2012

The patchy Method for the Infinite Horizon Hamilton-Jacobi-Bellman Equation and its Accuracy

Hunt, Thomas

THE PATCHY METHOD FOR THE INFINITE HORIZON HAMILTON-JACOBI-BELLMAN EQUATION AND ITS ACCURACY [*]

THOMAS HUNT[†] AND ARTHUR J. KRENER[†]

Abstract. We introduce a modification to the patchy method of Navasca and Krener for solving the stationary Hamilton Jacobi Bellman equation. The numerical solution that we generate is a set of polynomials that approximate the optimal cost and optimal control on a partition of the state space. We derive an error bound for our numerical method under the assumption that the optimal cost is a smooth strict Lyupanov function. The error bound is valid when the number of subsets in the partition is not too large.

Key words. stationary Hamilton Jacobi Bellman equation, infinite horizon optimal control problem, patchy solutions

AMS subject classifications. 49-04, 49J15, 49J20, 49L99, 49M37

1. Introduction. In this paper, we introduce a numerical scheme for solving the infinite horizon optimal control problem of minimizing the integral

$$\int_0^\infty l(x, u)dt \tag{1.1a}$$

of a Lagrangian $l(x, u)$ subject to the controlled dynamics

$$\dot{x} = f(x, u) \qquad\qquad x(0) = x^0 \tag{1.1b}$$

where f and l are both smooth, and l is strictly convex in the control $u \in \mathbb{R}^m$ for all states $x \in \mathbb{R}^n$. The solution that we are interested in is an optimal cost $\pi : \mathbb{R}^n \to \mathbb{R}$, which is the minimum value of (1.1a) incurred by driving the state from an initial value of x to the origin. The optimal control that drives the state to the origin is given in feedback form by $u = \kappa(x)$. Our solution is a set of polynomials that approximate the optimal cost and optimal control on a partition of a subset of the state space.

Our method is a modification of the original patchy method [7], which we altered to obtain the error bound in Theorem 3.6. Like the original patchy method, it is an extension of the Cauchy-Kovalevskaya method [4], the fast marching method [13, 9], the patchy technique of Ancona and Bressan [2], and Al'brekht's power series method [1].

Al'brekht's method is an algorithm for computing a series solution to the optimal cost that must be centered at the origin of the state space. A drawback to the requirement that the series solution be centered at the origin is, even when the optimal cost and optimal control are smooth over the entire domain of interest, their power series solutions may be local in nature. This means that outside of some region containing the origin, it may be impractical to compute the number of terms necessary to achieve a desired accuracy. The patchy method that we present uses Al'brekht's series solution as an initialization step, and generates a new series solution centered at a point away from the origin. Doing this enlarges the region where we have a valid approximation to the optimal cost and optimal control.

[†]Naval Postgraduate School, 833 Dyer Road Monterey, CA 93943-5216 (twhunt@nps.edu, ajkrener@nps.edu)

1.1. The Hamilton-Jacobi-Bellman PDE. It is widely known that if the optimal control problem (1.1) has a smooth optimal cost $\pi(x)$, and the optimal control can be put in feedback form $u = \kappa(x)$, then the optimal control and optimal cost satisfy the Hamilton-Jacobi-Bellman (HJB) equation

$$0 = \min_u \left(\frac{\partial \pi}{\partial x}(x) f(x, u) + l(x, u) \right)$$
$$\kappa(x) = \arg\min_u \left(\frac{\partial \pi}{\partial x}(x) \frac{\partial f}{\partial u}(x, u) + \frac{\partial l}{\partial u}(x, u) \right) \tag{1.2}$$

locally around the origin. Furthermore, if $\frac{\partial \pi}{\partial x}(x) f(x, u) + l(x, u)$ is strictly convex in u locally around $x = 0$, $u = 0$, then the HJB PDEs (1.2) simplify to

$$0 = \frac{\partial \pi}{\partial x}(x) f(x, \kappa(x)) + l(x, \kappa(x)) \tag{1.3a}$$

$$0 = \frac{\partial \pi}{\partial x}(x) \frac{\partial f}{\partial u}(x, \kappa(x)) + \frac{\partial l}{\partial u}(x, \kappa(x)). \tag{1.3b}$$

1.2. Solutions to the Hamilton-Jacobi-Bellman PDE. The existence of a solution of the optimal control problem (1.1) around the origin is determined by the leading order terms of the problem data f and l. Suppose the dynamics and Lagrangian have Taylor expansions

$$\dot{x} = Fx + Gu + \sum_{k=2}^{\infty} f^{[k]}(x, u)$$

$$l(x, u) = \frac{1}{2}\left(x^T Q x + u^T R u \right) + \sum_{k=3}^{\infty} l^{[k]}(x, u) \tag{1.4}$$

where $^{[k]}$ denotes degree k terms in the power series, R is positive definite, and Q is positive semidefinite. The optimal control problem is said to be *nice* if (F, G) is stabilizable and $(Q^{1/2}, F)$ is controllable.

Al'brekht [1] showed that when the optimal control problem (1.3) is nice, the optimal cost and optimal control have series solution centered at the origin of the state space. Lukes [6] showed that both these series solutions converge to the optimal cost and control under suitable conditions. Al'brekht's method is an algorithm for computing both these series. It works by computing the degree j and $j-1$ terms as a pair in the expansions of the optimal cost and optimal control, starting at $j = 2$ and terminating at some user specified degree. The original description of the algorithm is in Al'brekht's original work [1], and a generalization with notation consistent with this paper is in [7]. The Matlab implementation of Al'brekht's method that we use in the patchy method is available by request [5].

2. Derivation of the patchy method.

2.1. Optimal control problem assumptions. We will derive the patchy method for the special case of the optimal control problem where the control is scalar ($m = 1$), and the dynamics and lagrangian have the form

$$\dot{x} = f(x) + g(x)u \qquad\qquad l(x, u) = q(x) + \frac{1}{2}r(x)u^2 \tag{2.1}$$

where

$$f, g : \mathbb{R}^n \to \mathbb{R}^n \qquad\qquad q, r : \mathbb{R}^n \to \mathbb{R} \qquad\qquad u \in \mathbb{R}.$$

2

The patchy method does not require these assumptions, but they do significantly simplify its derivation. Furthermore, we assume

1. f, g, q, r have continuous partial derivatives of to degree $d + 2$ on X_c.
2. The optimal control problem is "nice", meaning it can be expressed in the form of (1.4), and (F, G) is stabilizable and $(F, Q^{1/2})$ is detectable. We also assume that Q is strictly positive definite.
3. π has continuous partial derivatives up to order $d + 2$ on X_c
4. $q \geq 0$ and $r > 0$ on X_c
5. The optimal cost is a strict Lyapunov function on X_c, meaning

$$\frac{\partial \pi}{\partial x}(x)\dot{x} = \frac{\partial \pi}{\partial x}(x)\big(f(x) + g(x)\kappa(x)\big) < 0 \tag{2.2}$$

for all x in $X_c \setminus \{0\}$.

Assumptions 1 and 2 guarantee that assumption 3 holds in some neighborhood of the origin. The Taylor series of π in this neighborhood can be computed by Al'brekht's method. The fifth assumption is essential in both the derivation of the patchy method and the proof of its error bound in Theorem 3.6.

2.2. Overview of the patchy method. Although the method works in higher dimensions, we will assume in the overview that the state space is two dimensional. This allows us to illustrate how we partition the state space. We first compute the degree $d + 1$ Taylor approximation to the optimal cost at the origin by Al'brekht's method, which we denote π^0. We assume that all truncated series solutions to the optimal cost and optimal control are polynomials of degree $d + 1$ and d. We then pick a sublevel set $\pi^0 \leq c$ on which π^0 has an acceptable level of error. We refer to this sublevel set as the *Al'brekht patch* and denote it as $\mathcal{A}$. The patchy method expands the domain of the numerical solution to a superset of the Al'brekht patch, which we refer to as the computational domain and denote it as X_c. The patchy method picks a *patch point*, denoted x^i, on the boundary of the Al'brekht patch and then computes an approximating polynomial to the optimal cost that is centered at the new patch point, which we denote as π^i. The core of the patchy algorithm is a method to compute the approximate partial derivatives of the optimal cost at x^i, thus computing an approximation to the Taylor polynomial centered at x^i of the optimal cost. The patchy method repeats this process until the origin is surrounded with a number of patch points.

The computational domain is partitioned into a set of *patches* during the course of the patchy algorithm. Each patch point is associated with a single patch, and the optimal cost and optimal control at any state that falls inside in the patch are computed by the approximating polynomials centered at the associated patch point. Every point in the computational domain is associated with exactly one patch, and we compute the approximate optimal cost and optimal control at the point by the series solutions centered at the associated patch point. In two dimensions, the boundaries of the patch are the boundary of the Al'brekht patch, a level curve of the optimal cost centered at the patch point, and two *lateral boundaries* that intersect the other two boundaries. Illustrations of the Al'brekht patch and a typical patch in a two dimensional state space are included in Figure 2.1.

Once the Patchy method has encircled the Al'brekht patch with new patches, it generates a new ring of patches and associated optimal cost polynomials, where the new patches form a ring that encloses the old set of patches.

As we will see, under some standard assumptions on the underlying optimal control problem, as well as the assumptions that the true solution is both sufficiently smooth and a strict Lyapunov function, we can derive error estimates on the patchy method. Unfortunately, these error estimates do not go to zero as the density of patches on the computational domain goes to infinity. To our knowledge, the only algorithm that provably solves the nonlinear Hamilton-Jacobi-Bellman equations in multiple dimensions with higher order accuracy is the finite difference scheme due to Szpiro and Dupuis [12]. Their method is provably second order accurate on subsets of the computational domain where the true solution is smooth.

2.3. Placing a new patch point. Given the *previous* patch point x^i and its associated approximate optimal cost π^i, the patchy method must determine where in the state space to place the *new* patch point x^{i+1}. If the computed optimal cost π^i is sufficiently close to the true optimal cost in a neighborhood of x^i, then by the assumption that the optimal cost is a strict Lyapunov function, the closed loop dynamics is not tangent to a level curve of π^i in the neighborhood. Thus the partial derivatives of the optimal cost can be calculated by the technique of this section at a point on an appropriately chosen level curve of π^i. The patchy method places x^{i+1} on a level curve of π^i under the constraint that the distance between x^i and x^{i+1} is less than h, the *maximum consecutive patch point distance*. We derive the maximum consecutive patch point distance in chapters 4 and 5.

2.4. Computing the optimal cost at a new patch point. Away from the origin, the HJB equations alone do not fully specify all the partial derivatives of the optimal cost at a given patch point. As a consequence, the patchy algorithm computes some partial derivatives of the optimal cost from the HJB equations, and computes the remaining partial derivatives from the previous optimal cost.

The computed optimal cost at the new patch point x^{i+1} is calculated by *inheritance*, meaning we set

$$\pi^{i+1}(x^{i+1}) = \pi^i(x^{i+1})$$

2.5. Computing the first order partial derivatives of the optimal cost away from the origin. First, we compute $\mathbf{n}$, the gradient direction of π^{i+1} at x^{i+1} from π^i by

$$\mathbf{n} \equiv \frac{1}{\left\|\frac{\partial \pi^i}{\partial x}(x^{i+1})\right\|} \frac{\partial \pi^i}{\partial x}(x^{i+1}). \tag{2.3}$$

If the level sets of the previous and new optimal costs are tangent, then the gradients of the previous and new optimal cost are collinear at the current patch point. In this case, there is some positive scalar z such that

$$\frac{\partial \pi^{i+1}}{\partial x}(x^{i+1}) = z\mathbf{n} \tag{2.4}$$

To compute z, we first define the scalars

$$f_n \equiv \mathbf{n} \cdot f(x^{i+1}) \qquad \text{and} \qquad g_n \equiv \mathbf{n} \cdot g(x^{i+1}). \tag{2.5}$$

We can derive a formula for the optimal control in terms of the gradient of the optimal cost by solving (1.3b) under the assumption (2.1). If we substitute this formula for

the optimal control into (1.3a), then it reduces to the scalar quadratic equation

$$0 = -\frac{g_n^2}{2r(x^{i+1})}z^2 + f_n z + q(x^{i+1}) \tag{2.6}$$

under (2.4) and (2.5). If the computed and exact gradients of the optimal cost are sufficiently close, then it follows from the assumption that the exact optimal cost is a strict Lyapunov function that the quadratic equation (2.6) has exactly one strictly positive root. We denote this positive root as z_+, so the computed gradient of the new optimal cost and the computed new optimal control at the new patch point are

$$\frac{\partial \pi^{i+1}}{\partial x}(x^{i+1}) = z_+\mathbf{n} \qquad \kappa^{i+1}(x^{i+1}) = -\frac{1}{r(x^{i+1})}\frac{\partial \pi^{i+1}}{\partial x}(x^{i+1})g(x^{i+1}) \tag{2.7}$$

2.6. Computing higher order partial derivatives of the optimal cost away from the origin. Let $\dot{x}^{i+1}$ denote the computed *optimal direction* at the new patch point calculated from the first order partial derivatives of the computed optimal cost. If the exact and computed gradients of the optimal cost are sufficiently close at the new patch point, then it follows from the assumption that the optimal cost is a strict Lyapunov function that the computed optimal direction is nonzero. The computed optimal direction is calculated from the formula

$$\dot{x}^{i+1} = f(x^{i+1}) + g(x^{i+1})\kappa^{i+1}(x^{i+1}).$$

We set

$$\hat{V}^1 = \frac{1}{\|\dot{x}^{i+1}\|}\dot{x}^{i+1} \qquad\qquad \hat{V} \equiv \begin{bmatrix} \hat{V}^1 \cdots \hat{V}^n \end{bmatrix}$$

where $\hat{V}^1, \hat{V}^2, \ldots, \hat{V}^n$ form an orthonormal basis of $\mathbb{R}^n$. This can be done with a Householder reflector [3, ch. 5.1.2]. We will compute the higher order partial derivatives of the optimal cost under the change of variables

$$x = x^{i+1} + \hat{V}\hat{\xi}$$

and then recover the partial derivatives with respect to the original state space variables. In the new variables, the HJB equations (1.3) become

$$0 = \frac{\partial \hat{\pi}}{\partial \hat{\xi}}(\hat{\xi})\big(\hat{f}(\hat{\xi}) + \hat{g}(\hat{\xi})\hat{\kappa}(\hat{\xi})\big) + \hat{q}(\hat{\xi}) + \frac{1}{2}\hat{r}(\hat{\xi})\big(\hat{\kappa}(\hat{\xi})\big)^2 \tag{2.8a}$$

$$0 = \frac{\partial \hat{\pi}}{\partial \hat{\xi}}(\hat{\xi})\hat{g}(\hat{\xi}) + \hat{r}(\hat{\xi})\hat{\kappa}(\hat{\xi}) \tag{2.8b}$$

where

$$\begin{aligned}
\hat{\pi}(\hat{\xi}) &\equiv \pi(x^{i+1} + \hat{V}\hat{\xi}) & \hat{\kappa}(\hat{\xi}) &\equiv \kappa(x^{i+1} + \hat{V}\hat{\xi}) \\
\hat{f}(\hat{\xi}) &\equiv \hat{V}^T f(x^{i+1} + \hat{V}\hat{\xi}) & \hat{g}(\hat{\xi}) &\equiv \hat{V}^T g(x^{i+1} + \hat{V}\hat{\xi}) \\
\hat{q}(\hat{\xi}) &\equiv q(x^{i+1} + \hat{V}\hat{\xi}) & \hat{r}(\hat{\xi}) &\equiv r(x^{i+1} + \hat{V}\hat{\xi})
\end{aligned} \tag{2.9}$$

To derive a formula for the *characteristic* second order partial derivatives of the optimal cost, we evaluate the derivative of the first HJB equation in (2.8a) with respect to $\hat{\xi}_j$ at $\hat{\xi} = 0$, and arrive at

$$0 = \|\dot{x}\|\frac{\partial^2 \hat{\pi}}{\partial \hat{\xi}_j \partial \hat{\xi}_1}(0) + \frac{\partial \hat{\pi}}{\partial \hat{\xi}}(0)\Big(\frac{\partial \hat{f}}{\partial \hat{\xi}_j}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_j}(0)\hat{\kappa}(0)\Big) + \frac{\partial \hat{q}}{\partial \hat{\xi}_j}(0) + \frac{1}{2}\frac{\partial \hat{r}}{\partial \hat{\xi}_j}(0)\big(\hat{\kappa}(0)\big)^2 \tag{2.10}$$

where $\dot{x}$ denotes the optimal direction at x^{i+1}. The terms involving the partial derivative of $\hat{\kappa}$ with respect to $\hat{\xi}_j$ drop out due to (2.8b). We compute the *characteristic* second order partial derivatives of $\hat{\pi}^{i+1}$ by substituting in the computed approximations to the optimal direction and gradient of the optimal cost into (2.10), and solving for the unknown second order partial derivatives of the optimal cost. This yields the formula

$$\frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_j \partial \hat{\xi}_1}(0) = -\frac{1}{\|\dot{x}^{i+1}\|}\left[\frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0)\left(\frac{\partial \hat{f}}{\partial \hat{\xi}_j}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_j}(0)\hat{\kappa}^{i+1}(0)\right) + \frac{\partial \hat{q}}{\partial \hat{\xi}_j}(0)\right.$$
$$\left. + \frac{1}{2}\frac{\partial \hat{r}}{\partial \hat{\xi}_j}(0)\left(\hat{\kappa}^{i+1}(0)\right)^2\right]. \quad (2.11)$$

The characteristic second order partial derivatives are a strict subset of all the second order partial derivatives, since the HJB equations do not specify all the second order partial derivatives. We call the remaining second order partial derivatives *non characteristic*. We compute them by inheritance, meaning for $2 \le j_1 \le j_2 \le n$ we set

$$\frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) = \frac{\partial^2}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}\left[\pi^i(x^{i+1} + \hat{V}\hat{\xi})\right]_{\hat{\xi}=0}. \quad (2.12)$$

To get the approximate first order partial derivatives of the optimal control, we follow the same process. We differentiate the second HJB equation(2.8b) with respect to $\hat{\xi}_j$ at $\hat{\xi} = 0$, and after substituting in the computed partial derivatives of the optimal control and cost, we obtain the formula

$$\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_j}(0) = -\frac{1}{\hat{r}(0)}\left[\frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_j \partial \hat{\xi}}(0)\hat{g}(0) + \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0)\frac{\partial \hat{g}}{\partial \hat{\xi}_j}(0)\hat{\kappa}^{i+1}(0) + \frac{\partial \hat{r}}{\partial \hat{\xi}_j}(0)\hat{\kappa}^{i+1}(0)\right].$$
$$(2.13)$$

We compute the remaining higher order partial derivatives in an analogous fashion. Their formulas are in §9.2.

We now introduce new notation to describe how to recover the partial derivatives with respect to the original state space variables from the partial derivatives with respect to the new variables. They are computed by

$$\frac{\partial^2 \pi^{i+1}}{\partial x_{j_1} \partial x_{j_2}}(x^{i+1}) = \frac{\partial}{\partial \hat{\xi}} \otimes \frac{\partial}{\partial \hat{\xi}}\left[\hat{\pi}^{i+1}(\hat{\xi})\right]_{\hat{\xi}=0}(\hat{V}^T e_{j_1} \otimes \hat{V}^T e_{j_2})$$
$$\vdots \quad (2.14)$$
$$\frac{\partial^k \pi^{i+1}}{\partial x_{j_1} \cdots \partial x_{j_k}}(x^{i+1}) = \frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}}\left[\hat{\pi}^{i+1}(\hat{\xi})\right]_{\hat{\xi}=0}(\hat{V}^T e_{j_1} \otimes \cdots \otimes \hat{V}^T e_{j_k})$$

where e_j is the jth column of the identity matrix. Each formula in (2.14) is a standard dot product, written as the product of a row and column vector. The symbol $\otimes$ has two meanings in (2.14). It is the standard Kronecker product if it appears between two matrices. The row vector $\frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}}[\hat{\pi}^{i+1}(\hat{\xi})]$ is new notation for the *Kronecker derivative*, which is a bookkeeping mechanism. In the first equation in (2.14), if the state space is two dimensional then the Kronecker derivative is defined as

$$\frac{\partial}{\partial \hat{\xi}} \otimes \frac{\partial}{\partial \hat{\xi}}\left[\hat{\pi}^{i+1}(\hat{\xi})\right]_{\hat{\xi}=0} = \left[\frac{\partial^2 \hat{\pi}}{\partial \hat{\xi}_1^2}(\hat{\xi}) \quad \frac{\partial^2 \hat{\pi}}{\partial \hat{\xi}_1 \partial \hat{\xi}_2}(\hat{\xi}) \quad \frac{\partial^2 \hat{\pi}}{\partial \hat{\xi}_1 \partial \hat{\xi}_2}(\hat{\xi}) \quad \frac{\partial^2 \hat{\pi}}{\partial \hat{\xi}_2^2}(\hat{\xi})\right]_{\hat{\xi}=0}.$$

6

We give the full definition of the Kronecker derivative in the appendix.

2.7. Computing the optimal cost and optimal control away from a patch point. To compute the optimal cost or optimal control at a point in the state space that is not a patch point, we need a means to determine which patch the point belongs to. In other words, we need to know which polynomials π^i and κ^i to use to calculate the optimal cost and optimal control at some $\bar{x}$ in the state space. There are two types of patches, *typical* patches, and a single Al'brekht patch. The boundary of the Al'brekht patch is a level curve of π^0 that encloses the entire patch. The optimal cost and optimal control at a point inside the Al'brekht patch are computed from π^0 and κ^0. A typical patch is associated with the patch point $x^{i+1} \neq 0$, and the optimal cost and optimal control at every point in the patch is computed from π^{i+1} and κ^{i+1}. When the state space is two dimensional, a typical patch has four intersecting boundaries that enclose the patch. Two of the boundaries are level curves of the optimal cost, and the other two boundaries are *lateral* boundaries. The patch point x^{i+1} lies on the optimal cost level curve of π^i that defines one of the patch boundaries. The other optimal cost level curve patch boundary is a level curve of π^{i+1}. Each lateral boundary is a straight line segment that intersects the level curve of π^i at a point x_b, and is oriented so that it points in the opposite direction of the optimal direction $f(x_b) + g(x_b)\kappa^i(x_b)$. When the state space is two dimensional, the *patch boundary point* x_b is placed on a level curve of the optimal cost so that it is halfway between two adjacent patch points that lie on the same level curve. The optimal direction at x_b may be computed from the approximate optimal cost associated with either adjacent patch point. Illustrations of the two types of patches are in Figure 2.1.
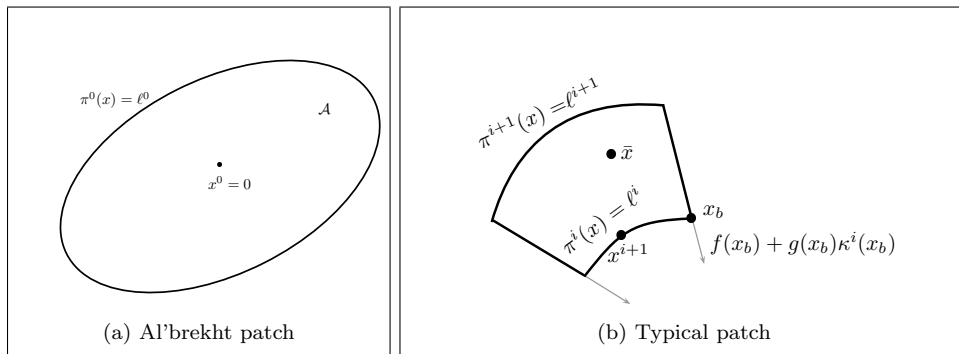

(a) Al'brekht patch (b) Typical patch

FIG. 2.1. *Al'brekht patch and typical patch*

3. Analytical error bound for the patchy method. The difference between the exact and computed optimal cost grows as a power of the maximum consecutive patch point distance times a term that grows exponentially as the patchy method moves along a sequence of consecutive patch points. We first develop some concepts and lemmas that we will need before making the more rigorous statement of the error bound in Theorem 3.6 at the end of this section.

3.1. Sequence of consecutive patch points. The error bound is predicated on the notion of a *sequence of consecutive patch points*.

DEFINITION 3.1. *The patch points x^i and x^{i+1} are said to be* consecutive *if the computed optimal cost polynomial centered at x^{i+1} was computed from information*

contained in the computed optimal cost polynomial centered at x^i. The origin is always the first patch point in a sequence of consecutive patch points. A sequence of three consecutive patch points is illustrated in Figure 3.1.
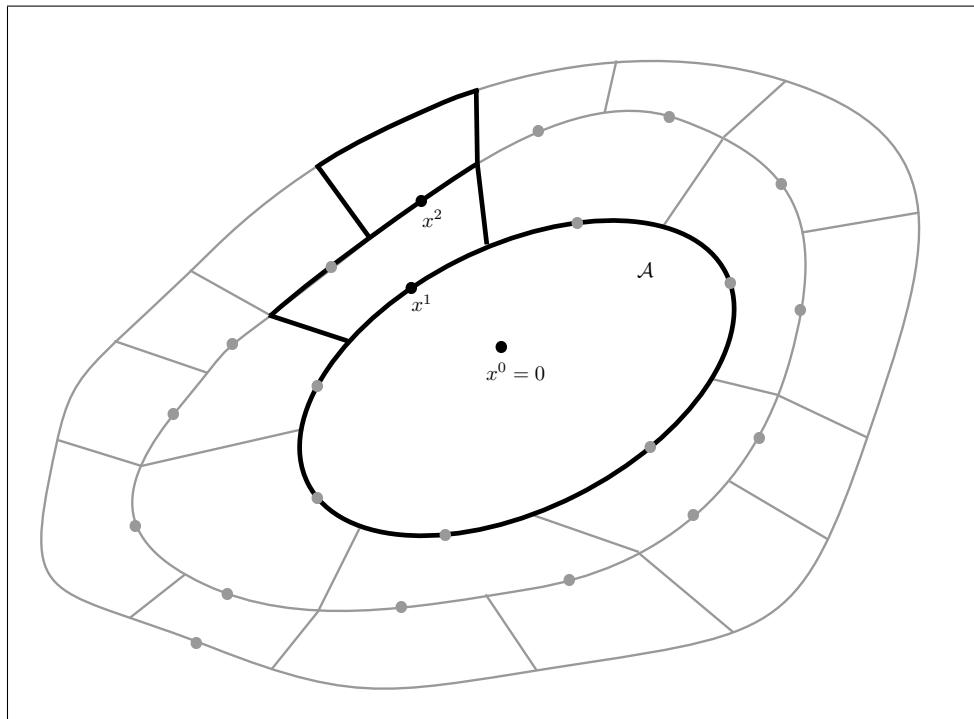

FIG. 3.1. *A sequence of three consecutive patch points x^0, x^1, x^2*

3.2. Notation. The patchy algorithm takes a truncated series solution to the optimal cost centered at the previous patch point, and generates a new series solution to the optimal cost centered at the next patch point. Thus, it can be interpreted as a mapping from a set of polynomial coefficients associated with the previous patch point to a set of polynomial coefficients associated with the new patch point. These coefficients are approximate partial derivatives of the optimal cost evaluated at the appropriate patch point, so we introduce notation to reflect this.

$\hat{C}_j^i$ denotes the vector of computed j^{th} partial derivatives of the optimal cost at x^i. For example, if the state space is two dimensional

$$\hat{C}_2^i = \left[\frac{\partial^2 \hat{\pi}}{\partial x_1^2}(x^i) \quad \frac{\partial^2 \hat{\pi}}{\partial x_1 \partial x_2}(x^i) \quad \frac{\partial^2 \hat{\pi}}{\partial x_2^2}(x^i) \right]$$

$\hat{C}^i$ is the vector of all partial derivatives of the computed optimal cost at x^i up to order $d+1$, that is

$$\hat{C}^i = \begin{bmatrix} \hat{C}_0^i & \hat{C}_1^i & \cdots & \hat{C}_{d+1}^i \end{bmatrix}$$

We denote with ϕ_j the part of the patchy method that computes the jth order partial derivatives of the optimal cost at the current patch point from the partial derivatives of the optimal cost evaluated at the previous patch point, meaning

$$\hat{C}_j^{i+1} = \phi_j(\hat{C}^i)$$

C_j^i denotes the vector of exact j^{th} partial derivatives of the optimal cost π at x^i, and C^i has a definition that is analogous to $\hat{C}^i$.

3.3. Local truncation vector. DEFINITION 3.2. C^i *denotes the vector of exact partial derivatives of the optimal cost at* x^i, *and* C_j^{i+1} *denotes the exact partial derivatives of order* j *of the optimal cost at* x^{i+1}, *where* x^i *and* x^{i+1} *are consecutive patch points. The local truncation vector* τ_j^i *is*

$$\tau_j^i \equiv C_j^{i+1} - \phi_j(C^i)$$

which is the difference between the exact j^{th} *order partial derivatives of* π *at* x^{i+1}, *and the computed partial derivatives of* π *at* x^{i+1} *that are generated by feeding* ϕ_j *the exact partial derivatives of of* π *centered at* x^i.

3.4. The Lipschitz condition. DEFINITION 3.3. *Suppose* C_j^i *and* $\hat{C}_j^i$ *are vectors of exact and approximate partial derivatives of order* j *of the optimal cost* π *at* x^i. *The function* ϕ *is said to satisfy the Lipschitz condition if there exists a* maximum *multiplier* $M_{max} > 0$ *and a* maximum *consecutive patch point distance* $h > 0$, *such that for all* $M \leq M_{max}$ *and all* $\|x^{i+1} - x^i\| \leq h$, *then*

$$\|\phi_k(C^i) - \phi_k(\hat{C}^i)\|_2 \leq LM\|x^{i+1} - x^i\|_2^{d+2-k} \text{ for } 0 \leq k \leq d+1$$

whenever

$$\|C_j^i - \hat{C}_j^i\|_2 \leq M\|x^{i+1} - x^i\|_2^{d+2-j} \text{ for } 0 \leq j \leq d+1$$

The following lemma will help us establish that the patchy method satisfies the Lipschitz condition.

LEMMA 3.4. *Polynomials satisfy the Lipschitz condition of Definition 3.3 with* $M_{max} \leq \infty$. *We may write any degree* $d+1$ *polynomial as*

$$P(C, x) \equiv \sum_{\ell=0}^{d+1} C_\ell \cdot m_\ell(x).$$

where C_ℓ *is a coefficient vector as defined in §3.2 and* $m_\ell(x)$ *is a vector of* ℓ^{th} *degree monomials so that all the degree* ℓ *terms of the polynomial are* $C_\ell \cdot m_\ell(x)$. *Thus, if* $\|C_\ell - \hat{C}_\ell\|_2 \leq M\|x\|_2^{d+2-\ell}$, *then there exists* $L_{inh} < \infty$ *such that*

$$\left| \frac{\partial^k P}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(C, x) - \frac{\partial^k P}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(\hat{C}, x) \right| \leq L_{inh} M\|x\|_2^{d+2-k}$$

Proof. Since $P(C, x) - P(\hat{C}, x) = P(C - \hat{C}, x)$, *we may assume without a loss of generality that* $\hat{C} = 0$, *and* $\|C_\ell\|_2 \leq M\|x\|_2^{d+2-\ell}$. *By Cauchy-Schwarz,*

$$\left| \frac{\partial^k P}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(C, x) \right| \leq \sum_{\ell=k}^{d+1} \|C_\ell\| \left\| \frac{\partial^k m_\ell}{\partial x_{j_1} \partial x_{j_2} \cdots \partial x_{j_k}}(x) \right\|.$$

Each entry in the kth *order partial derivative of* m_ℓ *is* $\mathcal{O}(\|x\|^{\ell-k})$, *and by assumption* $\|C_\ell\|_2 \leq M\|x\|_2^{d+2-\ell}$, *so each term in the sum is* $\mathcal{O}(\|x\|^{d+2-k})$. $\square$

The next lemma establishes a bound on the difference between the Taylor polynomial coefficients of the optimal cost π, and the coefficients of the computed optimal cost π^i.

LEMMA 3.5. *Suppose there exists a maximum consecutive patch point distance $h > 0$ such that ϕ satisfies the Lipschitz condition with constants L and M_{max}. Also, suppose there exists a local truncation error constant T such that for any patch points x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a radius of h of each other,*

$$\|\tau_j^i\| \leq Th^{d+2-j}$$

where the approximating polynomials to the optimal cost are degree $d + 1$. Let $\hat{C}_j^i$ denote the vector of jth order partial derivatives of the optimal cost computed by the patchy method, where i is the sequence index for a sequence of consecutive patch points. Let C_j^i denote the corresponding vector of exact partial derivatives. The growth in the difference between the true and computed j^{th} partial derivatives of the optimal cost is bounded by

$$\|C_j^i - \hat{C}_j^i\| \leq \frac{L^i - 1}{L - 1}Th^{d+2-j} \text{ for } 1 \leq i \leq N$$

where N is the smallest integer such that

$$\frac{L^N - 1}{L - 1}T \leq M_{max}$$

Proof. The proof proceeds by induction on the sequence index i, and is similar in spirit to the construction of an error bound between the true solution of an ODE and a solution computed by a one step method [11, ch. 7.2]. We assume that all arithmetic is exact. It follows from the definition of the local truncation vector that the difference between the true and computed j^{th} order partial derivatives is bounded by

$$\|C_j^{i+1} - \hat{C}_j^{i+1}\| \leq \|\phi_j(C^i) - \phi_j(\hat{C}^i)\| + \|\tau_j^i\|$$

At the origin, the patchy algorithm computes the exact partial derivatives of the optimal cost by Al'brekht's method, so $\hat{C}_j^0 = C_j^0$ for $0 \leq j \leq d+1$. Therefore the theorem holds in the base case of the induction proof. Now, suppose that at the ith patch point in the sequence of consecutive patch points,

$$\|C_j^i - \hat{C}_j^i\| \leq \frac{L^i - 1}{L - 1}Th^{d+2-j} \qquad and \qquad \frac{L^i - 1}{L - 1}T \leq M_{max}$$

for $0 \leq j \leq d+1$. Then

$$\|\phi_j(C^i) - \phi_j(\hat{C}^i)\| \leq L\frac{L^i - 1}{L - 1}Th^{d+2-j}$$

for $0 \leq j \leq d+1$ since ϕ satisfies the Lipschitz condition. Therefore, at the next patch point in the sequence, the difference between the partial derivatives is bounded by

$$\|C_j^{i+1} - \hat{C}_j^{i+1}\| \leq L\frac{L^i - 1}{L - 1}Th^{d+2-j} + Th^{d+2-j} = \frac{L^{i+1} - 1}{L - 1}Th^{d+2-j}.$$

Thus, the conclusion of the lemma holds. □

3.5. Analytical bound on absolute error of the optimal cost. THEOREM 3.6. *Suppose there exists a maximum consecutive patch point distance $h > 0$ such that ϕ satisfies the Lipschitz condition. Also, suppose there exists a local truncation error constant T such that for any consecutive patch points x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a radius h of each other, $\|\tau_j^i\| \leq Th^{d+2-j}$, where the degree of all the optimal cost approximating polynomials is $d + 1$. Finally, assume that the maximum patch diameter is proportional to the maximum consecutive patch point distance. Then there exist constants L and K such that for any x in the ith patch in a sequence, the error between the exact and computed optimal cost is bounded by*

$$|\pi(x) - \pi^i(x)| \leq K\Big(\frac{L^i - 1}{L - 1} + 1\Big)h^{d+2}$$

and the difference between corresponding k^{th} order partial derivatives is bounded by

$$\left|\frac{\partial^k \pi}{\partial x_{j_1} \cdots \partial x_{j_k}}(x) - \frac{\partial^k \pi^i}{\partial x_{j_1} \cdots \partial x_{j_k}}(x)\right| \leq K\Big(\frac{L^i - 1}{L - 1} + 1\Big)h^{d+2-k}.$$

Proof. The basic idea behind the proof is to write the true optimal cost as the sum of its Taylor polynomial and a remainder term, then apply Lemma 3.5 to bound the difference between the coefficients of π^i and the coefficients of the Taylor polynomial. By Lemma 3.5, there is some Lipschitz constant L such that the exact and computed k^{th} order partial derivatives of the optimal cost at x^i are bounded by

$$\|C_k^i - \hat{C}_k^i\| \leq \frac{L^i - 1}{L - 1}Th^{d+2-k}$$

where T is the local truncation constant from the theorem statement. Let $\bar{x}$ be any point in the ith patch in a sequence of consecutive patches, and let D denote the maximum patch diameter for all the patches in the computational domain X_c. By Lemma 3.4, there exists L_{inh} such that

$$\left|\frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}}\big[P(C^i - \hat{C}^i, x - x^i)\big]_{x=\bar{x}}\right| \leq L_{inh}\frac{L^i - 1}{L - 1}TD^{d+2-k}$$

By assumption, the optimal cost is smooth on X_c, so it follows from the definition of the Taylor remainder constant R_T that

$$\left|\frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}}\big[\pi(x - x^i) - P(C^i, x - x^i)\big]_{x=\bar{x}}\right| \leq R_T D^{d+2-k}.$$

Therefore

$$\left|\frac{\partial^k \pi}{\partial x_{j_1} \cdots \partial x_{j_k}}(x) - \frac{\partial^k \pi^i}{\partial x_{j_1} \cdots \partial x_{j_k}}(x)\right|$$

$$\leq \left|\frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}}\big[P(C^i - \hat{C}^i, x - x^i)\big]_{x=\bar{x}}\right|$$

$$+ \left|\frac{\partial^k}{\partial x_{j_1} \cdots \partial x_{j_k}}\big[\pi(x - x^i) - P(C^i, x - x^i)\big]_{x=\bar{x}}\right|$$

$$\leq \Big(L_{inh}\frac{L^i - 1}{L - 1}T + R_T\Big)\Big(\frac{D}{h}\Big)^{d+2-k}h^{d+2-k}.$$

11

The error bound from the theorem statement follows by setting $K = (D/h)^{d+2} \max(TL_{inh}, R_T)$. ∎
Our assumption that the patch diameter D is proportional to the maximum consecutive patch point distance h guarantees that D/h is bounded from above by a constant.
∎

In sections 4-7, we list sufficient conditions that guarantee that the local truncation error of the patchy algorithm has the desired order, and the patchy algorithm satisfies the Lipschitz condition, thus guaranteeing the error bound in the previous theorem.

4. Local truncation error for first order partial derivatives of the optimal cost. We state Theorem 4.1 whose content implies the first order local truncation vector is bounded by $\|\tau_1^i\| \leq Th^{d+1}$, and defer its proof until after we have established some supporting lemmas. In this section, we always assume that the conditions on the optimal control problem in §2.1 hold.

THEOREM 4.1. *Suppose π solves the HJB equations (1.3) under the assumptions in §2.1. Let $\bar{\pi}$ denote the Taylor polynomial of the optimal cost centered at the previous patch point, and let $\frac{\partial \pi^{i+1}}{\partial x}(x^{i+1})$ denote the computed gradient of the optimal cost at x^{i+1}, computed from the coefficients of $\bar{\pi}$. There exists a ball centered at the origin in which the Al'brekht patch must lie, a maximum consecutive patch point distance, and a constant T such that*

$$\left\|\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \pi^{i+1}}{\partial x}(x^{i+1})\right\| \leq T\left\|\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\right\|. \tag{4.1}$$

whenever the distance from x^{i+1} to the previous patch point is less than the maximum consecutive patch point distance . The maximum consecutive patch point distance h_T must be less than the consecutive patch point distance that we derive in Lemma 4.3, and small enough to ensure $\|\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\| \leq M_\Delta$, where

$$M_\Delta < \inf_{X_c \backslash \mathcal{A}} \left\|\frac{\partial \pi}{\partial x}(x)\right\|$$

$$M_\Delta < \frac{\inf_{X_c \backslash \mathcal{A}} |\frac{\partial \pi}{\partial x}(x)f(x)|}{\sup_{X_c \backslash \mathcal{A}} \|f(x)\|}$$

$$M_\Delta < \frac{\inf_{X_c \backslash \mathcal{A}} |\frac{\partial \pi}{\partial x}(x)g(x)|}{\sup_{X_c \backslash \mathcal{A}} \|g(x)\|}$$

$$M_\Delta < \frac{1}{2} \sup_{X_c \backslash \mathcal{A}} \left(\left\|\frac{\partial \pi}{\partial x}(x)\right\| \left(\|f(x)\|^2 + 2\frac{q(x)}{r(x)}\|g(x)\|^2\right)\right.$$

$$\left. + \|f(x)\|^2 + \|g(x)\|^2\right) \inf_{X_c \backslash \mathcal{A}} \left(\frac{\partial \pi}{\partial x}(x)\dot{x}\right)^2$$

The assumptions in §2.1 guarantee that each right hand side of the previous four inequalities is strictly positive.

COROLLARY 4.2. *There exists a maximum consecutive patch point distance h_T and local truncation constant T such that for any two consecutive patch points x^i and x^{i+1} in the computational domain X_c within a distance of h_T of each other,*

$$\left\|\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \pi^{i+1}}{\partial x}(x^{i+1})\right\| \leq T\|x^{i+1} - x^i\|^{d+1}. \tag{4.2}$$

where $\frac{\partial \pi^{i+1}}{\partial x}(x^{i+1})$ denotes the computed gradient of the optimal cost at x^{i+1}, computed from the coefficients of the Taylor polynomial centered at the previous patch point.

Equivalently

$$\|\tau_1^i\| = \left\|C_1^{i+1} - \phi_1(C^i)\right\| \leq T\|x^{i+1} - x^i\|^{d+1}.$$

It is helpful to think of the quantities f_n and g_n from (2.5), and the solution to the quadratic equation z from (2.6) as functions of a nonzero vector w, so we define

$$f_n(w) \equiv \frac{w}{\|w\|} \cdot f(x^{i+1})$$

$$g_n(w) \equiv \frac{w}{\|w\|} \cdot g(x^{i+1})$$

$$z(w) \equiv \begin{cases} -\frac{q(x^{i+1})}{f_n(w)} & g_n(w) = 0 \\ \frac{f_n(w) + \sqrt{f_n(w)^2 + 2\frac{q(x^{i+1})}{r(x^{i+1})}g_n(w)^2}}{g_n(w)^2} r(x^{i+1}) & g_n(w) \neq 0 \end{cases}$$

(4.3)

As a consequence of the assumption that $\frac{\partial \pi}{\partial x}(x)\dot{x} < 0$, $f_n(w)$ and $g_n(w)$ cannot both be zero away from the origin if w is sufficiently close to $\frac{\partial \pi}{\partial x}(x^{i+1})$. Therefore, $z(w)$ is well defined. The difference between the true and computed first order partial derivatives of π can be rewritten as

$$\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \pi^{i+1}}{\partial x}(x^{i+1}) = \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{z\left(\frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\right)}{\|\frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\|} \frac{\partial \bar{\pi}}{\partial x}(x^{i+1})$$

and the main inequality (4.1) from the statement of Theorem 4.1 will follow from the fact that

$$\frac{z\left(\frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\right)}{\|\frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\|} = 1 + \mathcal{O}\left(\left\|\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \bar{\pi}}{\partial x}(x^{i+1})\right\|\right)$$

(4.4)

as long as the distance from x^{i+1} to the previous patch point is less than the maximum consecutive patch point distance.

To prove (4.4), we define the helper function

$$R(t) \equiv \frac{z\left(\frac{\partial \pi}{\partial x}(x^{i+1}) + t\left(\frac{\partial \bar{\pi}}{\partial x}(x^{i+1}) - \frac{\partial \pi}{\partial x}(x^{i+1})\right)\right)}{\left\|\frac{\partial \pi}{\partial x}(x^{i+1}) + t\left(\frac{\partial \bar{\pi}}{\partial x}(x^{i+1}) - \frac{\partial \pi}{\partial x}(x^{i+1})\right)\right\|}$$

(4.5)

so that we can appeal to the mean value theorem, and rewrite the left-hand side of (4.4) as $1 + \frac{dR}{dt}(\xi)$ for some $0 < \xi < 1$. If R is differentiable, then its derivative is bounded by

$$\left|\frac{dR}{dt}(t)\right| \leq$$

$$\left(\left\|\frac{\partial z}{\partial w}\left(\frac{\partial \pi}{\partial x}(x) + t\left(\frac{\partial \bar{\pi}}{\partial x}(x) - \frac{\partial \pi}{\partial x}(x)\right)\right)\right\| + \frac{z\left(\frac{\partial \pi}{\partial x}(x) + t\left(\frac{\partial \bar{\pi}}{\partial x}(x) - \frac{\partial \pi}{\partial x}(x)\right)\right)}{\left\|\frac{\partial \pi}{\partial x}(x) + t\left(\frac{\partial \bar{\pi}}{\partial x}(x) - \frac{\partial \pi}{\partial x}(x)\right)\right\|}\right)$$

$$\times \frac{\left\|\frac{\partial \bar{\pi}}{\partial x}(x) - \frac{\partial \pi}{\partial x}(x)\right\|}{\left\|\frac{\partial \pi}{\partial x}(x) + t\left(\frac{\partial \bar{\pi}}{\partial x}(x) - \frac{\partial \pi}{\partial x}(x)\right)\right\|}$$

(4.6)

13

Our assumption that the optimal cost is a strict Lyupanov function is the key fact that guarantees that there exists a maximum consecutive patch point distance such that $\frac{dR}{dt}(t)$ is well defined for $0 \le t \le 1$ and (4.4) holds. This is the main content of Lemmas 4.3, 4.4, 4.5, and 4.6.

The next lemma guarantees that the denominators in the bound on $\frac{dR}{dt}(t)$ (4.6) are bounded away from zero.

LEMMA 4.3. *Let $\bar{\pi}$ denote the Taylor polynomial centered at the previous patch point. If π has continuous first order partial derivatives on X_c and is a strict Lyapunov function on X_c, then there exists a maximum consecutive patch point distance h' and constant T', such that*

$$\frac{\left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\|}{\left\|\frac{\partial \pi}{\partial x}(x) + t\left(\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right)\right\|} \le T'\left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\|.$$

whenever $0 \le t \le 1$ and the distance from x to the previous patch point is less than h'.

Proof. We split the proof into two cases, depending on whether the previous patch point is the origin or not. In either case, we will use the fact that if $0 \le \|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\| < \|\frac{\partial \pi}{\partial x}(x)\|$, then

$$\frac{\left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\|}{\left\|\frac{\partial \pi}{\partial x}(x) + t\left(\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right)\right\|} \le \frac{\left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\|}{\|\frac{\partial \pi}{\partial x}(x)\| - \|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\|}$$

for all $0 \le t \le 1$.

If the previous patch point is the origin, then we need establish the existence of a deleted neighborhood of radius h' centered at the origin such that

$$\left|\left\|\frac{\partial \pi}{\partial x}(x)\right\| - \left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\|\right| > 0$$

for all x in the neighborhood. $\frac{\partial \pi}{\partial x}$ is nonzero away from the origin under the assumption that π is a strict Lyapunov function on X_c, so we need to establish that the difference between the gradients of π and $\bar{\pi}$ goes to zero faster than the gradient of π as x goes to zero. The assumption (§2.1, Assumption 2) that $Q \succ 0$ guarantees that the leading order linear term of $\frac{\partial \pi}{\partial x}(x)$ is nonzero for all nonzero x. Thus $\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\|/\|\frac{\partial \pi}{\partial x}(x)\| \to 0$ as $x \to 0$, so there is a radius $h' > 0$ such that

$$\left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\| < \left\|\frac{\partial \pi}{\partial x}(x)\right\| \text{ for all } x \text{ in } 0 < \|x\| \le h'.$$

Since $\frac{\partial \pi}{\partial x}$ and $\frac{\partial \bar{\pi}}{\partial x}(x)$ are continuous and $r \le \|x\| \le h'$ is compact, then for any r in $0 < r < h'$,

$$\sup_{r \le \|x\| \le h'}\left|\left\|\frac{\partial \pi}{\partial x}(x)\right\| - \left\|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\right\|\right|^{-1} < \infty.$$

Thus, the lemma holds if the previous patch point is the origin, and the distance from x^{i+1} to the origin is no greater than h'.

For this lemma to apply to the patchy method, the Al'brekht patch $\mathcal{A}$ must lie in the neighborhood of the origin of radius h', and the neighborhood of radius r must lie completely inside the Al'brekht patch.

We now consider the case where the previous patch point x^i is not the origin, so its distance to the origin is at least r. $\|\frac{\partial \pi}{\partial x}(x)\|$ is bounded away from zero on $\overline{X_c \setminus \mathcal{A}}$, and $\lim_{x \to x^i} \|\frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x)\| = 0$, so there exists an h' such that

$$\sup_{\substack{\overline{X_c \setminus \mathcal{A}} \\ \|x-x^i\| \leq h'}} \left| \left\| \frac{\partial \pi}{\partial x}(x) \right\| - \left\| \frac{\partial \pi}{\partial x}(x) - \frac{\partial \bar{\pi}}{\partial x}(x) \right\| \right|^{-1} < \infty$$

and we have established the validity of the lemma when the previous patch point is not the origin. $\square$

Our next task is to establish the existence of M_Δ so that $z(\frac{\partial \pi}{\partial x}(x) + \Delta)/\|\frac{\partial \pi}{\partial x}(x) + \Delta\|$ is uniformly bounded for all $\|\Delta\| \leq M_\Delta$ and all x in X_c. This is the content of lemma 4.4.

LEMMA 4.4.

$$\sup_{\substack{\overline{X_c \setminus \mathcal{A}} \\ \|\Delta\| \leq M_\Delta}} \frac{|z(\frac{\partial \pi}{\partial x}(x) + \Delta)|}{\|\frac{\partial \pi}{\partial x}(x) + \Delta\|} < \infty$$

for any nonzero M_Δ satisfying

$$M_\Delta < \inf_{\overline{X_c \setminus \mathcal{A}}} \left\| \frac{\partial \pi}{\partial x}(x) \right\|$$

$$M_\Delta < \frac{\inf_{\mathcal{S}} |\frac{\partial \pi}{\partial x}(x)f(x)|}{\sup_{\mathcal{S}} \|f(x)\|} \tag{4.7}$$

$$M_\Delta < \frac{\inf_{\mathcal{S}^C} |\frac{\partial \pi}{\partial x}(x)g(x)|}{\sup_{\mathcal{S}^C} \|g(x)\|}$$

where

$$\mathcal{S} \equiv \left\{ x \in \overline{X_c \setminus \mathcal{A}} \ \middle| \ \left| \frac{\partial \pi}{\partial x}(x)g(x) \right| \leq \sqrt{\frac{1}{2} \inf_{\overline{X_c \setminus \mathcal{A}}} r(x) \inf_{\overline{X_c \setminus \mathcal{A}}} \left| \frac{\partial \pi}{\partial x}(x)\dot{x} \right|} \right\}.$$

The assumption that π is a strict Lyupanov function and the definition of $\mathcal{S}$ implies that both $\sup_{\mathcal{S}} \|f(x)\|$ and $\sup_{\mathcal{S}^C} \|g(x)\|$ are strictly greater than zero, so M_Δ is well defined.

Proof. Let $w = \frac{\partial \pi}{\partial x}(x) + \Delta$, if $w \cdot g(x) \neq 0$, then by the formula for $z(w)$ in (4.3),

$$\frac{z(w)}{\|w\|} = \frac{w \cdot f(x) + \sqrt{(w \cdot f(x))^2 + 2\frac{q(x)}{r(x)}(w \cdot g(x))^2}}{(w \cdot g(x))^2} r(x) \tag{4.8}$$

Since $(\frac{\partial \pi}{\partial x}(x) + \Delta)g(x)$ appears in the denominator of (4.8), we split the proof into the two cases where x is in $\mathcal{S}$ or its complement.

If x is in $\mathcal{S}$, then by the definition of $\mathcal{S}$ and the assumption that π is a strict Lyupanov function

$$\frac{\partial \pi}{\partial x}(x)f(x) \leq -\frac{1}{2} \inf_{\overline{X_c \setminus \mathcal{A}}} \left| \frac{\partial \pi}{\partial x}(x)\dot{x} \right| < 0.$$

15

Therefore, if M_Δ satisfies the second condition in (4.7), then there is a strictly negative uniform upper bound on $(\frac{\partial \pi}{\partial x}(x) + \Delta) f(x)$ for all x in $\mathcal{S}$ and all $\|\Delta\| \leq M_\Delta$. It follows that for all x in $\mathcal{S}$, the square root term in (4.8) has the Taylor expansion with remainder

$$\sqrt{(w \cdot f(x))^2 + 2\frac{q(x)}{r(x)}(w \cdot g(x))^2} = -(w \cdot f(x))\left(1 + \frac{q(x)}{r(x)}\left(\frac{w \cdot g(x)}{w \cdot f(x)}\right)^2 \frac{1}{\sqrt{1 + \xi}}\right)$$

for some $\xi \geq 0$. So, for all x in $\mathcal{S}$ and all $\|\Delta\| \leq M_\Delta$

$$\frac{|z(\frac{\partial \pi}{\partial x}(x) + \Delta)|}{\|\frac{\partial \pi}{\partial x}(x) + \Delta)\|} = \frac{q(x)}{(\frac{\partial \pi}{\partial x}(x) + \Delta)f(x)} \frac{1}{\sqrt{1 + \xi}} \leq \frac{q(x)}{(\frac{\partial \pi}{\partial x}(x) + \Delta)f(x)} \tag{4.9}$$

which also holds in the case where $(\frac{\partial \pi}{\partial x}(x) + \Delta)g(x) = 0$, and consequently $z(\frac{\partial \pi}{\partial x}(x) + \Delta)$ solves a linear equation (2.6). The supremum of the right-hand side of (4.9) over $\mathcal{S}$ and all $\Delta \leq M_\Delta$ is finite. The first two conditions on M_Δ (4.7) guarantee that the infimum of $\|(\frac{\partial \pi}{\partial x}(x) + \Delta)f(x)\|$ is strictly greater than zero over $\mathcal{S}$ and all Δ bounded above by M_Δ in norm. The function q is continuous on the compact set $\overline{X_c \setminus \mathcal{A}} \supseteq \mathcal{S}$, therefore its supremum over $\mathcal{S}$ is bounded.

We now turn to the case where x is in $\mathcal{S}^C$. In this case, the only way $z(w)/\|w\|$ (4.8) can become unbounded is if $(\frac{\partial \pi}{\partial x}(x) + \Delta)g(x)$ can become arbitrarily small, but this cannot happen. By the definition of $\mathcal{S}$, the infimum of $|\frac{\partial \pi}{\partial x}(x)g(x)|$ over $\mathcal{S}^C$ is strictly positive, and the infimum of $|(\frac{\partial \pi}{\partial x}(x) + \Delta)g(x)|$ over $\mathcal{S}^C$ and all Δ bounded above in norm by M_Δ is also strictly positive.
□

Our final task in establishing Theorem 4.1 is to show that there exists an M_Δ such that the gradient of $z(w)$ exists and is uniformly bounded over $\frac{\partial \pi}{\partial x}(\overline{X_c \setminus \mathcal{A}}) + \Delta$ for $\|\Delta\| \leq M_\Delta$. This is the content of Lemmas 4.5 and 4.6 .

LEMMA 4.5. *Under the assumptions on the optimal control problem of §2.1, the function z that solves (2.6) is continuously differentiable in a nonempty neighborhood of every point in $\frac{\partial \pi}{\partial x}(\overline{X_c \setminus \mathcal{A}})$.*

Proof. By the implicit function theorem [8, ch. 9], a solution to $ax^2 + bx + c = 0$ is a continuously differentiable function of its coefficients in a neighborhood of a_0, b_0, and c_0 if x_0 solves the quadratic equation with coefficients a_0, b_0, and c_0, and $2a_0 x_0 + b_0 \neq 0$. To conclude that $z(w)$ is continuously differentiable in a neighborhood of each point in $\frac{\partial \pi}{\partial x}(\overline{X_c \setminus \mathcal{A}})$, we must verify that $-\frac{1}{r(x)}g_n(\frac{\partial \pi}{\partial x}(x))^2 z(\frac{\partial \pi}{\partial x}(x)) + f_n(\frac{\partial \pi}{\partial x}(x))$ is nonzero for every x in $\frac{\partial \pi}{\partial x}(\overline{X_c \setminus \mathcal{A}})$.

As a function of x, $\frac{\partial \pi}{\partial x}\dot{x}$ is continuous under the assumptions that f, g, r, and $\frac{\partial \pi}{\partial x}$ are continuous on $\overline{X_c \setminus \mathcal{A}}$. Furthermore, $\frac{\partial \pi}{\partial x}\dot{x}$ is nonzero on $\overline{X_c \setminus \mathcal{A}}$ by the assumption that π is a strict Lyapunov function, so there is a nonzero uniform lower bound on both $|\frac{\partial \pi}{\partial x}(x)\dot{x}|$ and $\|\frac{\partial \pi}{\partial x}\|$ on $\overline{X_c \setminus \mathcal{A}}$. It then follows from the identity

$$-\frac{1}{r(x)}g_n\left(\frac{\partial \pi}{\partial x}(x)\right)^2 z\left(\frac{\partial \pi}{\partial x}(x)\right) + f_n\left(\frac{\partial \pi}{\partial x}(x)\right) = \frac{\frac{\partial \pi}{\partial x}(x)\dot{x}}{\left\|\frac{\partial \pi}{\partial x}(x)\right\|},$$

that $-\frac{1}{r(x)}g_n(\frac{\partial \pi}{\partial x}(x))^2 z(\frac{\partial \pi}{\partial x}(x)) + f_n(\frac{\partial \pi}{\partial x}(x))$ is nonzero for every x in $\frac{\partial \pi}{\partial x}(\overline{X_c \setminus \mathcal{A}})$. □

We now verify that there exists an $M_\Delta > 0$ such that $\|\frac{\partial z}{\partial w}(\frac{\partial \pi}{\partial x}(\overline{X_c \setminus \mathcal{A}}) + \Delta)\|$ is

uniformly bounded for all $\|\Delta\| \leq M_\Delta$.

LEMMA 4.6. *If M_Δ satisfies the conditions of Lemma 4.4 in (4.7) in addition to*

$$M_\Delta < \frac{1}{2} \sup_{\overline{X_c \setminus \mathcal{A}}} \left(\left\| \frac{\partial \pi}{\partial x}(x) \right\| \left(\|f(x)\|^2 + 2\frac{q(x)}{r(x)}\|g(x)\|^2 \right) + \|f(x)\|^2 + \|g(x)\|^2 \right)$$

$$\times \inf_{\overline{X_c \setminus \mathcal{A}}} \left(\frac{\partial \pi}{\partial x}(x)\dot{x} \right)^2$$

then

$$\sup_{\substack{\overline{X_c \setminus \mathcal{A}} \\ \|\Delta\| \leq M_\Delta}} \left\| \frac{\partial z}{\partial w}\left(\frac{\partial \pi}{\partial x}(x) + \Delta \right) \right\| < \infty.$$

Proof. Let $w = \frac{\partial \pi}{\partial x}(x) + \Delta$. By Lemma 4.5, $\frac{\partial z}{\partial w}$ exists in a neighborhood of $\frac{\partial \pi}{\partial x}(x)$ for all x in $\overline{X_c \setminus \mathcal{A}}$, and an implicit formula for it is

$$\frac{\partial z}{\partial w}(w) = \frac{1}{f_n(w) - \frac{1}{r(x)}g_n(w)^2 z(w)}$$

$$\times \left[\left(\frac{z(w)}{\|w\|_2} \right)^2 g_n(w)\left(\|w\|_2 g(x) - g_n(w)w \right) \right.$$

$$\left. + \frac{z(w)}{\|w\|_2}\left(g_n(w)\frac{w}{\|w\|_2} - f(x) \right) \right] \quad (4.10)$$

Both f and g are uniformly bounded on $\overline{X_c \setminus \mathcal{A}}$. Since $\Delta \leq M_\Delta < \inf_{\overline{X_c \setminus \mathcal{A}}}\|\frac{\partial \pi}{\partial x}(x)\|$, w is nonzero, and the terms $f_n(w)$ and $g_n(w)$ are both uniformly bounded on $\overline{X_c \setminus \mathcal{A}}$ for all $\Delta \leq M_\Delta$ by the supremums of f and g on $\overline{X_c \setminus \mathcal{A}}$. By Lemma 4.4, the term $z(w)/\|w\|$ is bounded.

The only remaining term that can potentially become unbounded is $(f_n(w) - \frac{1}{r(x)}g_n(w)^2 z(w))^{-1}$. This quantity can be rewritten as

$$\left(f_n(w) - \frac{g_n(w)^2}{r(x)}z(w) \right)^2 = f_n(w)^2 + 2\frac{q(x)}{r(x)}g_n(w)^2 \quad (4.11)$$

$$= \frac{1}{\|\frac{\partial \pi}{\partial x}(x) + \Delta\|^2}\left(\left(\frac{\partial \pi}{\partial x}(x)\dot{x} \right)^2 + E(x, \Delta) \right) \quad (4.12)$$

Equation (4.11) follows from the fact that both sides of the equation are formulas for the discriminant of the quadratic equation (2.6). If we solve the HJB equation (1.3a) for $q(x)$ in terms of $\frac{\partial \pi}{\partial x}(x)$ and the other problem data, and then substitute this formula into (4.11), then we arrive at (4.12). We defined M_Δ in the lemma statement so that $E(x, \Delta) = \mathcal{O}(\|\Delta\|)$ and the infimum of (4.12) is strictly greater than zero over $\overline{X_c \setminus \mathcal{A}}$ and all $\|\Delta\| \leq M_\Delta$. □

5. ϕ_1 satisfies the Lipschitz condition. THEOREM 5.1. *Under the assumptions on the optimal control problem of §2.1, there exists a maximum step size $h_L > 0$, maximum multiplier $M_{max} > 0$, and Lipschitz constant L_1 such that for all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a distance of h_L of each other, and for all $M \leq M_{max}$,*

$$\|\phi_1(C^i) - \phi_1(\hat{C}^i)\| \leq L_1 M \|x^{i+1} - x^i\|^{d+1}$$

17

whenever

$$\|C^i_j - \hat{C}^i_j\| \leq M\|x^{i+1} - x^i\|^{d+2-j} \ for \ 0 \leq j \leq d+1$$

where C^i is the vector of exact partial derivatives up to order $d+1$ of the optimal cost at x^i. The theorem holds for any M_{max}, h_L pair that satisfy maximum consecutive patch point distance multiplier condition of Definition 5.2.

DEFINITION 5.2. *The maximum consecutive patch point distance h_L and the maximum multiplier M_{max} are said to satisfy the maximum consecutive patch point distance multiplier condition if*

$$h_L > 0 \qquad M_{max} > 0 \qquad h_L \leq h_T \qquad (\sqrt{n}L_{inh}M_{max} + T)h_L^{d+1} \leq M_\Delta$$

where L_{inh} is defined in Lemma 3.4, h_T and T are the maximum consecutive step size and local truncation constant from Corollary 4.2 and M_Δ satisfies the inequalities of Theorem 4.1.

Proof. [Proof of Theorem 5.1] If $w = \frac{\partial P}{\partial x}(C, x^{i+1} - x^i)$ is nonzero and z is defined at w, then ϕ_1 can be expressed as

$$\phi_1(C) = z\Big(\frac{\partial P}{\partial x}(C, x^{i+1} - x^i)\Big)\frac{\frac{\partial P}{\partial x}(C, x^{i+1} - x^i)}{\|\frac{\partial P}{\partial x}(C, x^{i+1} - x^i)\|}$$

where P is the polynomial whose coefficients are held in C, and is defined in Lemma 3.4. We will show that there exists L' such that, if the conditions of the theorem are met, then

$$\left\|\frac{\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)}{\|\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)\|} - \frac{\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)}{\|\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)\|}\right\| \leq L'_1 M\|x^{i+1} - x^i\|^{d+1} \qquad (5.1)$$

and

$$\left|z\Big(\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)\Big) - z\Big(\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)\Big)\right| \leq L'_1 M\|x^{i+1} - x^i\|^{d+1} \qquad (5.2)$$

which implies the conclusion of the theorem.

We start with the inequality in (5.1). P and all its partial derivatives are linear in its first argument. If

$$\left\|\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) - \frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)\right\| < \left\|\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)\right\|$$

then

$$\left\|\frac{\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)}{\|\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)\|} - \frac{\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)}{\|\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)\|}\right\|$$

$$\leq 2\frac{\left\|\frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i)\right\|}{\|\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)\| - \|\frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i)\|}$$

$$\leq 2\frac{\sqrt{n}L_{\mathrm{inh}}M\|x^{i+1} - x^i\|^{d+1}}{\|\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)\| - \|\frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i)\|}$$

where the final inequality follows from Lemma 3.4. The inequality in (5.1) will follow after we establish that for some ϵ,

$$0 < \epsilon \leq \left| \left\| \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\| - \left\| \frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i) \right\| \right| \tag{5.3}$$

(5.3) also guarantees that $\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)$ is nonzero. C^i holds the exact partial derivatives of the optimal cost at x^i, so $P(C^i, x - x^i) = \bar{\pi}(x)$ where $\bar{\pi}$ is the Taylor polynomial centered at x^i. By Corollary 4.2,

$$\left| \left\| \frac{\partial \pi}{\partial x}(x^{i+1}) \right\| - \left\| \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\| \right| \leq \left\| \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\|$$

$$\leq T \| x^{i+1} - x^i \|^{d+1}.$$

By Lemma 3.4

$$\left\| \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) - \frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i) \right\| \leq \sqrt{n} L_{\text{inh}} M_{\max} \| x^{i+1} - x^i \|$$

Inequality 5.3 follows from the final strict inequality in

$$\left| \left\| \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\| - \left\| \frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i) \right\| \right|$$

$$= \left| \left\| \frac{\partial \pi}{\partial x}(x^{i+1}) \right\| + \left\| \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\| \right.$$

$$\left. - \left\| \frac{\partial \pi}{\partial x}(x^{i+1}) \right\| - \left\| \frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i) \right\| \right|$$

$$\geq \inf_{X_c \setminus \mathcal{A}} \left\| \frac{\partial \pi}{\partial x}(x) \right\| - (\sqrt{n} L_{\text{inh}} M_{\max} + T) h_L^{d+1}$$

$$> 0$$

The maximum patch point multiplier condition (Definition 5.2) was defined so the last inequality is strict.

We now prove that the inequality (5.2) holds by applying the mean value theorem to $z(w)$. If

$$\left\| \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) - \frac{\partial \pi}{\partial x}(x^{i+1}) \right\| \leq M_\Delta \tag{5.4}$$

and

$$\left\| \frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i) - \frac{\partial \pi}{\partial x}(x^{i+1}) \right\| \leq M_\Delta \tag{5.5}$$

then for all τ in $0 \leq \tau \leq 1$,

$$\left\| (1 - \tau)\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) + \tau \frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i) - \frac{\partial \pi}{\partial x}(x^{i+1}) \right\| \leq M_\Delta$$

19

meaning that every point on the line segment connecting $\frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i)$ and $\frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i)$ is at most a distance of M_Δ from $\frac{\partial \pi}{\partial x}(x^{i+1})$. Therefore, z is well defined at the endpoints, and by Lemma 4.6 there is a uniform bound on the magnitude of $\frac{\partial z}{\partial w}$ over all such lines. Inequality 5.2 will follow once we show that the conditions of the theorem guarantee that inequalities 5.4 and 5.5 hold. By assumption, $h_L \leq h_T$, so by Corollary 4.2 and Definition 5.2

$$\left\| \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\| \leq T h_L^{d+1} \leq M_\Delta$$

and

$$\left\| \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial P}{\partial x}(\hat{C}^i, x^{i+1} - x^i) \right\|$$
$$\leq \left\| \frac{\partial P}{\partial x}(C^i - \hat{C}^i, x^{i+1} - x^i) \right\| + \left\| \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial P}{\partial x}(C^i, x^{i+1} - x^i) \right\|$$
$$\leq (\sqrt{n} L_{\text{inh}} M_{\max} + T) h_L^{d+1}$$
$$\leq M_\Delta.$$

□

6. Local truncation error for second and higher order partial derivatives . THEOREM 6.1. *Let C^i and C^{i+1} denote vectors of exact partial derivatives of the optimal cost at the consecutive patch points x^i and x^{i+1} in the computational domain X_c. If the optimal control problem satisfies the assumptions of §2.1 , then there exists local truncation constant $T_2 < \infty$ and maximum consecutive patch point distance $h > 0$ such that for all $2 \leq j \leq d+1$,*

$$\|C_j^{i+1} - \phi_j(C^i)\| \leq T \|x^{i+1} - x^i\|_2^{d+2-j} \text{ whenever } \|x^{i+1} - x^i\| \leq h$$

We delay the proof of Theorem 6.1 until after we have developed an overview of its main ideas.

Let $\dot{x}^{i+1}$ denote the computed optimal direction at the current patch point $x = x^{i+1}$. There are two changes of variables associated with the computed and exact solutions at the current patch point. They are

$$x = x^{i+1} + \hat{V}\hat{\xi} \qquad \hat{V}^1 = \frac{\dot{x}^{i+1}}{\|\dot{x}^{i+1}\|} \qquad x = x^{i+1} + V\xi \qquad V^1 = \frac{\dot{x}}{\|\dot{x}\|}. \qquad (6.1)$$

We denote the computed and exact optimal cost and control under the changes of variables as

$$\hat{\pi}^{i+1}(\hat{\xi}) \equiv \pi^{i+1}(x^{i+1} + \hat{V}\hat{\xi}) \qquad\qquad \tilde{\pi}(\xi) \equiv \pi(x^{i+1} + V\xi)$$
$$\hat{\kappa}^{i+1}(\hat{\xi}) \equiv \kappa^{i+1}(x^{i+1} + \hat{V}\hat{\xi}) \qquad\qquad \tilde{\kappa}(\xi) \equiv \kappa(x^{i+1} + V\xi). \qquad (6.2)$$

We defined the *Kronecker derivative* notation $\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}$ in §2.6 and the appendix, and it is merely a bookkeeping mechanism for mixed partial derivatives. The patchy algorithm computes the kth order partial derivatives with respect to the change of variables of the optimal cost, then recovers the partial derivatives with respect to the original state space variables by the formula

$$\underbrace{\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}}_{k \text{ times}} [\pi^{i+1}(x)]_{x=x^{i+1}} = \underbrace{\frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}}}_{k \text{ times}} [\hat{\pi}^{i+1}(\hat{\xi})]_{\xi=0} \underbrace{\hat{V}^T \otimes \cdots \otimes \hat{V}^T}_{k \text{ times}}. \qquad (6.3)$$

The exact k^{th} order partial derivatives of the optimal cost with respect to the original state space variables are recovered by

$$\underbrace{\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}}_{k \text{ times}}[\pi(x)]_{x=x^{i+1}} = \underbrace{\frac{\partial}{\partial \xi} \otimes \cdots \otimes \frac{\partial}{\partial \xi}}_{k \text{ times}}[\tilde{\pi}(\xi)]_{\xi=0} V^T \underbrace{\otimes \cdots \otimes V^T}_{k \text{ times}}. \qquad (6.4)$$

There are two main ideas behind the proof of Theorem 6.1. The left hand side of (6.4) is the same for any choice of orthogonal matrix V in the change of variables $x = x^{i+1} + V\xi$. In the case of the computed partial derivatives of the optimal cost, the entries in $\frac{\partial}{\partial\hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\hat{\xi}}$ appearing on the right hand side of (6.3) are a mixture of characteristic and non characteristic partial derivatives. Despite this, the computed partial derivatives with respect to the original state space variables of the optimal cost are invariant under any choice of orthonormal vectors $\{\hat{V}^j\}_{j=1}^n$ in the change of variables $x = x^{i+1} + \hat{V}\hat{\xi}$ as long as $\hat{V}^1 = \dot{x}^{i+1}/\|\dot{x}^{i+1}\|$. This is the first main idea and is the content of Lemma 6.2, Corollary 6.3, and Lemma 6.4.

The second idea is that there exist two convenient changes of variables that simplify our task of establishing the desired bound on the difference between the partial derivatives with respect to the original state space variables of the exact and computed optimal cost by bounding the difference between the partial derivatives under the convenient changes of variables. The advantage of this approach is that we may treat the case of bounding the difference between the computed and exact characteristic partial derivatives independently of the non characteristic partial derivatives.

Once we have established these two facts, we will prove the theorem for the second order partial derivatives of the optimal cost and proceed by induction.

The relevant interpretation of Lemma 6.2 is that, taken as a group, the characteristic partial derivatives that the patchy algorithm computes at the current patch point can be expressed as an orthogonal transformation of a vector that depends only on partial derivatives of the problem data at the current patch point, and on partial derivatives of the optimal cost that are of strictly lower order than the characteristic partial derivatives being computed. Secondly, all these partial derivatives are with respect to the original state space variables.

LEMMA 6.2. *Suppose V is an orthogonal matrix whose first column is the exact normalized optimal direction, $V^1 = \dot{x}/\|\dot{x}\|$ at $x = x^{i+1}$. Let $\tilde{\pi}(\xi)$ denote the optimal cost under the change of variables $x = x^{i+1} + V\xi$. Then, at $x = x^{i+1}$, there exist row vectors $w^p \in \mathbb{R}^{n^{p-1}\times 1}$, independent of $V^2,\ldots,V^n$, such that the exact characteristic partial derivatives of the optimal cost satisfy*

$$\frac{\partial}{\partial\xi}\left[\frac{\partial\tilde{\pi}}{\partial\xi_1}(\xi)\right]_{\xi=0} = \left(w^2(x^{i+1})\right)^T V$$

$$\frac{\partial}{\partial\xi} \otimes \frac{\partial}{\partial\xi}\left[\frac{\partial\tilde{\pi}}{\partial\xi_1}(\xi)\right]_{\xi=0} = \left(w^3(x^{i+1})\right)^T V \otimes V$$

$$\vdots \qquad\qquad (6.5)$$

$$\frac{\partial}{\partial\xi} \otimes \cdots \otimes \frac{\partial}{\partial\xi}\left[\frac{\partial\tilde{\pi}}{\partial\xi_1}(\xi)\right]_{\xi=0} = \left(w^p(x^{i+1})\right)^T V \otimes \cdots \otimes V$$

where $w^k(x^{i+1})$ depends only on partial derivatives of the problem data and partial derivatives with respect to the original state space variables of the optimal cost of order $k-1$ and lower.

An analogous statement is true for the computed characteristic partial derivatives of the optimal cost.

Proof. To reduce the number of terms in some of the ensuing formulas, let $\tilde{F}$ and $\tilde{\ell}$ denote the dynamics and lagrangian as functions of the new coordinates.

$$\tilde{F}(\xi) \equiv \tilde{f}(\xi) + \tilde{g}(\xi)\tilde{\kappa}(\xi) \qquad\qquad \tilde{\ell}(\xi) \equiv \tilde{q}(\xi) + \frac{1}{2}\tilde{r}(\xi)\tilde{\kappa}(\xi)^2$$

where the tilde functions appearing on the right hand sides are defined by the change of variables $x = x^{i+1} + V\xi$.

We will prove the lemma in the representative case of the third order characteristic partial derivatives of the optimal cost. It follows from the HJB equations (2.8a) that a formula for the exact third order characteristic partial derivatives of the optimal cost is

$$\frac{\partial^3 \tilde{\pi}}{\partial \xi_{j_1} \partial \xi_{j_2} \partial \xi_1}(0) = -\frac{1}{\|\dot{x}\|}\left[\frac{\partial^2 \tilde{\pi}}{\partial \xi_{j_1} \partial \xi}(0)\frac{\partial \tilde{F}}{\partial \xi_{j_2}}(0) + \frac{\partial^2 \tilde{\pi}}{\partial \xi_{j_2} \partial \xi}(0)\frac{\partial \tilde{F}}{\partial \xi_{j_1}}(0) \right.$$
$$\left. + \frac{\partial \tilde{\pi}}{\partial \xi}(0)\frac{\partial^2 \tilde{F}}{\partial \xi_{j_1} \partial \xi_{j_2}}(0) + \frac{\partial^2 \tilde{\ell}}{\partial \xi_{j_1} \partial \xi_{j_2}}(0) \right]. \quad (6.6)$$

The right hand side of the formula (6.6) contains three distinctive types of terms characterized by the order of the partial derivatives of $\tilde{\pi}$, $\tilde{F}$, and $\tilde{\ell}$. The first type is the product of a second order partial derivative of $\tilde{\pi}$ and a first order partial derivative of $\tilde{F}$, the second type is the product of a first order partial of $\tilde{\pi}$ and a second order partial of $\tilde{F}$, and the final type is a second order partial of ℓ.

We will show that each type of term can be written as the product of a row vector that depends only on the partial derivatives of both $\pi(x)$ and $F(x)$ at $x = x^{i+1}$ and the column vector $V^{j_1} \otimes V^{j_2}$. We tackle the first type of term first. A formula for $\frac{\partial^2 \tilde{\pi}}{\partial \xi_{j_1} \partial \xi}(0)\frac{\partial \tilde{F}}{\partial \xi_{j_2}}(0)$ is

$$\frac{\partial^2 \tilde{\pi}}{\partial \xi_{j_1} \partial \xi}(0)\frac{\partial \tilde{F}}{\partial \xi_{j_2}}(0) = \left[\sum_{\sigma=1}^{n}\left(\frac{\partial}{\partial x}\left[\frac{\partial \pi}{\partial x_\sigma}(x) \right]_{x=x^{i+1}} \right) \otimes \left(\frac{\partial}{\partial x}\left[F_\sigma(x) \right]_{x=x^{i+1}} \right) \right]V^{j_1} \otimes V^{j_2}.$$
$$(6.7)$$

The quantity on the right hand side of (6.7) depends only on second order and lower partial derivatives of $\pi(x)$ and $F(x)$ evaluated at x^{i+1}. To derive a formula for $\frac{\partial^2 \tilde{\pi}}{\partial \xi_{j_2} \partial \xi}(0)\frac{\partial \tilde{F}}{\partial \xi_{j_1}}(0)$ with the desired factor $V^{j_1} \otimes V^{j_2}$, we use the fact that for every permutation (i_1, i_2) of (j_1, j_2), there exists a permutation matrix P, that depends only on the permutation and not the choice of basis vectors, such that $V^{i_1} \otimes V^{i_2} = P(V^{j_1} \otimes V^{j_2})$. Therefore, a relevant formula is

$$\frac{\partial^2 \tilde{\pi}}{\partial \xi_{j_2} \partial \xi}(0)\frac{\partial \tilde{F}}{\partial \xi_{j_1}}(0) =$$
$$\left[\sum_{\sigma=1}^{n}\left(\frac{\partial}{\partial x}\left[\frac{\partial \pi}{\partial x_\sigma}(x) \right]_{x=x^{i+1}} \right) \otimes \left(\frac{\partial}{\partial x}\left[F_\sigma(x) \right]_{x=x^{i+1}} \right)P^T \right]V^{j_1} \otimes V^{j_2}$$

Although tedious, we can derive formulas for the other terms in (6.6) that are the product of a row vector that depends only on the second order and lower partial

derivatives of the optimal cost and problem data with respect to the original state space variables, and the column vector V^{j_1} and V^{j_2}. It follows that there exists $w^3(x^{i+1}) \in \mathbb{R}^{n^2 \times 1}$ such that

$$\frac{\partial \tilde{\pi}^3}{\partial \xi_{j_1} \partial \xi_{j_2} \partial \xi_1}(0) = \left(w^3(x^{i+1})\right)^T V^{j_1} \otimes V^{j_2}$$

so the formula for $\frac{\partial}{\partial \xi} \otimes \frac{\partial}{\partial \xi} \left[\frac{\partial \tilde{\pi}}{\partial \xi_1}(\xi)\right]_{\xi=0}$ in the theorem statement follows.

The proof for the computed characteristic partial derivatives is analogous. □

The relevant interpretation of Corollary 6.3 is that, under two changes of variables that share the normalized computed optimal direction at the current patch point as their first basis vector, the two sets of computed characteristic partial derivatives of the optimal cost at the current patch point are related to each other by an orthogonal transformation.

COROLLARY 6.3. *If $\tilde{V}$ and $\hat{V}$ are are orthogonal matrices whose first columns are the normalized computed optimal direction at $x = x^{i+1}$, that is $\tilde{V}^1 = \hat{V}^1 = \dot{x}^{i+1}/\|\dot{x}^{i+1}\|$, then the approximate second and higher order approximate characteristic partial derivatives of the optimal cost computed under the changes of variables*

$$x = x^{i+1} + \tilde{V}\tilde{\xi} \qquad\qquad x = x^{i+1} + \hat{V}\hat{\xi}$$

are related by the orthogonal transformation

$$\frac{\partial}{\partial \tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \tilde{\xi}} \left[\frac{\partial \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_1}(\tilde{\xi})\right]_{\tilde{\xi}=0} = \frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}} \left[\frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}_1}(\hat{\xi})\right]_{\hat{\xi}=0} \hat{V}^T \tilde{V} \otimes \cdots \otimes \hat{V}^T \tilde{V}.$$

Proof. The proof for the case of the approximate third order partials of the optimal cost is representative. By Lemma 6.2, there exists a single vector w^{i+1} that depends only on the previously computed lower order partial derivatives of $\pi^{i+1}(x)$ at $x = x^{i+1}$ such that the characteristic partials under both coordinate transformations are given by

$$\frac{\partial}{\partial \tilde{\xi}} \otimes \frac{\partial}{\partial \tilde{\xi}} \left[\frac{\partial \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_1}(\tilde{\xi})\right]_{\tilde{\xi}=0} = \left(w^{i+1}\right)^T \tilde{V} \otimes \tilde{V}$$

and

$$\frac{\partial}{\partial \hat{\xi}} \otimes \frac{\partial}{\partial \hat{\xi}} \left[\frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}_1}(\hat{\xi})\right]_{\hat{\xi}=0} = \left(w^{i+1}\right)^T \hat{V} \otimes \hat{V}.$$

The conclusion of the corollary immediately follows from the identity

$$\left(w^{i+1}\right)^T \left(\tilde{V} \otimes \cdots \otimes \tilde{V}\right) = \left(w^{i+1}\right)^T \left(\hat{V} \otimes \cdots \otimes \hat{V}\right)\left(\hat{V}^T \tilde{V} \otimes \cdots \otimes \hat{V}^T \tilde{V}\right).$$

□

The content of Lemma 6.4 is that, under two orthogonal changes of variables that share the normalized computed optimal direction at the current patch point as their first basis vector, the two sets of computed characteristic and non characteristic partial derivatives are related by an orthogonal transformation. Thus, it follows immediately from (6.3) that the computed partial derivatives of the optimal cost with respect to the original state space variables at the current patch point are the same under two

different orthogonal changes of variables as long as the changes of variables share the normalized computed optimal direction as their first basis vector.

LEMMA 6.4. *If $\tilde{V}$ and $\hat{V}$ are orthogonal matrices whose first columns are the normalized optimal direction at $x = x^{i+1}$, that is $\tilde{V}^1 = \hat{V}^1 = \dot{x}^{i+1}/\|\dot{x}^{i+1}\|$, then the approximate second and higher order approximate partial derivatives of the optimal cost computed under the changes of variables*

$$x = x^{i+1} + \tilde{V}\tilde{\xi} \qquad\qquad x = x^{i+1} + \hat{V}\hat{\xi}$$

are related to each other by

$$\frac{\partial}{\partial\tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\tilde{\xi}}\big[\tilde{\pi}^{i+1}(\tilde{\xi})\big]_{\tilde{\xi}=0}\big(\tilde{V}^T \otimes \cdots \otimes \tilde{V}^T\big) =$$
$$\frac{\partial}{\partial\hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\hat{\xi}}\big[\hat{\pi}^{i+1}(\hat{\xi})\big]_{\hat{\xi}=0}\big(\hat{V}^T \otimes \cdots \otimes \hat{V}^T\big).$$

Proof. *The conclusion of the theorem will follow once we establish that the quantity*

$$\frac{\partial}{\partial\tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\tilde{\xi}}\big[\tilde{\pi}^{i+1}(\tilde{\xi})\big]_{\tilde{\xi}=0}$$
$$- \frac{\partial}{\partial\hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\hat{\xi}}\big[\hat{\pi}^{i+1}(\hat{\xi})\big]_{\hat{\xi}=0}\big(\hat{V}^T\tilde{V} \otimes \cdots \otimes \hat{V}^T\tilde{V}\big) \quad (6.8)$$

is zero. We will prove the lemma for the third order partial derivatives of the computed optimal cost, which is a representative case.

It follows immediately from the definition of the $\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}$ notation that each entry in the term appearing on the left of (6.8) can be extracted by multiplying by one of the standard basis vectors. Any standard basis vector in $\mathbb{R}^{n^3}$ can be expressed as a Kronecker product of three standard basis vectors in $\mathbb{R}^n$, so

$$\frac{\partial^3\tilde{\pi}^{i+1}}{\partial\tilde{\xi}_{j_1}\partial\tilde{\xi}_{j_2}\partial\tilde{\xi}_{j_3}}(0) = \frac{\partial}{\partial\tilde{\xi}} \otimes \frac{\partial}{\partial\tilde{\xi}} \otimes \frac{\partial}{\partial\tilde{\xi}}\big[\tilde{\pi}^{i+1}(\tilde{\xi})\big]_{\tilde{\xi}=0} e_{j_1} \otimes e_{j_2} \otimes e_{j_3},$$

where e_j denotes the jth column of the identity matrix. We will establish that the difference in (6.8) is zero by multiplying it by $e_{j_1} \otimes e_{j_2} \otimes e_{j_3}$ and showing that

$$\frac{\partial^3\tilde{\pi}^{i+1}}{\partial\tilde{\xi}_{j_1}\partial\tilde{\xi}_{j_2}\partial\tilde{\xi}_{j_3}}(0) - \frac{\partial}{\partial\hat{\xi}} \otimes \frac{\partial}{\partial\hat{\xi}} \otimes \frac{\partial}{\partial\hat{\xi}}\big[\hat{\pi}^{i+1}(\hat{\xi})\big]_{\hat{\xi}=0}\big(\hat{V}^T\tilde{V}^{j_1} \otimes \hat{V}^T\tilde{V}^{j_2} \otimes \hat{V}^T\tilde{V}^{j_3}\big)$$

is zero for any $1 \le j_1, j_2, j_3 \le n$, which is equivalent to showing that

$$\frac{\partial^3\tilde{\pi}^{i+1}}{\partial\tilde{\xi}_{j_1}\partial\tilde{\xi}_{j_2}\partial\tilde{\xi}_{j_3}}(0)$$
$$- \sum_{\sigma_1=1}^{n}\sum_{\sigma_2=1}^{n}\sum_{\sigma_3=1}^{n}\Big(\frac{\partial^3\hat{\pi}^{i+1}}{\partial\hat{\xi}_{\sigma_{j_1}}\partial\hat{\xi}_{\sigma_{j_2}}\partial\hat{\xi}_{\sigma_{j_3}}}(0)\big(\hat{V}^{\sigma_1}\cdot\tilde{V}^{j_1}\big)\big(\hat{V}^{\sigma_2}\cdot\tilde{V}^{j_2}\big)\big(\hat{V}^{\sigma_3}\cdot\tilde{V}^{j_3}\big)\Big) \quad (6.9)$$

is zero.

The partial derivatives of $\tilde{\pi}^{i+1}(\tilde{\xi})$ and $\hat{\pi}^{i+1}(\hat{\xi})$ at $\tilde{\xi} = \hat{\xi} = 0$ are computed by two different methods depending on whether they are characteristic or non characteristic

24

partial derivatives, so the proof that the quantity in (6.9) is zero naturally breaks up into two cases. In the characteristic case, at least one of the coordinate indices j_1, j_2, or j_3 is one, and we may assume that $j_3 = 1$ without a meaningful loss of generality. The columns of $\tilde{V}$ and $\hat{V}$ are both orthonormal sets, and $\tilde{V}^1 = \hat{V}^1$ by assumption, so $\hat{V}^{\sigma_3} \cdot \tilde{V}^1 = \delta_{1,\sigma_3}$, where δ_{1,σ_3} is the Kronecker delta. Therefore for any j_1, j_2 in $1 \leq j_1, j_2 \leq n$, the difference in (6.9) reduces to a quantity that depends only on characteristic partial derivatives of the optimal cost

$$\frac{\partial^3 \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_{j_1} \partial \tilde{\xi}_{j_2} \partial \tilde{\xi}_1}(0) - \sum_{\sigma_1=1}^{n} \sum_{\sigma_2=1}^{n} \left(\frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{\sigma_1} \partial \hat{\xi}_{\sigma_2} \partial \hat{\xi}_1}(0) \left(\hat{V}^{\sigma_1} \cdot \tilde{V}^{j_1}\right)\left(\hat{V}^{\sigma_2} \cdot \tilde{V}^{j_2}\right) \right)$$

$$= \frac{\partial^3 \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_{j_1} \partial \tilde{\xi}_{j_2} \partial \tilde{\xi}_1}(0) - \frac{\partial}{\partial \hat{\xi}} \otimes \frac{\partial}{\partial \hat{\xi}} \left[\frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}_1}(\hat{\xi})\right]_{\hat{\xi}=0} \left(\hat{V}^T \tilde{V}^{j_1} \otimes \hat{V}^T \tilde{V}^{j_2}\right) \qquad (6.10)$$

$$= 0.$$

The first equality follows from the definition of the $\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}$, and the second equality follows from Corollary 6.3. Therefore the quantity in (6.9) is zero whenever $\frac{\partial^3 \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_{j_1} \partial \tilde{\xi}_{j_2} \partial \tilde{\xi}_{j_3}}(0)$ is a characteristic partial derivative.

   We now show that the difference in (6.9) is zero in the non characteristic partial derivative case, that is, when none of the coordinate indices j_1, j_2, or j_3 are one. By a standard argument

$$\frac{\partial^3 \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_{j_1} \partial \tilde{\xi}_{j_2} \partial \tilde{\xi}_{j_3}}(0) =$$

$$\left[\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \left[\pi^i(x)\right]_{x=x^{i+1}} \left(\hat{V} \otimes \hat{V} \otimes \hat{V}\right)\right] \left[\hat{V}^T \tilde{V}^{j_1} \otimes \hat{V}^T \tilde{V}^{j_2} \otimes \hat{V}^T \tilde{V}^{j_3}\right]$$

The right hand side of the previous equation is the product of a row and column vector. Since none of j_1, j_2, or j_3 is one, the previous equation can be rewritten as

$$\frac{\partial^3}{\partial \tilde{\xi}_{j_1} \partial \tilde{\xi}_{j_2} \partial \tilde{\xi}_{j_3}} \left[\tilde{\pi}^{i+1}(\tilde{\xi})\right]_{\tilde{\xi}=0} =$$

$$\sum_{\sigma_1=2}^{n} \sum_{\sigma_2=2}^{n} \sum_{\sigma_3=2}^{n} \left(\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \left[\pi^i(x)\right]_{x=x^{i+1}} \left(\hat{V}^{\sigma_1} \otimes \hat{V}^{\sigma_2} \otimes \hat{V}^{\sigma_3}\right) \right)$$

$$\times \left(\left(\hat{V}^{\sigma_1} \cdot \tilde{V}^{j_1}\right)\left(\hat{V}^{\sigma_2} \cdot \tilde{V}^{j_2}\right)\left(\hat{V}^{\sigma_3} \cdot \tilde{V}^{j_3}\right) \right) \quad (6.11)$$

where the summations all start at two because $\hat{V}^1 \cdot \tilde{V}^{j_1} = \hat{V}^1 \cdot \tilde{V}^{j_2} = \hat{V}^1 \cdot \tilde{V}^{j_3} = 0$. Thus, the sum on the right hand side of (6.11) depends only on non characteristic partial derivatives of $\hat{\pi}^{i+1}(\hat{\xi})$ at $\hat{\xi} = 0$. Therefore when $j_1, j_2, j_3 \neq 1$, $\frac{\partial^3 \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_j \partial \tilde{\xi}_k \partial \tilde{\xi}_\ell}(0)$ can be expressed as a linear transformation of $\frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}} [\hat{\pi}^{i+1}(\hat{\xi})]_{\hat{\xi}=0}$, whose entries are composed of both characteristic and non characteristic partial derivatives. The conclusion of the theorem follows from the fact that (6.10) and (6.11) cover all of (6.9). $\square$

   We now turn to the task of constructing two convenient changes of variables, meaning we can pick V and $\hat{V}$ in (6.1) such that V and $\hat{V}^T$ are almost inverses of

each other. This is the content of Lemma 6.5, which in turn relies on Lemma 6.6, and Corollaries 6.7 and 6.8.

LEMMA 6.5. *Let $\dot{x}^{i+1}$ denote the optimal direction at $x = x^{i+1}$ computed by the patchy algorithm from the exact partial derivatives of the optimal cost at the previous patch point. If the optimal cost $\pi(x)$ is a strict Lyapunov function and the distance between the current and previous patch point is less than the maximum consecutive patch point distance $h > 0$, then there exists a constant $B < \infty$ and orthogonal matrices V and $\hat{V}$ such that*

$$V^1 = \frac{\dot{x}}{\|\dot{x}\|} \qquad\qquad \hat{V}^1 = \frac{\dot{x}^{i+1}}{\|\dot{x}^{i+1}\|}$$

and

$$\|\hat{V}^T V - I\| \leq B \|x^{i+1} - x^i\|^{d+1}$$

where the degree of the approximate optimal cost polynomial π^{i+1} is $d + 1$.

Proof. By Theorem 4.1, there is a constant $T < \infty$ and a maximum consecutive patch point distance $h > 0$ such that

$$\left\| \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \pi^{i+1}}{\partial x}(x^{i+1}) \right\| \leq T \|x^{i+1} - x^i\|^{d+1}.$$

Therefore, there exists an $\epsilon > 0$ such that, for any x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a radius of h of the previous patch point

$$\|\dot{x} - \dot{x}^{i+1}\| = \mathcal{O}\big(\|x^{i+1} - x^i\|^{d+1}\big) \qquad and \qquad 0 < \epsilon \leq \|\dot{x}\| - \|\dot{x} - \dot{x}^{i+1}\|$$

The result of the lemma follows from Lemma 6.6 where V and $\hat{V}$ are constructed, and Corollaries 6.7 and 6.8 where the bound on $\|\hat{V}^T V - I\|$ in the lemma statement is established. □

We construct orthogonal matrices V and $\hat{V}$ associated with the changes of variables for the exact and computed optimal cost such that $\hat{V}^T$ and V are almost inverses of each other. This is the content of Lemma 6.6, which is mostly a restatement of [10, pg. 73] in terms relevant to our problem.

LEMMA 6.6. *Suppose v and $\hat{v}$ are nonzero vectors in $\mathbb{R}^n$ such that $v \cdot \hat{v} > 0$, then there exist orthonormal bases $\{V^1, \ldots, V^n\}$ and $\{\hat{V}^1, \ldots, \hat{V}^n\}$ with the properties*

1. *$V^1 = v/\|v\|$ and $\hat{V}^1 = \hat{v}/\|\hat{v}\|$*
2. *$\sum_{j=2}^{n} \big(V^1 \cdot \hat{V}^j\big)^2 = \sum_{j=2}^{n} \big(\hat{V}^1 \cdot V^j\big)^2 = 1 - \big(V^1 \cdot \hat{V}^1\big)^2$*
3. *The sets $\{V^2, \ldots, V^n\}$ and $\{\hat{V}^2, \ldots, \hat{V}^n\}$ satisfy the orthogonality condition $\big(V^j \cdot \hat{V}^k\big) = 0$ whenever $j \neq k$ and $V^1 \cdot \hat{V}^1 \leq V^j \cdot \hat{V}^j \leq 1$ for $2 \leq j \leq n$*

Proof. The first property from the lemma statement is trivially achieved by setting V^1 and $\hat{V}^1$ to the unit vectors in the direction of v and $\hat{v}$.

Let W be any $n \times n - 1$ matrix whose columns are orthogonal to V^1, and define $\hat{W}$ analogously. Let M and $\hat{M}$ be the orthogonal matrices

$$M \equiv \begin{bmatrix} V^1 & W \end{bmatrix} \qquad\qquad \hat{M} \equiv \begin{bmatrix} \hat{V}^1 & \hat{W} \end{bmatrix}$$

then $\hat{M}^T M$ is orthogonal. The second statement of the lemma is just a restatement of the fact that the first row and column of $\hat{M}^T M$ both have unit two-norm.

The third statement in the lemma statement requires a special choice of bases for the subspaces orthogonal to V^1 and $\hat{V}^1$. Let $\hat{W}^T W$ have the singular value decomposition

$$\hat{W}^T W = \hat{U} \Sigma U^T$$

where Σ is an $n-1 \times n-1$ diagonal matrix. If we take the sets $\{V^j\}_{j=2}^n$ and $\{\hat{V}^j\}_{j=2}^n$ to be the columns of WU and $\hat{W}\hat{U}$, then $\{V^j\}_{j=2}^n$ and $\{\hat{V}^j\}_{j=2}^n$ are orthonormal bases for the span of the spaces orthogonal to V^1 and $\hat{V}^1$. Furthermore, it follows from the fact that $(\hat{W}\hat{U})^T WU = \Sigma$ that $\{V^j\}_{j=2}^n$ and $\{\hat{V}^j\}_{j=2}^n$ are biorthogonal and $V^j \cdot \hat{V}^j$ is a singular value of $\hat{W}^T W$ when $2 \le j \le n$.

Finally, we must verify that $V^1 \cdot \hat{V}^1 \le V^j \cdot \hat{V}^j \le 1$ for $2 \le j \le n$. Let N and $\hat{N}$ be the orthogonal matrices

$$N \equiv \begin{bmatrix} V^1 & WU \end{bmatrix} \qquad\qquad \hat{N} \equiv \begin{bmatrix} \hat{V}^1 & \hat{W}\hat{U} \end{bmatrix}$$

Their product has the form

$$\hat{N}^T N = \begin{bmatrix} \langle \hat{V}^1, V^1 \rangle & (\hat{V}^1)^T (WU) \\ (\hat{W}\hat{U})^T V^1 & \Sigma \end{bmatrix}.$$

Each column of of $\hat{N}^T N$ has unit norm, in particular the last column contains $\sigma_{min}(\hat{W}^T W)$,█ the smallest singular value of $\hat{W}^T W$, and thus satisfies

$$1 \le \sigma_{min}(\hat{W}^T W)^2 + \|(\hat{V}^1)^T W U\|^2$$
$$= \min_j (V^j \cdot \hat{V}^j)^2 + \|(\hat{V}^1)^T W U\|^2$$
$$\le \min_j (V^j \cdot \hat{V}_j)^2 + 1 - (V^1 \cdot \hat{V}^1)^2.$$

It then follows from the final inequality and the assumption that $(V^1 \cdot \hat{V}^1) > 0$ that $V^j \cdot \hat{V}^j \ge V^1 \cdot \hat{V}^1$ for $2 \le j \le n$. □

Corollaries 6.7 and 6.8 establish a bound on $\|\hat{V}^T V - I\|$ in terms of the difference between the true and computed optimal directions at the current patch point.

COROLLARY 6.7. *Let V and $\hat{V}$ denote the orthogonal matrices whose columns are $\{V^j\}_{j=1}^n$ and $\{\hat{V}^j\}_{j=1}^n$ from Lemma 6.6, then*

$$\|I - \hat{V}^T V\|_2 \le 2\sqrt{2}\sqrt{1 - V_1 \cdot \hat{V}_1}$$

Proof. Partition the matrices V and $\hat{V}$ as

$$V = \begin{bmatrix} V_1 & W \end{bmatrix} \qquad\qquad \hat{V} = \begin{bmatrix} \hat{V}_1 & \hat{W} \end{bmatrix}.$$

W and $\hat{W}$ were constructed so that their columns have unit norm and satisfy $W^j \cdot \hat{W}^k = 0$ whenever $j \ne k$. Let $\Sigma = \hat{W}^T W$, where Σ is diagonal and its diagonal entries are $\hat{W}_j \cdot W_j$. $I - \hat{V}V$ can be expressed as

$$I - \hat{V}^T V = \begin{bmatrix} 1 - V_1 \cdot \hat{V}_1 & 0 \\ 0 & I - \Sigma \end{bmatrix} + \begin{bmatrix} 0 & \hat{V}_1^T W \\ \hat{W}^T V_1 & 0 \end{bmatrix}. \qquad (6.12)$$

By Lemma 6.6, $\Sigma_{j,j} = \hat{W}_j \cdot W_j \geq \hat{V}_1 \cdot V_1$, so the norm of the diagonal matrix appearing in the right hand side of (6.12) is bounded by

$$\left\| \begin{bmatrix} 1 - V_1 \cdot \hat{V}_1 & 0 \\ 0 & I - \Sigma \end{bmatrix} \right\|_2 \leq 1 - \hat{V}_1 \cdot V_1 \qquad (6.13)$$

Let x be a unit column vector partitioned as $x = [x_1 \; x_2]^T$, then the following bound on the norm of the second matrix in (6.12) follows from the fact that $\|\hat{V}_1^T W\|^2 = \|\hat{W}^T V_1\|^2 = 1 - \hat{V}_1 \cdot V_1^2$

$$\begin{aligned}
\left\| \begin{bmatrix} 0 & \hat{V}_1^T W \\ \hat{W}^T V_1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_2^2 &= \|\hat{V}_1^T W x_2\|^2 + \|\hat{W}^T V_1 x_1\|^2 \\
&\leq \|\hat{V}_1^T W\|^2 \|x_2\|^2 + \|\hat{W}^T V_1\|^2 (1 - \|x_2\|^2) \\
&\leq 2(1 - \hat{V}_1 \cdot V_1)
\end{aligned} \qquad (6.14)$$

The conclusion of the lemma follows from (6.13) and (6.14). □

COROLLARY 6.8. *Suppose v is a nonzero vector, and $\|v - \hat{v}\|_2 < \|v\|_2$, then there exist orthogonal matrices V and $\hat{V}$ such that their first columns are*

$$V^1 = \frac{v}{\|v\|} \qquad\qquad \hat{V}^1 = \frac{\hat{v}}{\|\hat{v}\|}$$

and

$$\|I - \hat{V}^T V\| \leq 4 \frac{\|v - \hat{v}\|}{\|v\| - \|v - \hat{v}\|}$$

Proof. The corollary follows from Corollary 6.7 and the fact that

$$\left\| \frac{v}{\|v\|} - \frac{\hat{v}}{\|\hat{v}\|} \right\| \leq 2 \frac{\|v - \hat{v}\|}{\|v\| - \|v - \hat{v}\|} \quad \text{whenever } \|v - \hat{v}\| < \|v\|.$$

□

The relevant interpretation of Lemma 6.9 for the purpose of proving Theorem 6.1 is that $\hat{V}^T V$ and the repeated Kronecker product $\hat{V}^T V$ of the orthogonal matrices from the convenient changes of coordinates are both perturbations of the identity of the same order. This fact is useful for bounding the difference between partial derivatives of the same function under the two changes of coordinates.

LEMMA 6.9. *If A and B are matrices such that AB is in $\mathbb{R}^{n \times n}$ and $\|AB - I_n\|_2 \leq 1$, then*

$$\|(AB \otimes \cdots \otimes AB) - I_{n^k}\|_2 \leq (2^k - 1)\|AB - I_n\|_2$$

where the Kronecker product on the left hand side of the inequality is repeated k times, and I_n and I_{n^k} denote the identity matrices in $\mathbb{R}^{n \times n}$ and $\mathbb{R}^{n^p \times n^p}$.

Proof. The proof of the lemma rests on the fact that for any two matrices C and D, $\|C \otimes D\|_2 = \|C\|_2 \|D\|_2$. A standard induction argument establishes the conclusion of the lemma. □

We now turn back to the task of bounding the difference between an exact and computed partial derivative of the optimal cost with respect to the original state space

variables at the current patch point. This difference is given by the formula

$$\frac{\partial^k \pi}{\partial x_{j_1} \cdots \partial x_{j_k}}(x^{i+1}) - \frac{\partial^p \pi^{i+1}}{\partial x_{j_1} \cdots \partial x_{j_k}}(x^{i+1}) =$$

$$\left[\frac{\partial}{\partial \xi} \otimes \cdots \otimes \frac{\partial}{\partial \xi}[\tilde{\pi}(\xi)]_{\xi=0} - \frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}}[\hat{\pi}^{i+1}(\hat{\xi})]_{\hat{\xi}=0} \right.$$

$$\left. - \frac{\partial}{\partial \xi} \otimes \cdots \otimes \frac{\partial}{\partial \xi}[\tilde{\pi}(\xi)]_{\xi=0}\left(V^T \hat{V} \otimes \cdots \otimes V^T \hat{V} - I \right) \right]$$

$$\hat{V}^T e_{j_1} \otimes \cdots \otimes \hat{V}^T e_{j_k}. \quad (6.15)$$

There are a few things to note about the right hand side of (6.15). The first is that it is the product of a row vector on the left and column vector on the right, where the column vector $\hat{V}^T e_{j_1} \otimes \cdots \otimes \hat{V}^T e_{j_k}$ has unit norm since it is the column of an orthogonal matrix. It follows from the Cauchy-Schwarz inequality that the the right hand side of (6.15) is bounded by the norm of the row vector. This row vector has two terms, the first is the difference between the exact and computed partial derivatives of the optimal cost with respect to the changes of variables. The second is involves a vector composed of partial derivatives of the exact solution times a small perturbation of the identity. It follows from assumption that the optimal cost is smooth on the compact computational domain and by Lemma 6.5 that there exist $T' < \infty$ such that, whenever the current patch point is within a radius of the maximum consecutive patch point distance of the previous patch point, this term is bounded by

$$\left\| \frac{\partial}{\partial \xi} \otimes \cdots \otimes \frac{\partial}{\partial \xi}[\tilde{\pi}(\xi)]_{\xi=0}\left(V^T \hat{V} \otimes \cdots \otimes V^T \hat{V} - I \right) \right\|$$

$$\leq \left(2^k - 1 \right) \sup_{\overline{X_c \setminus \mathcal{A}}} \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[\pi(x)] \right\| \left\| V^T \hat{V} - I \right\| \quad (6.16)$$

$$\leq T' \|x^{i+1} - x^i\|^{d+1}.$$

This bound is independent of p, the order of the partial derivative, when $p \geq 2$. The remaining work is in establishing the existence of some $T'' < \infty$, such that under the convenient changes of variables

$$\left\| \frac{\partial}{\partial \xi} \otimes \cdots \otimes \frac{\partial}{\partial \xi}[\tilde{\pi}(\xi)]_{\xi=0} - \frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}}[\hat{\pi}^{i+1}(\hat{\xi})]_{\hat{\xi}=0} \right\|_2$$

$$\leq T'' \|x^{i+1} - x^i\|^{d+2-k} \quad (6.17)$$

whenever the distance between the current and previous patch point is less than the maximum consecutive patch point distance. We will bound the left hand side of (6.17) by bounding each of its entries by considering two cases, depending on whether the entry is the difference of exact and computed characteristic or non characteristic partial derivatives of the optimal cost.

With this approach in mind, we will need a bound on the partial derivatives of the problem data with respect to two changes of variables. This is the content of Lemma 6.10.

LEMMA 6.10. *Suppose all partial derivatives of $s : \mathbb{R}^n \to \mathbb{R}$ up to order k exist and are continuous in a neighborhood of $\bar{x}$ in $\mathbb{R}^n$. If $\tilde{V}$ and $\hat{V}$ are orthogonal matrices*

and $\tilde{s}$ and $\hat{s}$ are defined as

$$\tilde{s}(\tilde{\xi}) \equiv s(\bar{x} + \tilde{V}\tilde{\xi}) \qquad\qquad \hat{s}(\hat{\xi}) \equiv s(\bar{x} + \hat{V}\hat{\xi})$$

then

$$\left| \frac{\partial^k \tilde{s}}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{s}}{\partial \hat{\xi}_{j_1} \cdots \partial \hat{\xi}_{j_k}}(0) \right| \le$$

$$\left(2^k - 1\right) \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x)]_{x=\bar{x}} \right\|_2 \|\hat{V}^T \tilde{V} - I\|_2$$

If the components of $G : \mathbb{R}^n \to \mathbb{R}^n$ all have partial derivatives up to order p in a neighborhood of $\bar{x}$ in $\mathbb{R}^n$, and $\tilde{G}$ and $\hat{G}$ are defined as

$$\tilde{G}(\tilde{\xi}) \equiv \tilde{V}^T G(\bar{x} + \tilde{V}\tilde{\xi}) \qquad\qquad \hat{G}(\hat{\xi}) \equiv \hat{V}^T G(\bar{x} + \hat{V}\hat{\xi})$$

then

$$\left\| \frac{\partial^k \tilde{G}}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{G}}{\partial \hat{\xi}_{j_1} \cdots \partial \hat{\xi}_{j_k}}(0) \right\| \le$$

$$2^k n^{1/2} \max_\ell \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[G_\ell(x)]_{x=\bar{x}} \right\|_2 \|\hat{V}^T \tilde{V} - I\|_2.$$

Proof. A kth order partial derivative of $\tilde{s}$ can be expressed in Kronecker notation as

$$\frac{\partial^k \tilde{s}}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}(0) = \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x)]_{x=\bar{x}}(\tilde{V} \otimes \cdots \otimes \tilde{V})(e_{j_1} \otimes \cdots \otimes e_{j_k})$$

so after a sequence of algebraic manipulations, the difference between the partial derivatives of $\tilde{s}$ and $\hat{s}$ is

$$\left| \frac{\partial^k \tilde{s}}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{s}}{\partial \hat{\xi}_{j_1} \cdots \partial \hat{\xi}_{j_k}}(0) \right|$$

$$= \left| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s]_{x=\bar{x}}\left(V\hat{V}^T \otimes \cdots \otimes V\hat{V}^T - I\right)\left(\hat{V}^{j_1} \otimes \cdots \otimes \hat{V}^{j_k}\right) \right|$$

$$\le \left(2^k - 1\right) \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x)]_{x=\bar{x}} \right\|_2 \|\tilde{V}\hat{V}^T - I\|_2.$$

The last inequality follows from the Cauchy-Schwarz inequality, Lemma 6.9, and the facts that $\hat{V}^{j_1} \otimes \cdots \otimes \hat{V}^{j_k}$ has unit norm and $\|\tilde{V}\hat{V}^T - I\|_2 = \|\hat{V}^T\tilde{V} - I\|_2$.

The difference between the partial derivatives of G under the two changes of variables is bounded by

$$\left\| \frac{\partial^k}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}[G(\bar{x} + \tilde{V}\tilde{\xi})]_{\tilde{\xi}=0} - \frac{\partial^k}{\partial \hat{\xi}_{j_1} \cdots \partial \hat{\xi}_{j_k}}[G(\bar{x} + \hat{V}\hat{\xi})]_{\hat{\xi}=0} \right\|_2$$

$$\le \left(2^k - 1\right) n^{1/2} \max_\ell \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[G_\ell(x)]_{x=\bar{x}} \right\|_2 \|\tilde{V}\hat{V}^T - I\|_2.$$

This bound is derived by using the result of the first half of the lemma to bound the infinity norm of the difference between the partial derivatives. The second result of the lemma follows from the previous bound and the identity

$$\tilde{G}(\tilde{\xi}) - \hat{G}(\hat{\xi}) = \hat{V}^T\big[G(\bar{x} + \tilde{V}\tilde{\xi}) - G(\bar{x} + \hat{V}\hat{\xi}) + (\hat{V}\hat{V}^T - I)G(\bar{x} + \tilde{V}\tilde{\xi})\big].$$

□

The next corollary differs from the previous lemma in that it bounds the difference in partial derivatives of two different functions with respect to two changes of variables.

COROLLARY 6.11. *Suppose all partial derivatives of* $s, \bar{s} : \mathbb{R}^n \to \mathbb{R}$ *up to order* k *exist and are continuous in a neighborhood of* $\bar{x}$ *in* $\mathbb{R}^n$. *If* $\tilde{V}$ *and* $\hat{V}$ *are orthogonal matrices and* $\tilde{s}$ *and* $\hat{s}$ *are defined as*

$$\tilde{s}(\tilde{\xi}) \equiv s(\bar{x} + \tilde{V}\tilde{\xi}) \qquad\qquad \hat{s}(\hat{\xi}) \equiv \bar{s}(\bar{x} + \hat{V}\hat{\xi})$$

then

$$\left|\frac{\partial^k \tilde{s}}{\partial\tilde{\xi}_{j_1}\cdots\partial\tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{s}}{\partial\hat{\xi}_{j_1}\cdots\partial\hat{\xi}_{j_k}}(0)\right| \leq$$

$$\left\|\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x) - \bar{s}(x)]_{x=\bar{x}}\right\|_2$$

$$+ \left(2^k - 1\right)\left\|\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x)]_{x=\bar{x}}\right\|_2 \|\hat{V}^T\tilde{V} - I\|_2$$

Proof.

$$\left|\frac{\partial^k \tilde{s}}{\partial\tilde{\xi}_{j_1}\cdots\partial\tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{s}}{\partial\hat{\xi}_{j_1}\cdots\partial\hat{\xi}_{j_k}}(0)\right|$$

$$= \left|\left(\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x) - \bar{s}(x)]_{x=\bar{x}}\right)\left(\hat{V}^{j_1} \otimes \cdots \otimes \hat{V}^{j_k}\right)\right.$$

$$\left. + \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x)]_{x=\bar{x}}\left(\hat{V}^T\tilde{V} \otimes \cdots \otimes \hat{V}^T\tilde{V}\right)(e_{j_1} \otimes \cdots \otimes e_{j_k})\right|$$

$$\leq \left\|\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x) - \bar{s}(x)]_{x=\bar{x}}\right\|_2$$

$$+ \left(2^k - 1\right)\left\|\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}[s(x)]_{x=\bar{x}}\right\|_2 \|\hat{V}^T\tilde{V} - I\|_2$$

□

We now have all the ingredients necessary to prove Theorem 6.1.

Proof. [Proof of Theorem 6.1] We denote the problem data under the two changes of variables in (6.1) as

$$\begin{aligned}
\tilde{f}(\xi) &\equiv V^T f(x^{i+1} + V\xi) & \hat{f}(\hat{\xi}) &\equiv \hat{V}^T f(x^{i+1} + \hat{V}\hat{\xi}) \\
\tilde{g}(\xi) &\equiv V^T g(x^{i+1} + V\xi) & \hat{g}(\hat{\xi}) &\equiv \hat{V}^T g(x^{i+1} + \hat{V}\hat{\xi}) \\
\tilde{q}(\xi) &\equiv q(x^{i+1} + V\xi) & \hat{q}(\hat{\xi}) &\equiv \hat{q}(x^{i+1} + \hat{V}\hat{\xi}) \\
\tilde{r}(\xi) &\equiv r(x^{i+1} + V\xi) & \hat{r}(\hat{\xi}) &\equiv \hat{r}(x^{i+1} + \hat{V}\hat{\xi})
\end{aligned} \qquad (6.18)$$

We first consider the case of the second order partial derivatives of the optimal cost, so we must verify that the left hand side of (6.17) is $\mathcal{O}(\|x^{i+1} - x^i\|^d)$. The characteristic

31

second order partial derivatives of the optimal cost are given by (2.11) and the computed optimal control is given by (2.13). The exact second order partial derivatives of the optimal const and the optimal control satisfy the equations

$$\frac{\partial^2 \tilde{\pi}}{\partial \xi_j \partial \xi_1}(0) = -\frac{1}{\|\dot{x}\|} \left[\frac{\partial \tilde{\pi}}{\partial \xi}(0) \left(\frac{\partial \tilde{f}}{\partial \xi_j}(0) + \frac{\partial \tilde{g}}{\partial \xi_j}(0)\tilde{\kappa}(0) \right. \right.$$
$$\left. \left. + \frac{\partial \tilde{q}}{\partial \xi_j}(0) + \frac{1}{2}\frac{\partial \tilde{r}}{\partial \xi_j}(0)\tilde{\kappa}(0)^2 \right) \right] \quad (6.19)$$

and

$$\tilde{\kappa}(0) = -\frac{1}{\tilde{r}(0)} \left(\frac{\partial \tilde{\pi}}{\partial \xi}(0)\tilde{g}(0) \right) \quad (6.20)$$

The difference between $\frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_j \partial \hat{\xi}_1}(0)$ and $\frac{\partial^2 \tilde{\pi}}{\partial \xi_j \partial \xi_1}(0)$ can be bounded in terms of two types of differences. The first type involves terms that are the difference between the partial derivatives with respect to the two changes of variables of the problem data.

$$\begin{array}{ll}
\dfrac{\partial \tilde{f}}{\partial \xi}(0) - \dfrac{\partial \hat{f}}{\partial \hat{\xi}}(0) & \dfrac{\partial \tilde{g}}{\partial \xi}(0) - \dfrac{\partial \hat{g}}{\partial \hat{\xi}}(0) \\[2mm]
\dfrac{\partial \tilde{q}}{\partial \xi}(0) - \dfrac{\partial \hat{q}}{\partial \hat{\xi}}(0) & \dfrac{\partial \tilde{r}}{\partial \xi}(0) - \dfrac{\partial \hat{r}}{\partial \hat{\xi}}(0)
\end{array} \quad (6.21)$$

If the distance between the current and previous patch point is less than the maximum consecutive patch point distance h, then by Lemmas 6.5 and 6.10, there is a single bound that is proportional to $\|x^{i+1} - x^i\|^{d+1}$ that bounds each difference in (6.21).

The second type of difference involves terms that are the difference between exact and computed values of derivatives with respect to the changes of coordinates of the optimal cost. These terms are

$$\frac{\partial \tilde{\pi}}{\partial \xi}(0) - \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0) \qquad \tilde{\kappa}(0) - \hat{\kappa}^{i+1}(0) \qquad \frac{1}{\|\dot{x}^{i+1}\|} - \frac{1}{\|\dot{x}\|} \quad (6.22)$$

By Theorem 4.1, the difference between the computed and exact first order partial derivatives of the optimal cost is bounded by

$$\left\| \frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \pi^{i+1}}{\partial x}(x^{i+1}) \right\| \leq T'\|x^{i+1} - x^i\|^{d+1} \quad (6.23)$$

for some $T' < \infty$ and by Lemma 6.5, there exists $B < \infty$ such that

$$\|\hat{V}^T V - I\| \leq B\|x^{i+1} - x^i\|^{d+1}. \quad (6.24)$$

Each difference in (6.22) shares a single upper bound that is a linear combination of the right hand sides of (6.23) and (6.24). The first difference is bounded by

$$\left\| \frac{\partial \tilde{\pi}}{\partial \xi}(0) - \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0) \right\| = \left\| \left[\frac{\partial \pi}{\partial x}(x^{i+1}) - \frac{\partial \pi^{i+1}}{\partial x}(x^{i+1}) + \frac{\partial \pi}{\partial x}(x^{i+1})(V\hat{V}^T - I) \right] \hat{V} \right\|$$
$$\leq \left(T' + B \sup_{\overline{X_c \setminus \mathcal{A}}} \left\| \frac{\partial \pi}{\partial x}(x) \right\| \right) \|x^{i+1} - x^i\|^{d+1}$$

32

and the second difference obeys

$$\|\tilde{\kappa}(0) - \hat{\kappa}^{i+1}(0)\| \leq \left(T' \sup_{X_c\setminus\mathcal{A}} \frac{\|g(x)\|}{r(x)}\right)\|x^{i+1} - x^i\|^{d+1}.$$

By assumption, the distance between x^i and x^{i+1} is less than the maximum consecutive patch point distance h of Corollary 4.2 , so

$$\sup_{\substack{X_c\setminus\mathcal{A} \\ \|x^{i+1}-x^i\|\leq h}} \left(\|\dot{x}\|(\|\dot{x}\| - \|\dot{x} - \dot{x}^{i+1}\|)\right)^{-1} < \infty \tag{6.25}$$

and the third difference in (6.22) is bounded by

$$\left|\frac{1}{\|\dot{x}\|} - \frac{1}{\|\dot{x}^{i+1}\|}\right| \leq \left(T' \sup_{\substack{X_c\setminus\mathcal{A} \\ \|x^{i+1}-x^i\|\leq h}} \frac{\|g(x)\|^2}{r(x)\left(\|\dot{x}\|(\|\dot{x}\| - \|\dot{x} - \dot{x}^{i+1}\|)\right)}\right)\|x^{i+1} - x^i\|^{d+1}$$

Our next task is to demonstrate the contribution from the second order non characteristic partial derivatives to (6.17) is $\mathcal{O}(\|x^{i+1} - x^i\|^d)$. Let $\hat{\pi}^{i+1}(\hat{\xi}) \equiv \bar{\pi}(x^{i+1} + \hat{V}\hat{\xi})$ where $\bar{\pi}$ denotes the Taylor polynomial of the optimal cost centered at $x = x^i$. If neither j_1 nor j_2 are one, then by Corollary 6.11, the difference between the non characteristic partial derivatives is bounded by

$$\left|\frac{\partial^2\tilde{\pi}}{\partial\xi_{j_1}\partial\xi_{j_2}}(0) - \frac{\partial^2\hat{\pi}^{i+1}}{\partial\hat{\xi}_{j_1}\partial\hat{\xi}_{j_2}}(0)\right| \leq (2^2 - 1)\left\|\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x}[\pi(x)]_{x=x^{i+1}}\right\|\|V^T\hat{V} - I\|$$

$$+ (2^2 - 1)\left\|\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x}[\pi(x) - \bar{\pi}(x)]_{x=x^{i+1}}\right\| \tag{6.26}$$

The first term on the right hand side of the inequality (6.26) is $\mathcal{O}(\|x^{i+1} - x^i\|^{d+1})$ by the assumption that π is smooth on the compact computational domain, and (6.24). Since $\bar{\pi}$ is the Taylor polynomial of π centered at x^i, then each entry in $\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x}[\pi(x) - \bar{\pi}(x)]_{x=x^{i+1}}$ is bounded by $R_T\|x^{i+1} - x^i\|^d$, so the second term on the right hand side of the inequality in (6.26) is $\mathcal{O}(\|x^{i+1} - x^i\|^d)$.

We can conclude that the bound in (6.17) holds for $k = 2$, so the conclusion of the theorem holds in the case of the second order partial derivatives of the optimal cost.

By the formula for $\frac{\partial\hat{\kappa}^{i+1}}{\partial\hat{\xi}_j}(0)$ in (2.13) and the same reasoning we just employed, we can conclude that the difference between $\frac{\partial\hat{\kappa}^{i+1}}{\partial\hat{\xi}_j}(0)$ and $\frac{\partial\tilde{\kappa}}{\partial\hat{\xi}_j}(0)$ is $\mathcal{O}(\|x^{i+1} - x^i\|^d)$.

The proof that the higher order partial derivatives satisfy the inequality in (6.17) is analogous to the second order case. We outline the proof of the third order case, which is representative of the higher order cases. The computed third order characteristic partial derivatives with respect to the change of variables of the optimal cost is calculated from the formula in §9.2, and the exact characteristic partial derivatives have an analogous formula. As in the case of the second order characteristic partial derivatives of the optimal cost, the bound on the difference between the exact and computed third order characteristic partial derivatives depends linearly on two types of differences. The first type of difference is the difference between the partial derivatives of problem data functions with respect to the two changes of variables, and each one of these differences is $\mathcal{O}(\|x^{i+1} - x^i\|^{d-1})$. The other type of difference

33

is the difference between the exact and computed partial derivatives with respect to the two changes of variables of the solution. These differences are

$$
\begin{array}{cc}
\dfrac{1}{\|\dot{x}\|} - \dfrac{1}{\|\dot{x}^{i+1}\|} & \\[2mm]
\dfrac{\partial \tilde{\pi}}{\partial \xi}(0) - \dfrac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0) & \tilde{\kappa}(0) - \hat{\kappa}^{i+1}(0) \\[2mm]
& \dfrac{\partial \tilde{\kappa}}{\partial \xi_j}(0) - \dfrac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_j}(0) \\[2mm]
\dfrac{\partial^2 \tilde{\pi}}{\partial \xi_j \partial \xi}(0) - \dfrac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_j \partial \hat{\xi}}(0) &
\end{array}
\tag{6.27}
$$

Each one of these differences was treated in the second order case and is $\mathcal{O}(\|x^{i+1} - x^i\|^{d-1})$.

In the case of bounding the non characteristic partial derivatives, if none of the coordinate indices j_1, j_2, nor j_3 are one, then we may apply Corollary 6.11 to the difference between the non characteristic partial derivatives

$$
\begin{aligned}
\left| \frac{\partial^3 \tilde{\pi}}{\partial \xi_{j_1} \partial \xi_{j_2} \partial \xi_{j_3}}(0) \right. & \left. - \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) \right| \\
& \leq \left(2^3 - 1\right) \left\| \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} [\pi(x)]_{x=x^{i+1}} \right\| \left\| V^T \hat{V} - I \right\| \\
& + \left(2^3 - 1\right) \left\| \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} [\pi(x) - \bar{\pi}(x)]_{x=x^{i+1}} \right\|
\end{aligned}
\tag{6.28}
$$

The first term on the right hand side of the inequality in (6.28) is $\mathcal{O}(\|x^{i+1} - x^i\|^{d+1})$ by the assumption that π is smooth on the compact computational domain, and (6.24). Since $\bar{\pi}$ is the Taylor polynomial of π centered at x^i, then each entry in $\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} [\pi(x) - \bar{\pi}(x)]_{x=x^{i+1}}$ is bounded by $R_T \|x^{i+1} - x^i\|^{d-1}$, so the second term on the right hand side of the inequality in (6.28) is $\mathcal{O}(\|x^{i+1} - x^i\|^{d-1})$.
□

7. Lipschitz condition for second and higher order partial derivatives. In this section we prove that ϕ_j satisfies the Lipschitz condition when $j \geq 2$ under the conditions on the optimal control problem in §2.1, most importantly that the optimal cost is a strict Lyupanov function.

THEOREM 7.1. *Under the conditions on the optimal control problem in §2.1, there exists a maximum step size $h_L > 0$, maximum multiplier $M_{max} > 0$, and Lipschitz constant L_2 such that for all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a distance of h_L of each other, and for all $M \leq M_{max}$,*

$$
\|\phi_k(C^i) - \phi_k(\hat{C}^i)\| \leq L_2 M \|x^{i+1} - x^i\|^{d+2-k} \text{ for } 2 \leq k \leq d+1 \tag{7.1}
$$

whenever

$$
\|C_j^i - \hat{C}_j^i\| \leq M \|x^{i+1} - x^i\|^{d+2-j} \text{ for } 0 \leq j \leq d+1
$$

where C^i is the vector of exact partial derivatives up to order $d+1$ of the optimal cost at x^i.

The proof mirrors the proof of the local truncation error result of Theorem 6.1. We will delay the proof of the theorem until after we have introduced some notation and

34

established some relevant bounds. We denote the two approximate optimal directions at x^{i+1} computed from C^i and $\hat{C}^i$ as

$$\dot{x}_c^{i+1} \equiv f(x^{i+1}) - \frac{\phi_1(C^i) \cdot g(x^{i+1})}{r(x^{i+1})} g(x^{i+1})$$

and

$$\dot{x}_{\hat{c}}^{i+1} \equiv f(x^{i+1}) - \frac{\phi_1(\hat{C}^i) \cdot g(x^{i+1})}{r(x^{i+1})} g(x^{i+1})$$

As in the case of the local truncation error result, we will work with two changes of variables that we choose to be convenient. These changes of coordinates are

$$x = x^{i+1} + \tilde{V}\tilde{\xi} \qquad\qquad \tilde{V}^1 = \frac{\dot{x}_c^{i+1}}{\|\dot{x}_c^{i+1}\|} \qquad\qquad (7.2)$$

and

$$x = x^{i+1} + \hat{V}\hat{\xi} \qquad\qquad \hat{V}^1 = \frac{\dot{x}_{\hat{c}}^{i+1}}{\|\dot{x}_{\hat{c}}^{i+1}\|} \qquad\qquad (7.3)$$

We denote the approximate optimal cost and optimal control polynomials under the two changes of variables as

$$\tilde{\pi}^{i+1}(\tilde{\xi}) \qquad\qquad\qquad \hat{\pi}^{i+1}(\hat{\xi})$$
$$\tilde{\kappa}^{i+1}(\tilde{\xi}) \qquad\qquad\qquad \hat{\kappa}^{i+1}(\hat{\xi})$$

and we denote the problem data functions under the two changes of variables as

$$
\begin{aligned}
\tilde{f}(\tilde{\xi}) &\equiv \tilde{V}^T f(x^{i+1} + \tilde{V}\tilde{\xi}) & \hat{f}(\hat{\xi}) &\equiv \hat{V}^T f(x^{i+1} + \hat{V}\hat{\xi}) \\
\tilde{g}(\tilde{\xi}) &\equiv \tilde{V}^T g(x^{i+1} + \tilde{V}\tilde{\xi}) & \hat{g}(\hat{\xi}) &\equiv \hat{V}^T g(x^{i+1} + \hat{V}\hat{\xi}) \\
\tilde{q}(\tilde{\xi}) &\equiv q(x^{i+1} + \tilde{V}\tilde{\xi}) & \hat{q}(\hat{\xi}) &\equiv q(x^{i+1} + \hat{V}\hat{\xi}) \\
\tilde{r}(\tilde{\xi}) &\equiv r(x^{i+1} + \tilde{V}\tilde{\xi}) & \hat{r}(\hat{\xi}) &\equiv r(x^{i+1} + \hat{V}\hat{\xi})
\end{aligned}
\qquad (7.4)
$$

In the course of proving Theorem 7.1, we will need a guarantee that certain quantities depending on the computed solution are finite. Lemma 7.2 provides this guarantee.

LEMMA 7.2. *If the optimal control problem satisfies the assumptions of §2.1, then there exists an $M_{max} > 0$ and a maximum consecutive patch point distance $h_2 > 0$ such that, for all $M \leq M_{max}$ and for all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$,*

$$\inf \|\dot{x}_c^{i+1}\| > 0$$
$$\inf \left(\|\dot{x}_c^{i+1}\| - \|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\| \right) > 0 \qquad (7.5)$$

whenever

$$\|x^{i+1} - x^i\| \leq h_2 \text{ and } \|C_j^i - \hat{C}_j^i\| \leq M\|x^{i+1} - x^i\|^{d+2-j} \text{ for } 0 \leq j \leq d+1$$

where the infimums are taken over all $M \leq M_{max}$ and all x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ such that $x^{i+1} - x^i \leq h_2$. C^i holds the exact partial derivatives of π at $x^i \in \overline{X_c \setminus \mathcal{A}}$.

Proof. The exact optimal direction at x^{i+1} is $\dot{x} = f(x^{i+1}) + g(x^{i+1})\kappa(x^{i+1})$. C^i holds the exact partial derivatives of the optimal cost at x^i up to order $d + 1$, so by Theorem 4.1,

$$\lim_{x^{i+1} \to x^i} \left((f(x^{i+1}) + g(x^{i+1})\kappa(x^{i+1})) - \dot{x}_c^{i+1} \right) = 0. \tag{7.6}$$

The assumption that π is a strict Lyapunov function guarantees that the infimum over $\overline{X_c \setminus \mathcal{A}}$ of $\|\dot{x}\|$ is strictly greater than zero. The limit in (7.6) is uniform, so for all ϵ_1 in $0 < \epsilon_1 < 1$, there exists a $\delta_1 > 0$ independent of x^i and x^{i+1} such that

$$\|f(x^{i+1}) + g(x^{i+1})\kappa(x^{i+1}) - \dot{x}_c^{i+1}\| \leq \epsilon_1 \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\|$$

Therefore

$$\begin{aligned}
\|\dot{x}_c^{i+1}\| &\geq \|f(x^{i+1}) + g(x^{i+1})\kappa(x^{i+1})\| - \|f(x^{i+1}) + g(x^{i+1})\kappa(x^{i+1}) - \dot{x}_c^{i+1}\| \\
&\geq (1 - \epsilon_1) \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\| \\
&> 0
\end{aligned}$$

Thus we have established the first inequality of the lemma statement.

We will assume for the rest of the proof that $M \leq M_{max}$. By Theorem 5.1, if $\|C_j^i - \hat{C}_j^i\| \leq M\|x^{i+1} - x^i\|^{d+2-j}$ for $0 \leq j \leq d + 1$, then

$$\lim_{x^{i+1} \to x^i} \dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1} = 0. \tag{7.7}$$

The limit in (7.7) is uniform, so for all ϵ_2 in $0 < \epsilon_2 < 1$, there exists a $\delta_2 > 0$ independent of x^i and x^{i+1} such that

$$\|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\| \leq \epsilon_2 \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\|$$

whenever

$$\|x^{i+1} - x^i\| \leq \delta_2 \text{ and } \|C_j^i - \hat{C}_j^i\| \leq M\|x^{i+1} - x^i\|^{d+2-j} \text{ for } 0 \leq j \leq d + 1.$$

Therefore, if $\|\dot{x}_c^{i+1}\| \geq (1 - \epsilon_1) \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\|$ where $0 < \epsilon_1 < 1$, then

$$\begin{aligned}
\|\dot{x}_c^{i+1}\| - \|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\| &\geq (1 - \epsilon_1) \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\| - \epsilon_2(1 - \epsilon_1) \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\| \\
&= (1 - \epsilon_1)(1 - \epsilon_2) \inf_{\overline{X_c \setminus \mathcal{A}}} \|\dot{x}\| \\
&> 0
\end{aligned}$$

whenever

$$\|x^{i+1} - x^i\| \leq \delta_1 \text{ and } \|C_j^i - \hat{C}_j^i\| \leq M\|x^{i+1} - x^i\|^{d+2-j} \text{ for } 0 \leq j \leq d + 1$$

so the second infimum of the lemma statement has been established. We take h_2 to be the smaller of δ_1 and δ_2. $\square$

Corollary 7.3 provides a guarantee that any term depending on the difference between the normalized optimal directions at x^{i+1} computed from C^i and $\hat{C}^i$ does not affect the order of the main bound (7.1) in Theorem 7.1.

COROLLARY 7.3. *If the optimal cost π is a strict Lyapunov function on the computational domain X_c, then there exists a Lipschitz constant L', maximum multiplier $M_{max} > 0$ and a maximum consecutive patch point distance $h_2 > 0$ such that for all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a distance of h_2 of each other, and for all $M \leq M_{max}$,*

$$\left\| \frac{\dot{x}_c^{i+1}}{\|\dot{x}_c^{i+1}\|} - \frac{\dot{x}_{\hat{c}}^{i+1}}{\|\dot{x}_{\hat{c}}^{i+1}\|} \right\| \leq L'M\|x^{i+1} - x^i\|^{d+1}$$

whenever

$$\|C_j^i - \hat{C}_j^i\| \leq M\|x^{i+1} - x^i\|^{d+2-j} \text{ for } 0 \leq j \leq d+1$$

where C^i holds the exact partial derivatives of π at $x^i \in \overline{X_c \setminus \mathcal{A}}$.

Proof. The bound in the corollary statement follows from the Lemma 7.2 and the inequality

$$\left\| \frac{\dot{x}_c^{i+1}}{\|\dot{x}_c^{i+1}\|} - \frac{\dot{x}_{\hat{c}}^{i+1}}{\|\dot{x}_{\hat{c}}^{i+1}\|} \right\| \leq \frac{\|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\|}{\|\dot{x}_c^{i+1}\| - \|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\|}.$$

□

As in the case of proving the local truncation error result of Theorem 6.1, we will pick the changes of coordinates in (7.2) and (7.3) so that the orthogonal matrices $\tilde{V}$ and $\hat{V}$ are almost inverses of each other. This is the content of Corollary 7.4.

COROLLARY 7.4. *If the optimal cost π is a strict Lyapunov function on the computational domain X_c, then there exists a Lipschitz constant L', maximum multiplier $M_{max} > 0$ and a maximum consecutive patch point distance $h_2 > 0$, and orthogonal matrices $\hat{V}$ and $\tilde{V}$ such that their first columns are*

$$\tilde{V}^1 = \frac{\dot{x}_c^{i+1}}{\|\dot{x}_c^{i+1}\|} \qquad\qquad \hat{V}^1 = \frac{\dot{x}_{\hat{c}}^{i+1}}{\|\dot{x}_{\hat{c}}^{i+1}\|}$$

so that for all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a distance of h_2 of each other and all $M \leq M_{max}$,

$$\|\hat{V}^T\tilde{V} - I\| \leq L'M\|x^{i+1} - x^i\|^{d+1}$$

whenever

$$\|C_j^i - \hat{C}_j^i\| \leq M\|x^{i+1} - x^i\|^{d+2-j} \text{ for } 0 \leq j \leq d+1$$

where C^i holds the exact partial derivatives of π at $x^i \in \overline{X_c \setminus \mathcal{A}}$, then

Proof. For a sufficiently small maximum consecutive patch point distance h_2, Lemma 7.2 guarantees $\dot{x}_c^{i+1}$ and $\dot{x}_{\hat{c}}^{i+1}$ are close enough so that Lemma 6.6, and Corollaries 6.7 and 6.8 imply the bound on $\hat{V}^T\tilde{V} - I$ of the corollary statement. □

As in the proof of Theorem 6.1, we bound the differences between the computed partial derivatives of the optimal cost with respect to the two changes of coordinates, and then show that the order of this bound does not change after recovering the two

sets of partial derivatives with respect to the original state space coordinates and compare their differences. Lemma 7.5 provides a bound on the difference between the characteristic partial derivatives, which are computed by inheritance.

LEMMA 7.5. *Suppose the orthogonal matrices $\tilde{V}$ and $\hat{V}$ from the changes of coordinates in (7.2) and (7.3) are the ones from Corollary 7.4. If the optimal cost π is a strict Lyapunov function on the computational domain X_c, then there exists a Lipschitz constant $L'_{inh} < \infty$, $M_{max} > 0$ and a maximum consecutive patch point distance $h_2 > 0$ such that the partial derivatives computed by inheritance satisfy*

$$\left| \frac{\partial^k \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \cdots \partial \hat{\xi}_{j_k}}(0) \right| \le L'_{inh} \|x^{i+1} - x^i\|^{d+2-k}$$

whenever

$$\|x^{i+1} - x^i\| \le h_2 \ \text{and} \ \|C^i_j - \hat{C}^i_j\| \le M \|x^{i+1} - x^i\|^{d+2-j} \ \text{for} \ 0 \le j \le d+1$$

Proof. P is linear in its first argument, and $\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}$ is linear, so by Lemma 3.4

$$\left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(C^i, x - x^i)]_{x=x^{i+1}} \right.$$
$$\left. - \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(\hat{C}^i, x - x^i)]_{x=x^{i+1}} \right\|$$
$$= \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(C^i - \hat{C}^i, x)]_{x=x^{i+1}-x^i} \right\|$$
$$\le n^{k/2} L_{inh} M \|x^{i+1} - x^i\|^{d+2-k}$$

where $\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}$ is repeated k times.

It follows from the local truncation error result of Theorems 4.1 and 6.1 that there is some $T < \infty$ such that

$$\left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(C^i, x - x^i) - \pi(x) + \pi(x)]_{x=x^{i+1}} \right\|$$
$$\le T \|x^{i+1} - x^i\|^{d+2-k} + \sup_{X_c \setminus \mathcal{A}} \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [\pi(x)]_{x=x^{i+1}} \right\|$$

where the supremum is finite since π is smooth by assumption on the compact computational domain. It follows from Corollary 7.4 and Lemma 6.9 that

$$\left\| \tilde{V}\hat{V}^T \otimes \cdots \otimes \tilde{V}\hat{V}^T - I \right\| \le (2^k - 1) L'_1 M \|x^{i+1} - x^i\|^{d+1}$$

The conclusion of the lemma follows by applying the previous three inequalities to the following bound on the difference between the two computed characteristic partial

derivatives of the optimal cost.

$$\left| \frac{\partial^k \tilde{\pi}^{i+1}}{\partial \tilde{\xi}_{j_1} \cdots \partial \tilde{\xi}_{j_k}}(0) - \frac{\partial^k \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \cdots \partial \hat{\xi}_{j_k}}(0) \right|$$

$$= \left| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(C^i, x)]_{x=x^{i+1}} \tilde{V}^{j_1} \otimes \cdots \otimes \tilde{V}^{j_k} \right.$$

$$\left. - \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(\hat{C}^i, x)]_{x=x^{i+1}} \hat{V}^{j_1} \otimes \cdots \otimes \hat{V}^{j_k} \right|$$

$$\leq \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(C^i - \hat{C}^i, x)]_{x=x^{i+1}} \right\|$$

$$+ \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [P(C^i, x)]_{x=x^{i+1}} \left(\tilde{V}\hat{V}^T \otimes \cdots \otimes \tilde{V}\hat{V}^T - I \right) \right\|$$

□

We are now in a position to prove Theorem 7.1.

Proof. [Proof of Theorem 7.1] The proof of the theorem mirrors the proof of the local truncation error result in Theorem 6.1. We may assume that the orthogonal matrices $\tilde{V}$ and $\hat{V}$ appearing in the changes of coordinates of (7.2) and (7.3) are convenient, that is they are taken from Corollary 7.4. Starting with the computed second order partial derivatives, we will show that the main bound (7.1) of the theorem holds, by bounding the individual entries of $\phi_k(C^i) - \phi_k(\hat{C}^i)$. We do this by considering the cases of the computed characteristic and non characteristic partial derivatives of the optimal cost independently, and then proceed by induction for the higher order partial derivatives.

We denote the computed optimal cost polynomials computed from the coefficient sets C^i and $\hat{C}^i$ and centered at x^{i+1} as π_c^{i+1} and $\pi_{\hat{c}}^{i+1}$. They are given by the formulas

$$\pi_c^{i+1}(x) \equiv P(\phi(C^i), x - x^{i+1}) \qquad \pi_{\hat{c}}^{i+1}(x) \equiv P(\phi(\hat{C}^i), x - x^{i+1})$$

At x^{i+1}, the difference between a pth order partial derivative of the two computed optimal cost polynomials is bounded by

$$\left| \frac{\partial^p \pi_c^{i+1}}{\partial x_{j_1} \cdots \partial x_{j_p}}(x^{i+1}) - \frac{\partial^p \pi_{\hat{c}}^{i+1}}{\partial x_{j_1} \cdots \partial x_{j_p}}(x^{i+1}) \right|$$

$$\leq \left\| \frac{\partial}{\partial \tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \tilde{\xi}} [\tilde{\pi}^{i+1}(\tilde{\xi})]_{\tilde{\xi}=0} \right\| \left\| \tilde{V}\hat{V}^T \otimes \cdots \otimes \tilde{V}\hat{V}^T - I \right\|$$

$$+ \left\| \frac{\partial}{\partial \tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \tilde{\xi}} [\tilde{\pi}^{i+1}(\tilde{\xi})]_{\tilde{\xi}=0} - \frac{\partial}{\partial \hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \hat{\xi}} [\hat{\pi}^{i+1}(\hat{\xi})]_{\hat{\xi}=0} \right\| \tag{7.8}$$

The following two inequalities will imply that the first term on the right hand side of (7.8) has an upper bound that is proportional to $M\|x^{i+1} - x^i\|^{d+2-k}$ for any $k \geq 2$. Since $\tilde{\pi}^{i+1}$ is computed from the set of coefficients that are exact at the previous patch point, it follows from Theorem 6.1 that there is some $T < \infty$ such that

$$\left\| \frac{\partial}{\partial \tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial \tilde{\xi}} [\tilde{\pi}^{i+1}(\tilde{\xi})]_{\tilde{\xi}=0} \right\| \leq \sup_{X_c \setminus \mathcal{A}} \left\| \frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x} [\pi(x)] \right\| + T\|x^{i+1} - x^i\|^{d+2-k}. \tag{7.9}$$

By Corollary 7.4 and Lemma 6.9,

$$\left\| \tilde{V}\hat{V}^T \otimes \cdots \otimes \tilde{V}\hat{V}^T - I \right\| \leq (2^k - 1)L_1' M \|x^{i+1} - x^i\|^{d+1}. \tag{7.10}$$

39

We now prove that the second term on the right hand side of (7.8) has an upper bound that is proportional to $M\|x^{i+1} - x^i\|^{d+1}$. Each entry in

$$\frac{\partial}{\partial\tilde{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\tilde{\xi}}[\tilde{\pi}^{i+1}(\tilde{\xi})]_{\tilde{\xi}=0} - \frac{\partial}{\partial\hat{\xi}} \otimes \cdots \otimes \frac{\partial}{\partial\hat{\xi}}[\hat{\pi}^{i+1}(\hat{\xi})]_{\hat{\xi}=0}$$

is either the difference between characteristic or non characteristic partial derivatives. In the case of the non characteristic partial derivatives, Lemma 7.5 guarantees that the difference is at most

$$\left|\frac{\partial^k\tilde{\pi}^{i+1}}{\partial\tilde{\xi}_{j_1}\cdots\partial\tilde{\xi}_{j_k}}(0) - \frac{\partial^k\hat{\pi}^{i+1}}{\partial\hat{\xi}_{j_1}\cdots\partial\hat{\xi}_{j_k}}(0)\right| \leq L'_{\text{inh}}M\|x^{i+1} - x^i\|^{d+2-k}$$

so the non characteristic partials make a contribution to the right hand side of (7.8) that is at most $L'_{\text{inh}}M\|x^{i+1} - x^i\|^{d+2-k}$. We now consider the contribution to (7.8) from the characteristic partial derivatives. We start with the second order partial derivatives and proceed by induction. At x^{i+1}, a characteristic partial derivative of the optimal cost is calculated from the coefficients held in C^i from the formula

$$\frac{\partial^2\tilde{\pi}}{\partial\xi_j\partial\xi_1}(0) = -\frac{1}{\|\dot{x}_c^{i+1}\|}\left[\frac{\partial\tilde{\pi}^{i+1}}{\partial\tilde{\xi}}(0)\left(\frac{\partial\tilde{f}}{\partial\tilde{\xi}_j}(0) + \frac{\partial\tilde{g}}{\partial\tilde{\xi}_j}(0)\tilde{\kappa}^{i+1}(0)\right.\right.$$
$$\left.\left. + \frac{\partial\tilde{q}}{\partial\tilde{\xi}_j}(0) + \frac{1}{2}\frac{\partial\tilde{r}}{\partial\tilde{\xi}_j}(0)\tilde{\kappa}^{i+1}(0)^2\right)\right] \quad (7.11)$$

and the corresponding characteristic partial derivative computed from the coefficients in $\hat{C}^i$ are computed by an analogous formula. The task of showing that the difference between the second order computed characteristic partial derivatives makes no more than an $\mathcal{O}(M\|x^{i+1} - x^i\|^d)$ contribution to the right hand side of (7.8) reduces to showing that the difference between the partial derivatives of the problem data functions with respect to the two changes of variables

$$\begin{array}{cc}
\dfrac{\partial\tilde{f}}{\partial\tilde{\xi}}(0) - \dfrac{\partial\hat{f}}{\partial\hat{\xi}}(0) & \dfrac{\partial\tilde{g}}{\partial\tilde{\xi}}(0) - \dfrac{\partial\hat{g}}{\partial\hat{\xi}}(0) \\[3mm]
\dfrac{\partial\tilde{q}}{\partial\tilde{\xi}}(0) - \dfrac{\partial\hat{q}}{\partial\hat{\xi}}(0) & \dfrac{\partial\tilde{r}}{\partial\tilde{\xi}}(0) - \dfrac{\partial\hat{r}}{\partial\hat{\xi}}(0)
\end{array} \quad (7.12)$$

and the difference between each of the computed quantities

$$\frac{\partial\tilde{\pi}^{i+1}}{\partial\tilde{\xi}}(0) - \frac{\partial\hat{\pi}^{i+1}}{\partial\hat{\xi}}(0) \qquad \tilde{\kappa}^{i+1}(0) - \hat{\kappa}^{i+1}(0) \qquad \frac{1}{\|\dot{x}_c^{i+1}\|} - \frac{1}{\|\dot{x}_{\hat{c}}^{i+1}\|} \quad (7.13)$$

are all proportional to $M\|x^{i+1} - x^i\|^d$ or higher. In fact, they are all proportional to $M\|x^{i+1} - x^i\|^{d+1}$. Lemma 6.10 bounds the difference between the partial derivatives of the problem data functions in terms of $\hat{V}^T\tilde{V} - I$. Corollary 7.4 then implies that for some Lipschitz constant $L < \infty$

$$\left\|\frac{\partial\tilde{g}}{\partial\tilde{\xi}}(0) - \frac{\partial\hat{g}}{\partial\hat{\xi}}(0)\right\| \leq LM\|x^{i+1} - x^i\|^{d+1}$$

and and the same bound bounds the other differences in (7.12).

We now show that there is some Lipschitz constant $L > 0$ such that each difference in (7.13) is bounded by $LM\|x^{i+1} - x^i\|^{d+1}$. By Theorem 5.1, there exists an L such that

$$\left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1}) - \frac{\partial \pi_{\hat{c}}^{i+1}}{\partial x}(x^{i+1})\right\| \leq LM\|x^{i+1} - x^i\|^{d+1}$$

and by the inequalities (7.9) and (7.10), there is some other L

$$\left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1})(\tilde{V}\hat{V}^T - I)\right\| \leq LM\|x^{i+1} - x^i\|^{d+1}.$$

It then follows from

$$\left\|\frac{\partial \tilde{\pi}^{i+1}}{\partial \tilde{\xi}}(0) - \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0)\right\| = \left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1})\tilde{V} - \frac{\partial \pi_{\hat{c}}^{i+1}}{\partial x}(x^{i+1})\hat{V}\right\|$$

$$\leq \left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1}) - \frac{\partial \pi_{\hat{c}}^{i+1}}{\partial x}(x^{i+1})\right\|$$

$$+ \left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1})(\tilde{V}\hat{V}^T - I)\right\|$$

that for some $L < \infty$,

$$\left\|\frac{\partial \tilde{\pi}^{i+1}}{\partial \tilde{\xi}}(0) - \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0)\right\| \leq LM\|x^{i+1} - x^i\|^{d+1}. \tag{7.14}$$

The differences between the computed optimal controls is at most

$$|\tilde{\kappa}^{i+1}(0) - \hat{\kappa}^{i+1}(0)| = \frac{1}{r(x^{i+1})}\left|\left(\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1}) - \frac{\partial \pi_{\hat{c}}^{i+1}}{\partial x}(x^{i+1})\right)g(x^{i+1})\right|$$

$$\leq \sup_{X_c \setminus \mathcal{A}} \frac{\|g(x)\|}{r(x)}\left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1}) - \frac{\partial \pi_{\hat{c}}^{i+1}}{\partial x}(x^{i+1})\right\| \tag{7.15}$$

$$\leq \sup_{X_c \setminus \mathcal{A}} \frac{\|g(x)\|}{r(x)}LM\|x^{i+1} - x^i\|^{d+1}$$

Lemma 7.2 guarantees that for all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$,

$$\inf_{\|x^{i+1}-x^i\|\leq h_2}\|\dot{x}_c^{i+1}\|(\|\dot{x}_c^{i+1}\| - \|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\|) > 0 \tag{7.16}$$

whenever

$$\|x^{i+1} - x^i\| \leq h_2 \text{ and } \|C_k^i - \hat{C}_k^i\| \leq M\|x^{i+1} - x^i\|^{d+2-k} \text{ for } 0 \leq k \leq d+1$$

Therefore, for some L, the final difference in (7.13) is bounded by

$$\left|\frac{1}{\|\dot{x}_c^{i+1}\|} - \frac{1}{\|\dot{x}_{\hat{c}}^{i+1}\|}\right|$$

$$\leq \frac{\sup\|g(x)\|^2/r(x)}{\inf\left(\|\dot{x}_c^{i+1}\|(\|\dot{x}_c^{i+1}\| - \|\dot{x}_c^{i+1} - \dot{x}_{\hat{c}}^{i+1}\|)\right)}\left\|\frac{\partial \pi_c^{i+1}}{\partial x}(x^{i+1}) - \frac{\partial \pi_{\hat{c}}^{i+1}}{\partial x}(x^{i+1})\right\| \tag{7.17}$$

$$\leq LM\|x^{i+1} - x^i\|^{d+1}$$

where the supremum is taken over $\overline{X_c \setminus \mathcal{A}}$, and the infimum is taken over all x^i and x^{i+1} in $\overline{X_c \setminus \mathcal{A}}$ within a distance of h_2 of each other. Their ratio is then absorbed into L.

The proof that the difference between every corresponding characteristic partial derivatives of order $k > 2$ shares a single upper bound that is proportional to $M\|x^{i+1} - x^i\|^{d+2-k}$ proceeds by induction and follows the same form as the $k = 2$ case.
□

8. Numerical results. We tested the patchy algorithm on the following non-linear test problem.

Find

$$\pi(x^0) = \min_u \frac{1}{2} \int_0^\infty \left(\sin^2(x_1) + \left(x_2 - \frac{1}{3}x_1^3 \right)^2 + u^2 \right) dt$$

subject to the dynamics

$$\dot{x}_1 = \left(x_2 - \frac{1}{3}x_1^3 \right) \sec(x_1)$$
$$\dot{x}_2 = \left(x_1^2 x_2 - \frac{1}{3}x_1^5 \right) \sec(x_1) + u$$

For any integer k, the optimal dynamics have a rest point at $(x_1, x_2) = (k\pi, \frac{1}{3}(k\pi)^3)$, and the optimal cost is zero, and the presence of $\sec(x_1)$ makes $\dot{x}$ blow up near $x_1 = -\pi/2$ and $x_1 = \pi/2$.

Under a change of coordinates, the test problem is the linear-quadratic regulator problem

Find

$$\pi(y^0) = \min_u \frac{1}{2} \int_0^\infty (y^T y + u^2) dt$$

subject to the dynamics

$$\dot{y}_1 = y_2$$
$$\dot{y}_2 = u$$

and the change of variables is

$$y_1 = \sin(x_1)$$
$$y_2 = x_2 - \frac{1}{3}x_1^3. \tag{8.1}$$

We chose this nonlinear test problem because we can calculate the exact solution to the linear-quadratic regulator problem exactly by solving the associated algebraic Riccati matrix equation [7, p. 252]. We can then calculate the exact solution to the nonlinear test problem by applying the change of coordinates to the solution of the

linear-quadratic regulator problem. Doing so gives us the solution to the nonlinear optimal control problem

$$\pi(x) = \frac{\sqrt{3}}{2}\left[\left(\sin(x_1) + \left(x_2 - \frac{1}{3}x_1^3\right)\right)^2 + \frac{2}{3}\left(x_2 - \frac{1}{3}x_1^3\right)^2\right]$$

The problem data functions f, g, q, and r and the solution π are all smooth on $\{(x_1, x_2) \,|\, -\pi/2 < x_1 < \pi/2, -\infty < x_2 < \infty\}$. Furthermore, π is a strict Lyapunov function on this set. If we denote the change of variables (8.1) $y = T(x)$, and if the matrix $\frac{\partial T}{\partial x}(x)$ is invertible at x, then $\frac{\partial \pi}{\partial x}(x)\dot{x}$ reduces to the following quadratic form in $T(x)$

$$\frac{\partial \pi}{\partial x}(x)\dot{x} = T(x)^T \begin{bmatrix} -1 & 0 \\ -\sqrt{3} & -2 \end{bmatrix} T(x),$$

which is negative definite. Therefore, the computed solutions π^i obey the error bound of Theorem 3.6.

We computed the patchy solution on 73 patches including the Al'brekht patch. The maximum consecutive patch point distance was $h \approx .54$. The degree of the computed optimal cost and optimal control polynomials was four and three. To compute the absolute error, we created a 100×100 grid of equally spaced points in the square $-1 \leq x_1, x_2 \leq 1$, and then computed the difference between the exact and computed optimal cost at each grid point that fell inside a patch. The maximum absolute error at a grid point was approximately 5×10^{-3}. The theoretical absolute error bound grows along a sequence of consecutive patch points as $K(\frac{L^i-1}{L-1} + 1)h^5 \approx K(\frac{L^i-1}{L-1} + 1)(4.6 \times 10^{-2})$. The computed and exact solutions are displayed in Figure 8.1. The absolute error is displayed in Figure 8.2.

To verify the validity of the analytical error bound, we computed the patchy solution on five concentric level sets of the computed cost. We doubled the number of patch points on each level set, starting with eight points on the boundary of the Al'brekht patch. We then computed the absolute error of optimal cost at each patch point and then found the sequence of patch points that terminates with the greatest absolute error. The level sets of the optimal cost behave badly near the singularities at $x_1 = \pm\pi/2$, which limited the size of the computational domain for this test problem. In this case, $h \approx .5$, so the absolute error bound grows as $K(\frac{L^i-1}{L-1} + 1)h^5 \approx K(\frac{L^i-1}{L-1} + 1)(3.1 \times 10^{-2})$. The absolute error and the sequence of patch points yielding the worst case error are shown in Figure 8.3.
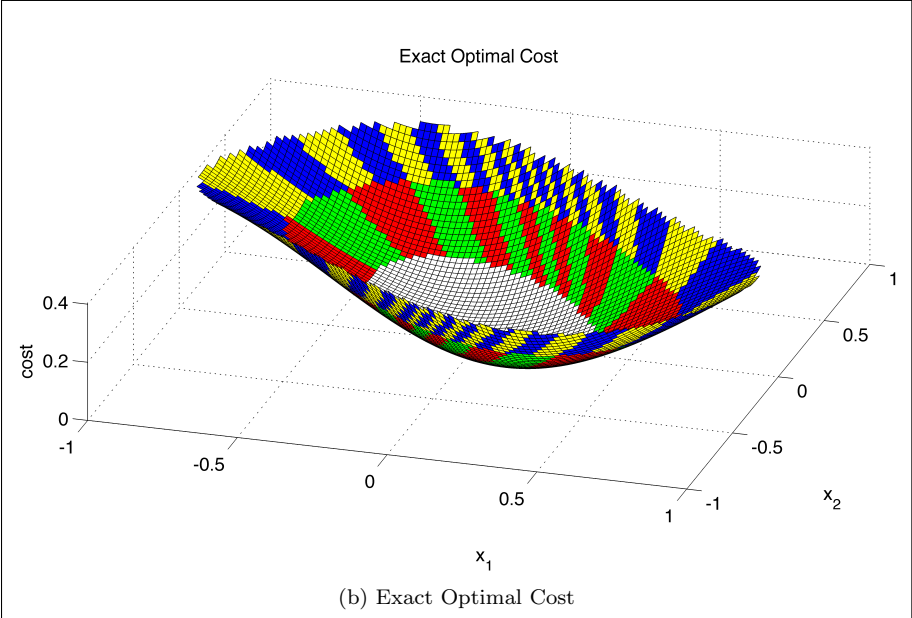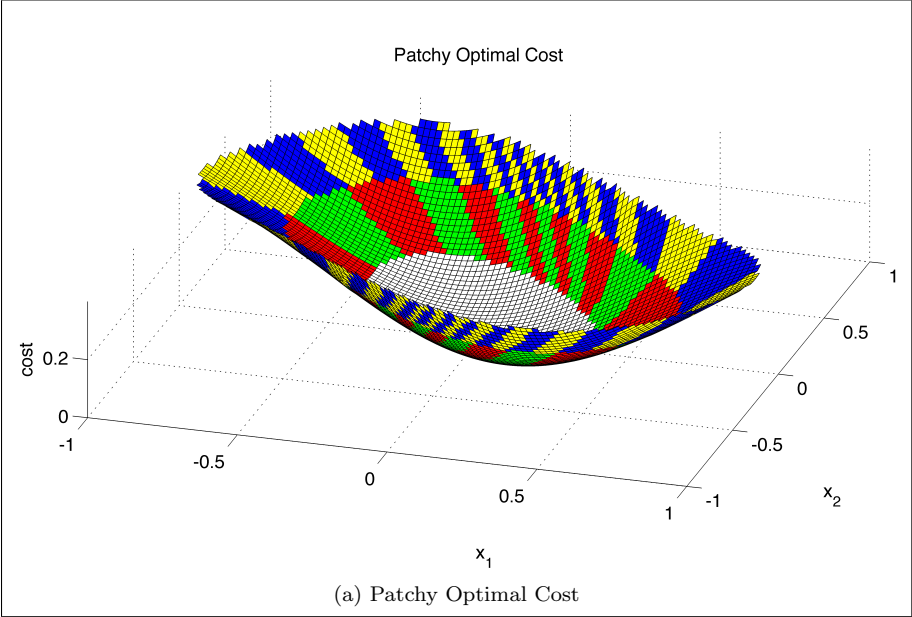
(a) Patchy Optimal Cost

(b) Exact Optimal Cost

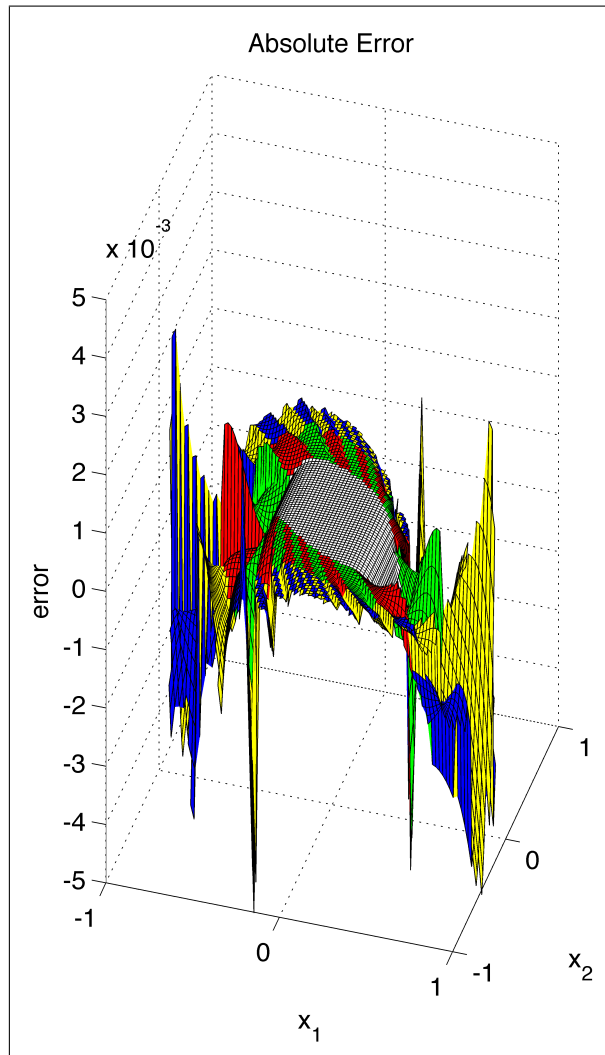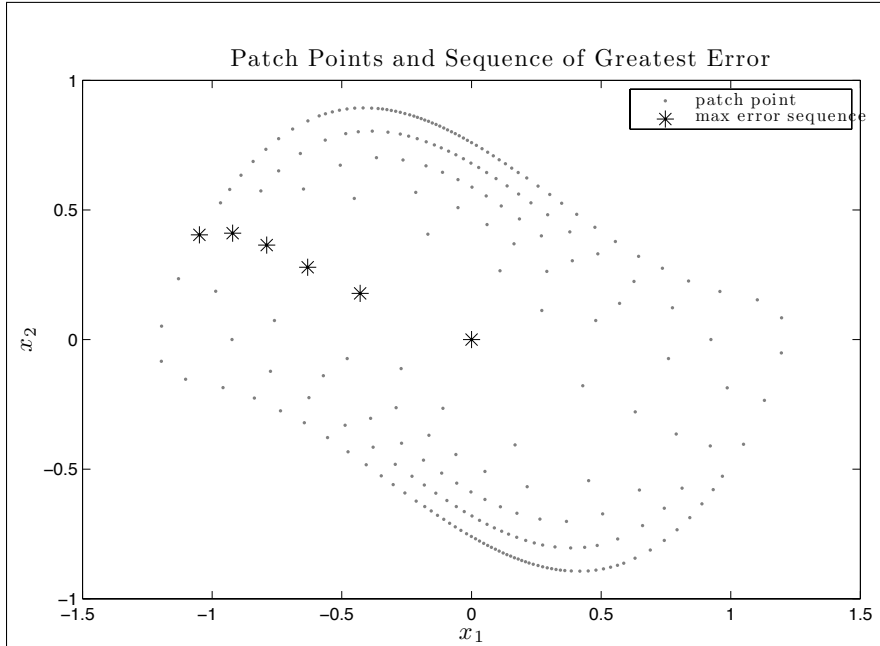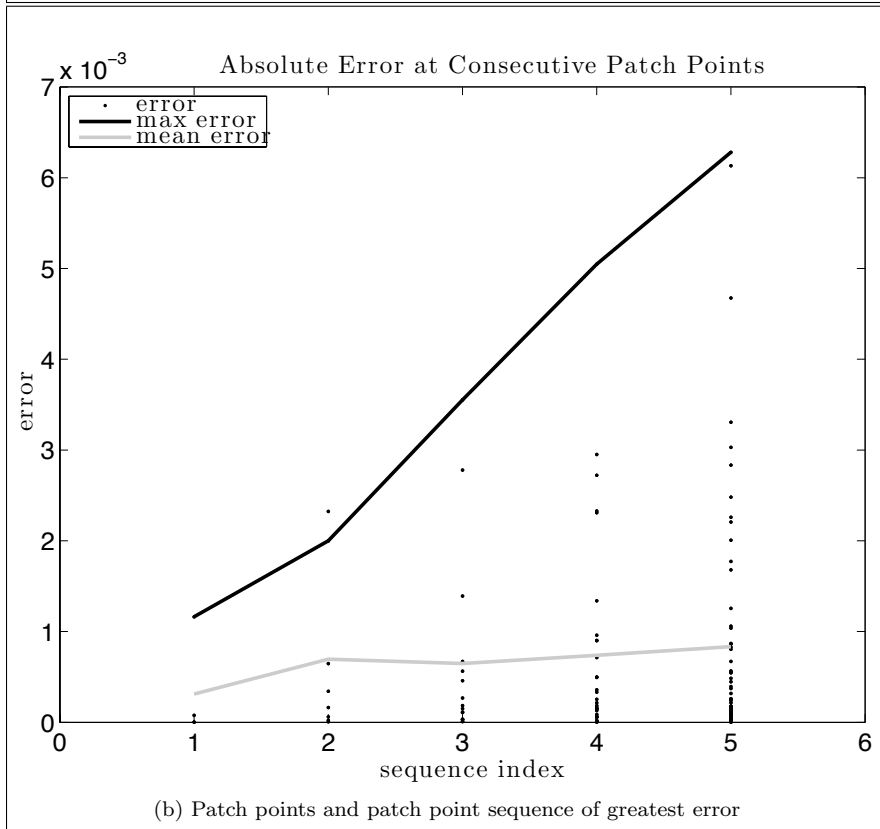Fig. 8.1. *Patchy and exact optimal costs*

FIG. 8.2. *Absolute Error* $\pi(x) - \pi^i(x)$

(a) Optimal cost absolute error at sequence of consecutive patch points.

(b) Patch points and patch point sequence of greatest error

FIG. 8.3. *Optimal cost absolute error at sequences of patch points.* $h \approx .5$

9. Appendix.

9.1. Kronecker derivative notation. DEFINITION 9.1. *Let* $s : \mathbb{R}^n \to \mathbb{R}^{1 \times m}$

$$s(x) = \begin{bmatrix} s_1(x) & s_2(x) & \cdots & s_m(x) \end{bmatrix},$$

The differential operator is $\frac{\partial}{\partial x} \otimes s(x)$ *is shorthand for the* $1 \times mn$ *row vector*

$$\frac{\partial}{\partial x} \otimes s(x) \equiv \begin{bmatrix} \frac{\partial s}{\partial x_1}(x) & \frac{\partial s}{\partial x_2}(x) & \cdots & \frac{\partial s}{\partial x_n}(x) \end{bmatrix}$$

The higher order operator $\frac{\partial}{\partial x} \otimes \cdots \otimes \frac{\partial}{\partial x}$ *is defined recursively in the natural way. In the special case where* $s : \mathbb{R}^2 \to \mathbb{R}$, *then* $\frac{\partial s}{\partial x} : \mathbb{R}^2 \to \mathbb{R}^{1 \times 2}$, *and*

$$\frac{\partial}{\partial x}[s(x)] = \begin{bmatrix} \frac{\partial s}{\partial x_1}(x) & \frac{\partial s}{\partial x_2}(x) \end{bmatrix}$$

$$\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x}[s(x)] \equiv \begin{bmatrix} \frac{\partial^2 s}{\partial x_1^2}(x) & \frac{\partial^2 s}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 s}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 s}{\partial x_2^2}(x) \end{bmatrix}$$

$$\frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x} \otimes \frac{\partial}{\partial x}[s(x)] \equiv \begin{bmatrix} \frac{\partial^3 s}{\partial x_1^3}(x) & \frac{\partial^3 s}{\partial x_1^2 \partial x_2}(x) & \frac{\partial^3 s}{\partial x_1^2 \partial x_2}(x) & \frac{\partial^3 s}{\partial x_1 \partial x_2^2}(x) & \cdots \\ \frac{\partial^3 s}{\partial x_1^2 \partial x_2}(x) & \frac{\partial^3 s}{\partial x_1 \partial x_2^2}(x) & \frac{\partial^3 s}{\partial x_1 \partial x_2^2}(x) & \frac{\partial^3 s}{\partial x_2 \partial x_1}(x) & \cdots \\ \frac{\partial^3 s}{\partial x_2^3}(x) & & & & \end{bmatrix}$$

$$\vdots$$

9.2. Partial derivative formulas for the optimal control and characteristic partial derivatives of the optimal cost.

$$
\frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_1}(0) = -\frac{1}{\|\dot{x}^{i+1}\|} \left[\frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}}(0) \left(\frac{\partial \hat{f}}{\partial \hat{\xi}_{j_2}}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_2}}(0) \hat{\kappa}^{i+1}(0) \right) \right.
$$

$$
+ \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0) \left(\frac{\partial \hat{f}}{\partial \hat{\xi}_{j_1}}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_1}}(0) \hat{\kappa}^{i+1}(0) \right)
$$

$$
+ \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0) \left(\frac{\partial^2 \hat{f}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) + \frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) \hat{\kappa}^{i+1}(0) \right)
$$

$$
+ \frac{\partial^2 \hat{q}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) + \frac{1}{2} \frac{\partial^2 \hat{r}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) \hat{\kappa}^{i+1}(0)^2
$$

$$
\left. - \frac{\partial \hat{\kappa}}{\partial \hat{\xi}_{j_1}}(0) \frac{\partial \hat{\kappa}}{\partial \hat{\xi}_{j_2}}(0) \right]
$$

$$\frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) = -\frac{1}{\hat{r}(0)} \left[\frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0)\hat{g}(0) + \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}}(0)\frac{\partial \hat{g}}{\partial \hat{\xi}_{j_2}}(0) \right.$$

$$+ \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0)\frac{\partial \hat{g}}{\partial \hat{\xi}_{j_1}}(0)$$

$$+ \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0)\frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) + \frac{\partial^2 r}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0)\hat{\kappa}^{i+1}(0)$$

$$+ \frac{\partial \hat{r}}{\partial \hat{\xi}_{j_1}}(0)\hat{\kappa}^{i+1}(0)$$

$$\left. + \frac{\partial \hat{r}}{\partial \hat{\xi}_{j_1}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2}}(0) + \frac{\partial \hat{r}}{\partial \hat{\xi}_{j_2}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1}}(0) \right]$$

$$\frac{\partial^4 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3} \partial \hat{\xi}_1}(0) =$$

$$-\frac{1}{\|\dot{x}^{i+1}\|} \left[\frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0)\left(\frac{\partial \hat{f}}{\partial \hat{\xi}_{j_3}}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_3}}(0)\hat{\kappa}^{i+1}(0) \right) \right.$$

$$+ \frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0)\left(\frac{\partial \hat{f}}{\partial \hat{\xi}_{j_3}}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_3}}(0)\hat{\kappa}^{i+1}(0) + \hat{g}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2}}(0) \right)$$

$$+ \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}}(0)\left(\frac{\partial^2 \hat{f}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) + \frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0)\hat{\kappa}^{i+1}(0) + \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_3}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2}}(0) \right.$$

$$\left. + \hat{g}(0)\frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) \right)$$

$$+ \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_3} \partial \hat{\xi}}(0)\left(\frac{\partial^2 \hat{f}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0) + \frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0)\hat{\kappa}^{i+1}(0) \right)$$

$$+ \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0)\left(\frac{\partial^2 \hat{f}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_3}}(0) + \frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_3}}(0)\hat{\kappa}^{i+1}(0) \right)$$

$$+ \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0)\left(\frac{\partial^3 \hat{f}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) + \frac{\partial^3 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_2}}(0)\hat{\kappa}^{i+1}(0) \right)$$

$$+ \frac{\partial^3 \hat{q}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) + \frac{1}{2}\frac{\partial^3 \hat{r}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0)\left(\hat{\kappa}^{i+1}(0) \right)^2$$

$$+ \frac{\partial^2 \hat{r}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0)\hat{\kappa}^{i+1}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_3}}(0) + \frac{\partial \hat{r}}{\partial \hat{\xi}_{j_3}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2}}(0)$$

$$- \hat{r}(0)\frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_3}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2}}(0) - \hat{r}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1}}(0)\frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0)$$

$$\left. - \left(\frac{\partial \hat{r}}{\partial \hat{\xi}_{j_2}}(0)\frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1}}(0) + \hat{r}(0)\frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0)\frac{\partial \kappa^{i+1}}{\partial \hat{\xi}_{j_3}}(0) \right) \right]$$

$$\frac{\partial^3 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) =$$

$$-\frac{1}{\hat{r}(0)} \left[\frac{\partial^4 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3} \partial \hat{\xi}_{j_4}}(0) \hat{g}(0) \right.$$

$$+ \frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}}(0) \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_3}}(0) + \frac{\partial^3 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_3} \partial \hat{\xi}}(0) \frac{\partial \hat{g}}{\partial \hat{\xi}_{j_2}}(0)$$

$$+ \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}}(0) \frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) + \frac{\partial^2 \hat{\pi}^{i+1}}{\partial \hat{\xi}_{j_3} \partial \hat{\xi}}(0) \frac{\partial^2 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0)$$

$$+ \frac{\partial \hat{\pi}^{i+1}}{\partial \hat{\xi}}(0) \frac{\partial^3 \hat{g}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0)$$

$$+ \frac{\partial^2 \hat{r}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) \frac{\partial \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1}}(0) + \frac{\partial \hat{r}}{\partial \hat{\xi}_{j_2}}(0) \frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) + \frac{\partial \hat{r}}{\partial \hat{\xi}_{j_3}}(0) \frac{\partial^2 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2}}(0)$$

$$\left. + \hat{r}(0) \frac{\partial^3 \hat{\kappa}^{i+1}}{\partial \hat{\xi}_{j_1} \partial \hat{\xi}_{j_2} \partial \hat{\xi}_{j_3}}(0) \right]$$

REFERENCES

[1] E. G. Al'brekht, *On the optimal stabilization of nonlinear systems*, Journal of Applied Mathematics and Mechanics, 25 (1961), pp. 836 – 844.

[2] Fabio Ancona and Alberto Bressan, *Nearly time optimal stabilizing patchy feedbacks*, Annales de l'Institut Henri Poincare (C) Non Linear Analysis, 24 (2007), pp. 279 – 310.

[3] Gene H. Golub and Charles F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 1996.

[4] Fritz John, *Partial Differential Equations*, Springer-Verlag, 4th ed., 1982.

[5] Arthur J. Krener, *Nonlinear systems toolbox v. 1.0.* Available by request from ajkrener@nps.edu.

[6] D. L. Lukes, *Optimal regulation of nonlinear dynamical systems*, SIAM Journal on Control, 7 (1969), pp. 75–100.

[7] Carmeliza Navasca and Arthur J. Krener, *Modeling, Estimation and Control: Festschrift in Honor of Giorgio Picci on the Occasion of his Sixty-Fifth Birthday*, Springer-Verlag, 2007.

[8] Walter Rudin, *Principles of Mathematical Analysis*, McGraw-Hill, Inc., third ed., 1976.

[9] J.A. Sethian, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, 1999.

[10] G.W. Stewart, *Matrix Algorithms Volume 1: Basic Decompositions*, SIAM, 1998.

[11] Josef Stoer and Roland Bulirsch, *Introduction to Numerical Analysis*, Springer, second ed., 1992.

[12] Adam Szpiro and Paul Dupuis, *Second order numerical methods for first order hamilton-jacobi equations*, SIAM Journal on Numerical Analysis, 40 (2003), pp. pp. 1136–1183.

[13] J.N. Tsitsiklis, *Efficient algorithms for globally optimal trajectories*, Automatic Control, IEEE Transactions on, 40 (1995), pp. 1528 –1538.