) - V(x_{j_k}(s_{j_k})) \\
+ \sum_{k \in \{2, \ldots, K\}} [V(x_{j_k}(0)) - V(x_{j_{k-1}}(s_{j_{k-1}}))] - \varepsilon
\]
\[
\geq V(x) - V(x_{j_k}(s_{j_k})) - (2K - 1)\varepsilon
\]

w.p.1.

Next consider \( \liminf_{h \to 0} g(\xi^h(\rho^h + \tau^h)) \). As previously noted, \( \liminf_{h \to 0} g(\xi^h(\rho^h + \tau^h)) \in B \) if \( \rho < \tau \). If \( \tau \leq \rho \), there are two possibilities. Recall that \( \rho^h \) is the controlled stopping time and that \( \tau^h \) is the first time the process enters \( \mathcal{M} \) or leaves \( \mathcal{D} \). If \( \rho^h < \tau^h \) or \( \rho^h \geq \tau^h \) and \( \xi^h(\tau^h) \notin \mathcal{M} \), then \( g(\xi^h(\rho^h + \tau^h)) \in B \). If \( \rho^h \geq \tau^h \) and \( \xi^h(\tau^h) \in \mathcal{G}_q \subset \mathcal{M} \), then as observed previously \( \xi^h(\tau^h_{j_{K-1}}) \in \mathcal{N}_{\theta/2}(\mathcal{G}_q') \) and therefore \( V(\xi^h(\tau^h_{j_{K-1}})) \leq g(\xi^h(\rho^h + \tau^h)) + \varepsilon \). Now \( j_K \) is the last index for which there is a transition between different \( N_\theta(\mathcal{G}_q') \). Thus, in general,

\[
\liminf_{h \to 0} g(\xi^h(\rho^h + \tau^h)) \geq V(x_{j_k}(s_{j_k})) - \varepsilon.
\]

Combining (5.24) and (5.25) gives

\[
\lim_{h \to 0} V^h(x) \geq V(x) - 2\varepsilon E_x K.
\]

Now by (5.21) we also have

\[
\lim_{h \to 0} V^h(x) \geq \tilde{c} E_x K,
\]

where \( \tilde{c} > 0 \) is independent of \( \varepsilon > 0 \). Thus \( E_x K \) has a bound that is independent of \( \varepsilon > 0 \). Sending \( \varepsilon \to 0 \) gives the desired lower bound \( \lim_{h \to 0} V^h(x) \geq V(x) \). \( \square \)

**Remark.** Dupuis would like to acknowledge an error in the proof of the lower bound as it appears in [14], Chapter 13. That proof considered convergence in the special case of vertical light only. In the proof, the distinction between the paths which move between neighborhoods of different \( \mathcal{G}_q \) and those that do not was omitted, making the assertion of a uniform upper bound on \( J \) incorrect. The correct assertion is the uniform upper bound on \( K \), as noted previously.
The next two proofs are of Propositions 3.1 and 3.2, respectively. The first proposition shows that if the initial conditions are the same, then so are the approximations to $V$ that are produced when either of the algorithms from Sections 2 or 3 is used. The second proposition provides a characterization of the approximation in terms of the initial condition, and proves that the iterates of any of the algorithms are monotonic.

**Proof of Proposition 3.1.** We will distinguish the functions that were defined in Sections 3 and 2 by using the superscripts (1) and (2), respectively. It is sufficient to consider only the case $\gamma_2 \leq 0$. For convenience we recall the various functions of interest. We have

$$L^{(1)}(x, \beta) = \frac{1}{2} \frac{|\beta_1|^2}{I^2(x)} + \frac{1}{2} \frac{\beta_2 + (1 - I^2(x))\gamma_2}{v(x)} + \frac{1}{2} (1 - I^2(x)),$$

(5.26)

$$L^{(2)}(x, \beta) = \begin{cases} \gamma_3^2 - \gamma_2 \beta_2 - \gamma_3 (I^2(x) - |\beta_1|^2 - |\beta_2 + \gamma_2|^2)^{1/2}, & \text{if } |\beta_1|^2 + |\beta_2 + \gamma_2|^2 \leq I^2(x), \\ \infty, & \text{if } |\beta_1|^2 + |\beta_2 + \gamma_2|^2 > I^2(x), \end{cases}$$

(5.27)

$$H^{(1)}(x, \alpha) = \sup_{\beta \in \mathbb{R}^2} \left[ -\langle \alpha, \beta \rangle - L^{(1)}(x, \beta) \right]$$

(5.28)

$$= \frac{1}{2} \left[ I^2(x) \alpha_1^2 + v(x) \alpha_2^2 + 2(1 - I^2(x))\gamma_2 \alpha_2 - (1 - I^2(x)) \right]$$

if $v(x) = I^2(x) - \gamma_2^2 \geq 0$,

$$H^{(1)}(x, \alpha) = \inf_{\beta_2} \sup_{\beta_1} \left[ -\langle \alpha, \beta \rangle - L^{(1)}(x, \beta) \right]$$

(5.29)

$$= \frac{1}{2} \left[ I^2(x) \alpha_1^2 + v(x) \alpha_2^2 + 2(1 - I^2(x))\gamma_2 \alpha_2 - (1 - I^2(x)) \right]$$

if $v(x) < 0$,

and finally

$$H^{(2)}(x, \alpha) = \sup_{\beta \in \mathbb{R}^2} \left[ -\langle \alpha, \beta \rangle - L^{(2)}(x, \beta) \right]$$

(5.30)

$$= I(x) \left( 1 + \|\alpha\|^2 - 2 \alpha_2 \gamma_2 \right)^{1/2} + \alpha_2 \gamma_2 - 1.$$
of the fixed points will be to relate the discrete equations that characterize a
fixed point back to these functions. For points $x$ where $\nu(x) > 0$, it is easy to
see that the convexity of both $H^{(1)}$ and $H^{(2)}$ in $\alpha$ implies

\begin{equation}
H^{(1)}(x, \alpha) \begin{cases}
> & 0 \\
< & \end{cases} H^{(2)}(x, \alpha) \begin{cases}
> & 0 \\
< & \end{cases}.
\end{equation}

(5.31)

For the points where $\nu(x) \leq 0$, the set \{ $\alpha$: $H^{(1)}(x, \alpha) = 0$ \} takes the form of a
hyperbola with two branches: one that opens in the positive $\alpha_2$ direction and
one that opens in the negative $\alpha_2$ direction. For the case $\gamma_2 \leq 0$ considered
here, the image irradiance equation must be satisfied with $f_x(x)$ taking a
value in the branch of the hyperbola that opens in the positive $\alpha_2$ direction.
This is because a value in the branch that opens in the negative $\alpha_2$ direction
corresponds to a surface normal that points away from the camera
direction, which is impossible if $f(\cdot)$ is a function. Even more to the point is
that the positive branch coincides precisely with the zeros of $H^{(2)}(x, \cdot)$.

A consequence of the convex duality formula used in the demonstration of
(5.40) is that

\begin{equation}
\tilde{H}^{(1)}(x, \alpha) = \inf_{\beta_2 \geq 0} \sup_{\beta_1} \left[ -\langle \alpha, \beta \rangle - L^{(1)}(x, \beta) \right]
= \sup_{\alpha_2 \geq 0} H^{(1)}(x, \alpha + (0, \alpha_2^*)).
\end{equation}

(5.32)

The function $\tilde{H}^{(1)}(x, \alpha)$ is automatically nonincreasing in $\alpha_2$ for each fixed
$\alpha_1$. Note that the zeros of the function $\tilde{H}^{(1)}(x, \alpha)$ correspond exactly to the
branch of the hyperbola that opens in the positive $\alpha_2$ direction. It follows that

\begin{equation}
\tilde{H}^{(1)}(x, \alpha) \begin{cases}
> & 0 \\
< & \end{cases} H^{(2)}(x, \alpha) \begin{cases}
> & 0 \\
< & \end{cases}.
\end{equation}

(5.33)

(It would indeed be possible to construct the algorithm of Section 3 directly
from the equation $\tilde{H}^{(1)}(x, f_x(x)) = 0$, which by (5.33) is equivalent to the
image irradiance equation.)

We now consider the equation for a fixed point of (3.7) with running cost
$L^{(1)}(x, \beta)$, and that for a fixed point of (2.14) with running cost $L^{(2)}(x, \beta)$. By
the statement "(3.7) [or (2.14)] holds with $w(\cdot)$ at $x$," we mean that (3.7) [or
(2.14)] holds at $x$ with $V_n^h$ and $V_n^{h+1}$ replaced by $w$. It will be convenient to
introduce some new notation. Define $Q_1 = \{ x: x_1 \geq 0, x_2 \geq 0 \}$, $Q_2 = \{ x:
= 0, x_2 \geq 0 \}$, $Q_3 = \{ x: x_1 \leq 0, x_2 \geq 0 \}$ and $Q_4 = \{ x: x_1 \geq 0, x_2 \leq 0 \}$. We
set $w_1(x) = (w(x + h(1, 0)) - w(x), w(x + h(0, 1)) - w(x))$ and $w_2(x) =
(w(x + h(-1, 0)) - w(x), w(x + h(0, 1)) - w(x))$, and define $w_3(x)$ and $w_4(x)$
analogously.

First assume $w(x) < g(x)$. If we insert the transition probabilities and
interpolation interval of Example 2.2, (2.14) implies

\begin{equation}
\left\{ \min_{i=1,2,3,4} \left[ \inf_{u \in Q_i\setminus\{0\}} \frac{L^{(2)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} \right] \right\} \wedge (L^{(2)}(x, 0)) = 0.
\end{equation}

(5.34)
SHAPE RECOVERY VIA OPTIMAL CONTROL

Recall that \( L^{(2)}(x, \beta) \geq 0 \) and that \( L^{(2)}(x, \beta) = 0 \) if and only if \( (x, u) \in \mathcal{S} \times (0) \). Thus for \( x \in \mathcal{S} \) we obtain

\[
(5.35) \quad \min_{i=1, 2, 3, 4} \left[ \inf_{u \in \mathcal{Q} \setminus \{0\}} \frac{L^{(2)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} \right] \geq 0,
\]

while for \( x \notin \mathcal{S} \) we have

\[
(5.36) \quad \min_{i=1, 2, 3, 4} \left[ \inf_{u \in \mathcal{Q} \setminus \{0\}} \frac{L^{(2)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} \right] = 0.
\]

Next consider the case \( w(x) = g(x) \). For this case we deduce that (5.35) holds for all \( x \).

We next simplify the inequalities by showing that we can eliminate the denominator. Consider any \( x \in \mathcal{S} \). Then for each \( i = 1, 2, 3, 4 \),

\[
(5.37) \quad \begin{align*}
\inf_{u \in \mathcal{Q} \setminus \{0\}} \frac{L^{(2)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} & \geq 0 \\
\iff & \inf_{u \in \mathcal{Q} \setminus \{0\}} \left[ L^{(2)}(x, u) + \langle u, w_i(x) \rangle \right] \geq 0.
\end{align*}
\]

Now consider any point \( x \notin \mathcal{S} \). Owing to the strictly positive lower bound on \( L^{(2)}(x, \cdot) \) for such points, for each \( i = 1, 2, 3, 4 \),

\[
(5.38) \quad \begin{align*}
\inf_{u \in \mathcal{Q} \setminus \{0\}} \frac{L^{(2)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} & \geq 0 \\
\iff & \inf_{u \in \mathcal{Q} \setminus \{0\}} \left[ L^{(2)}(x, u) + \langle u, w_i(x) \rangle \right] \geq 0.
\end{align*}
\]

It will be demonstrated below that (5.31), (5.33) and a convex duality formula imply

\[
(5.39) \quad \begin{align*}
\inf_{u \in \mathcal{Q} \setminus \{0\}} \left[ L^{(2)}(x, u) + \langle u, w_i(x) \rangle \right] \left\{ \geq \right\} 0 \\
\iff & \inf_{u \in \mathcal{Q} \setminus \{0\}} \left[ L^{(1)}(x, u) + \langle u, w_i(x) \rangle \right] \left\{ \geq \right\} 0
\end{align*}
\]

for \( i = 1, 2, 3, 4 \) when \( v(x) > 0 \) and

\[
(5.40) \quad \begin{align*}
\inf_{u \in \mathcal{Q} \setminus \{0\}} \left[ L^{(2)}(x, u) + \langle u, w_i(x) \rangle \right] \left\{ \geq \right\} 0 \\
\iff & \sup_{u_2 > 0} \inf_{u_1 \geq 0} \left[ L^{(1)}(x, u) + \langle u, w_i(x) \rangle \right] \left\{ \geq \right\} 0
\end{align*}
\]

for \( i = 1, 2 \) when \( v(x) \leq 0 \). Postponing temporarily the proofs of (5.39) and (5.40), we now complete the proof of the equivalence of the fixed points. It will be convenient to separate the cases \( \mathcal{S}_1 = \{ x : x \in \mathcal{S} \} \), \( \mathcal{S}_2 = \{ x : x \notin \mathcal{S} \} \), \( \mathcal{S}_3 = \{ x : x \notin \mathcal{S} \} \), \( \mathcal{S}_4 = \{ x : x \notin \mathcal{S} \} \), \( \mathcal{S}_5 = \{ x : x \notin \mathcal{S} \} \), \( \mathcal{S}_6 = \{ x : x \notin \mathcal{S} \} \), \( w(x) = g(x) \), \( v(x) > 0 \) and \( \mathcal{S}_6 = \{ x : x \notin \mathcal{S} \} \), \( w(x) = g(x) \), \( v(x) \leq 0 \) and \( \mathcal{S}_6 = \{ x : x \notin \mathcal{S} \} \), \( w(x) = g(x) \), \( v(x) \leq 0 \).
Assume (3.7) holds with $w(\cdot)$ at $x$. Note that (5.37) holds for $L^{(1)}$ as well as $L^{(2)}$. For $x \in \mathcal{A}'$, it is always true that (5.35) holds, and by (5.37) with $L^{(2)}$ replaced by $L^{(1)}$ and (5.39), equation (5.35) holds with $L^{(2)}$ replaced by $L^{(1)}$. Since $L^{(2)}(x, 0) = L^{(1)}(x, 0) = 0$, (3.7) is satisfied for $x \in \mathcal{A}'$. Next consider $x \in \mathcal{A}_2$. In this case (5.36) holds. Since $L^{(1)}(x, \cdot)$ has a strictly positive lower bound for $x \in \mathcal{A}_2$, (5.38) holds with $L^{(2)}$ replaced by $L^{(1)}$. If we now use (5.38) with $L^{(2)}$ replaced by $L^{(1)}$ and (5.39) we find that (5.36) holds with $L^{(2)}$ replaced by $L^{(1)}$. Since $L^{(1)}(x, 0) > 0$ for $x \in \mathcal{A}_2$, (3.7) holds for $x \in \mathcal{A}_2$.

Next consider $x \in \mathcal{A}_3$. The analogue of (5.37) that is appropriate for $L^{(1)}$ is

$$
\sup_{u_2 > 0} \inf_{(-1)^{i+1} u_1 \geq 0} \frac{L^{(1)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} \left\{ \begin{array}{c} > \ 0 \\ = \ 0 \end{array} \right. \quad (5.41)
$$

$$
\Leftrightarrow \sup_{u_2 > 0} \inf_{(-1)^{i+1} u_1 \geq 0} \left[ L^{(1)}(x, u) + \langle u, w_i(x) \rangle \right] \left\{ \begin{array}{c} > \ 0 \\ = \ 0 \end{array} \right. \quad \text{for } i = 1, 2.
$$

This equation is most easily verified by showing that strict positivity (respectively, strict negativity) of the left-hand side implies strict positivity (respectively, strict negativity) of the right-hand side, and conversely. For $x \in \mathcal{A}_3$ we have (5.36) for $i = 1, 2$, which by (5.40) and (5.41) implies

$$
\sup_{u_2 > 0} \inf_{(-1)^{i+1} u_1 \geq 0} \frac{L^{(1)}(x, u) + \langle u, w_i(x) \rangle}{|u_1| + |u_2|} = 0 \quad (5.42)
$$

for $i = 1, 2$. Thus (3.7) holds for $x \in \mathcal{A}_3$.

The proofs for $x \in \mathcal{A}_4$ and $x \in \mathcal{A}_5$ are very similar and are omitted. Thus we have proved that any fixed point of (2.14) is also a fixed point of (3.7). The proof of the reverse implication is essentially the same, and for the sake of brevity we give details only for the case $x \in \mathcal{A}_3$. For such points, (3.7) implies (5.42) for $i = 1, 2$. Tracing back through (5.40) and (5.41) we see that (5.36) holds for $i = 1, 2$. Since for $x \in \mathcal{A}_3$, $L^{(2)}(x, u) = \infty$ for $u \in Q_3 \cup Q_4$, this implies (2.14) at the point $x$.

Thus all that remains are the proofs of (5.39) and (5.40). Consider first (5.39) and the case $i = 1$. All other cases of $i$ may be treated similarly. Owing to continuity of the bracketed quantities below,

$$
\inf_{u \in Q_1 \setminus \{0\}} \left[ L^{(j)}(x, u) + \langle u, w_i(x) \rangle \right] = \inf_{u \in Q_1} \left[ L^{(j)}(x, u) + \langle u, w_i(x) \rangle \right] \quad (5.43)
$$

for $j = 1, 2$. Thus we can consider the infima over $Q_1$. Define

$$
I_{Q_1}(\beta) = \begin{cases} 
0, & \beta \in Q_1, \\
+\infty, & \beta \notin Q_1.
\end{cases}
$$

We can then write

$$
\inf_{u \in Q_1} \left[ L^{(j)}(x, u) + \langle u, w_i(x) \rangle \right] = \inf_{u \in \mathbb{R}^2} \left[ (L^{(j)}(x, u) + I_{Q_1}(u)) + \langle u, w_i(x) \rangle \right]
$$
for $j = 1, 2$. Define the Legendre transforms
\[
L^{(j),*}(x, \alpha) = \inf_{u \in \mathbb{R}^2} \left[ L^{(j)}(x, u) + \langle u, \alpha \rangle \right],
\]
\[
I^{*}_{Q_i}(\alpha) = \inf_{u \in \mathbb{R}^2} \left[ I_{Q_i}(u) + \langle u, \alpha \rangle \right].
\]

Note that the Legendre transform considered here differs by a sign convention from the one used previously. The relationship between the two transforms is simply
\[
H^{(j)}(x, \alpha) = -L^{(j),*}(x, \alpha)
\]
for $j = 1, 2$. One can easily calculate
\[
I^{*}_{Q_1}(\alpha) = \begin{cases} 
-\infty, & \alpha_1 < 0 \text{ or } \alpha_2 < 0, \\
0, & \text{else.}
\end{cases}
\]

For $x$ such that $v(x) > 0$, the domains of finiteness of $L^{(j)}(x, \cdot)$ and $I_{Q_i}(\cdot)$ have nonempty intersection for $j = 1, 2$. We can therefore apply the convex duality formula for the Legendre transform of a sum ([26], Theorem 16.4) to obtain
\[
\inf_{u \in Q_1} \left[ L^{(j)}(x, u) + \langle u, \alpha \rangle \right] = \sup_{\alpha^*} \left[ L^{(j),*}(x, \alpha - \alpha^*) + I^{*}_{Q_1}(\alpha^*) \right]
\]
\[
= \sup_{\alpha_1^* \geq 0 \atop \alpha_2^* \geq 0} L^{(j),*}(x, \alpha - \alpha^*).
\]

The last equation, (5.31), (5.43) and (5.44) then imply (5.39).

The proof of (5.40) is very similar. We consider only $i = 1$, since $i = 2$ is treated in the same way. An application of the same convex duality formula gives
\[
\inf_{u \in Q_1 \setminus \{0\}} \left[ L^{(2)}(x, u) + \langle u, \alpha \rangle \right] = \sup_{\alpha_1^* \geq 0 \atop \alpha_2^* \geq 0} -H^{(2)}(x, \alpha - \alpha^*)
\]
\[
= -\inf_{\alpha_1^* \geq 0} H^{(2)}(x, (\alpha_1 - \alpha_1^*, \alpha_2)),
\]
where the last equality follows from the fact that $H^{(2)}(x, \alpha)$ is nonincreasing in $\alpha_2$ for each fixed $\alpha_1$. A second application gives
\[
\sup_{u_2 \geq 0} \inf_{(-1)^{i+1}u_1 \geq 0} \left[ L^{(1)}(x, u) + \langle u, \alpha \rangle \right]
\]
\[
= -\sup_{\alpha_2^* \geq 0} \inf_{\alpha_1^* \geq 0} H^{(1)}(x, (\alpha_1 - \alpha_1^*, \alpha_2 + \alpha_2^*))
\]
\[
= -\inf_{\alpha_1^* \geq 0} \bar{H}^{(1)}(x, (\alpha_1 - \alpha_1^*, \alpha_2)).
\]

Thus (5.40) follows from (5.33), which completes the proof. \qed
Proof of Proposition 3.2. For each fixed \( x \in \mathcal{D}^h \) and \( i \in \mathbb{N} \), any of the Jacobi and Gauss–Seidel iterations we have defined may be written in one of the following forms:

\[
V_{i+1}^h(x) = \min \left[ \inf_u \left( c^h(x, u) + \sum_y p^h(x, y|u)w(y) \right), g(x) \right]
\]

or

\[
V_{i+1}^h(x) = \min \left[ \sup_{u_2} \inf_{u_1} \left( c^h(x, u) + \sum_y p^h(x, y|u)w(y) \right), g(x) \right].
\]

Here \( c^h(x, u) \) denotes the running cost and \( w(\cdot) \) is a function that depends on the particular type of iteration used as well as (in the Gauss–Seidel case) the ordering of the states. Note that for both (5.45) and (5.46) the right-hand sides are monotonically nondecreasing in \( w(\cdot) \) if we use the partial ordering of real valued functions on \( \mathcal{D}^h \) defined by \( w_1(\cdot) \leq w_2(\cdot) \) whenever \( w_1(x) \leq w_2(x) \) for all \( x \in \mathcal{D}^h \).

First consider the Jacobi iteration. The monotonicity property just described implies that \( V_{i+1}^h \leq V_i^h \) whenever \( V_i^h \leq V_{i-1}^h \). Since the initial condition satisfies \( V_0^h \geq g \) and since \( V_1^h \leq g \), we conclude \( V_{i+1}^h \leq V_i^h \) for all \( i \in \mathbb{N} \) by induction.

Next consider the Gauss–Seidel procedure, with a possibly different ordering for each iteration. Regardless of the ordering used on the first iteration, the fact that \( V_0^h \geq g \) implies \( V_1^h \leq V_0^h \). We will again complete the proof via an induction argument. Suppose that \( V_i^h \leq V_{i-1}^h \). Let \( <_i \) denote the ordering that is used on the \( i \)th iteration. Let \( x_1, x_2, \ldots \) denote the states of \( \mathcal{D}^h \), ordered according to \( <_{i+1} \). The values \( V_i^h(y), y \in \mathcal{D}^h \), are used to define \( V_{i+1}^h(x_i) \), while the values \( V_i^h(y), y <_i x_1, \) and \( V_{i-1}^h(y), x_1 <_i y \), were used to define \( V_i^h(x_1) \). Since \( V_i^h \leq V_{i-1}^h, V_{i+1}^h(x_i) \leq V_i^h(x_1) \). We now proceed by induction according to \( <_{i+1} \). Fix \( j \) and assume \( V_{i+1}^h(x_{i+1}) \leq V_i^h(x_k) \) for \( k < j \). The values \( V_{i+1}^h(y), y <_{i+1} x_j, \) and \( V_i^h(y), x_j <_{i+1} y \), are used to define \( V_{i+1}^h(x_j) \), while the values \( V_i^h(y), y <_i x_j, \) and \( V_{i-1}^h(y), x_j <_i y \), were used to define \( V_i^h(x_j) \). In all cases the values used to define \( V_{i+1}^h(x_{j+1}) \) are no larger than those used to define \( V_i^h(x_j) \). Thus \( V_{i+1}^h(x_j) \leq V_i^h(x_j) \). By induction on \( <_{i+1} \) we conclude \( V_{i+1}^h \leq V_i^h \), and by induction on the usual ordering on \( \mathbb{N} \) we obtain the monotonicity described in part 1 of the proposition.

Since the running costs are nonnegative for the control problem of Section 2, the proved monotonicity establishes the existence of \( V^h(x) = \lim_{i \to \infty} V_i^h(x) \) for both the Jacobi and Gauss–Seidel procedures. This is also the case for the control problem of Section 3 with vertical light or for \( v(x) \geq 0 \) for oblique light. When \( v(x) < 0 \) for that control problem, we first consider the case \( w(y) = 0 \) in (5.46). A simple calculation shows that

\[
\sup_{u_2} \inf_{u_1} (L^{(1)}(x, u)) > 0
\]
and therefore,
\[
\sup_{u_2 \gamma_2 < 0} \inf_{u_1} \left( L^{(1)}(x, u) \Delta t^h(u) \right) \geq 0.
\]

Since the probabilities are nonnegative and \( \sum_y p^h(x, y|u) = 1 \), this implies
\[
\sup_{u_2 \gamma_2 < 0} \inf_{u_1} \left( L^{(1)}(x, u) \Delta t^h + \sum_y p^h(x, y|u) w(y) \right) \geq w_{\min},
\]
where \( w_{\min} = \min_y w(y) \). Thus for the control problem of Section 3 and when \( \nu(x) < 0 \), \( V^n_i \) is bounded from below by \( \min_x V^n_0(x) \). This gives part 1 of the proposition.

We next turn to part 2. Let \( \bar{V}^h \) be any fixed point of (2.14) or (3.7) that satisfies \( \bar{V}^h(x) \leq V^n_0(x) \) for all \( x \in \mathcal{Q}^h \). An argument very similar to the one used to prove part 1 shows that
\[
\bar{V}^h(x) \leq V^n_i(x) \Rightarrow \bar{V}^h(x) \leq V^n_i(x).
\]
Therefore, by induction, \( \bar{V}^h(x) \leq V^h(x) \) for all \( x \in \mathcal{Q}^h \).

Acknowledgment. We would like to thank an anonymous referee for many helpful suggestions regarding the organization of this paper.

REFERENCES


LEFSCHETZ CENTER FOR DYNAMICAL SYSTEMS
DIVISION OF APPLIED MATHEMATICS
BROWN UNIVERSITY
PROVIDENCE, RHODE ISLAND 02912

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF MASSACHUSETTS
AMHERST, MASSACHUSETTS 01003