A Descent Algorithm for the Optimal Control of Constrained Nonlinear Switched Dynamical Systems: Appendix

Humberto Gonzalez
Ram Vasudevan
Maryam Kamgarpour
S. Shankar Sastry
Ruzena Bajcsy
Claire Tomlin

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2010-9
http://www.eecs.berkeley.edu/Pubs/TechRpts/2010/EECS-2010-9.html

January 26, 2010

A Descent Algorithm for the Optimal Control of Constrained Nonlinear Switched Dynamical Systems: Appendix

Humberto Gonzalez, Ram Vasudevan, Maryam Kamgarpour,
Shankar Sastry, Ruzena Bajcsy, and Claire Tomlin*

Abstract

This document is an appendix to the paper "A Descent Algorithm for the Optimal Control of Constrained Nonlinear Switched Dynamical Systems"[3], presented at the 13th International Conference on Hybrid Systems: Computation and Control (HSCC'10), and contains the proofs of all the propositions and theorems in Section 4 of that paper. The reader is advised to read this document after covering the first three sections of the original paper, as the notation and the problem formulation are not covered in this appendix. The numbering of the equations, propositions, and theorems is consistent with the numbering used in the original paper.

A Algorithm Analysis

A.1 Continuity of the Cost and Constraints

We must first check that the cost function, equation (12), under Assumptions 1 and 2 is continuous. We prove the continuity of the cost function by taking a sequence, $(\xi_j)_{j=1}^{\infty}$ converging to limit ξ, in our optimization space, and proving that the corresponding sequence of trajectories $(x_j(t))_{j=1}^{\infty}$ converge to trajectory $x(t)$ corresponding to ξ. This result proves the sequential continuity of our cost function, which implies continuity since $\mathcal{X}$ is a metric space.

Throughout this subsection we simplify the notation used for the functions μ, κ, π, and μ_f. Given $\xi_j = (\sigma_j, s_j, u_j) \in \mathcal{X}$, we define $\mu_j(i) = \mu(i; s_j)$, $\kappa_j(i) = \kappa(i; s_j)$, $\pi_j(i) = \pi(i; s_j)$, and $\mu_{f,j} = \mu_f(s_j)$. As usual, when the choice of $s \in \mathcal{S}$ is clear in context we use our standard notation.

Proposition 1. *Let $(\xi_j := (\sigma_j, s_j, u_j))_{j=1}^{\infty}$ be a convergent sequence in our optimization space, $\mathcal{X}$, and let $\xi := (\sigma, s, u)$ be its limit. Let $(x_j(t))_{j=1}^{\infty}$ be the corresponding trajectories (defined using Equation 11) associated with each ξ_j, with common initial condition x_0. The sequence $(x_j(t))_{j=1}^{\infty}$ converges pointwise to the trajectory $x(t)$ associated with ξ, for all t in $[0, \infty)$ with initial condition x_0.*

Proof. The proof is completed in steps, and employs an induction argument which is applicable since we assume that the number of switches between non-zero flows is finite. First fix an ϵ in $(0, \min\{1, \min_{s(i)>0}\{s(i)/2\}\})$. Since the sequence $(\xi_j)_{j=1}^{\infty}$ converges to ξ, we know there exists some j_0, such that for all j greater than j_0, $\|\xi_j - \xi\| < \epsilon$ (where we are using the norm induced

*All the authors are with the Department of Electrical Engineering and Computer Sciences, University of California, Berkeley. Emails: {hgonzale,ramv,maryamka,sastry,bajcsy,tomlin}@eecs.berkeley.edu

by our metric). Given our choice of ϵ, this means that for all j greater than j_0, $\sigma_j = \sigma$ and $|\mu_j(i) - \mu(i)| < \epsilon$ for all i. Without loss of generality we assume that $j_0 = 1$. We employ these two observations to define a partition, τ, of $[0, \infty)$ induced by the points, $\{\mu_j(i), \mu(i)\}_{i=0}^{\infty}$ for each j. This partition divides $[0, \infty)$ into two types of intervals:

- Intervals where the modes agree and are constant, but the initial conditions at the start of the interval differ by a value bounded by ϵ times a constant.
- Intervals where the initial conditions at the start of the interval differ by a value bounded by ϵ times a constant, the modes disagree but are constant, and the size of the interval is bounded by ϵ times a constant.

In order to apply the induction step, we need to find a method to bound the final condition of the trajectory for each of these intervals.

Step 1: Let $\tau_k := [a, b] \in \tau$. If $\pi_j(t) = \pi(t)$, $\forall t \in \tau_k$, and $\|x_j(a) - x(a)\| < C\epsilon$ for some $C > 0$, then $\exists \bar{C} > 0$ such that $\|x_j(t) - x(t)\| < \bar{C}\epsilon$, $\forall t \in \tau_k$.

Proof of Step 1: We begin by considering the norm difference of trajectories for $t \in [a, b]$ under our assumptions:

$$\|x_j(t) - x(t)\| \leq \|x_j(a) - x(a)\| + \int_a^t \|f_{\pi_j(r)}(x_j(r), u_j(r)) - f_{\pi(r)}(x(r), u(r))\| dr \tag{A.1}$$

Next, we employ the Lipschitz continuity assumption (our Lipschitz constant is called K_1 here) as prescribed by Assumption 1, since $\pi_j(t) = \pi(t)$, $\forall t \in \tau_k$:

$$\|x_j(t) - x(t)\| \leq \|x_j(a) - x(a)\| + K_1 \int_a^t \Big(\|x_j(r) - x(r)\| + \|u_j(r) - u(r)\| \Big) dr \tag{A.2}$$

Next, we employ the Bellman-Gronwall Inequality to bound the integral term in the previous inequality.

$$\|x_j(t) - x(t)\| \leq e^{K_1(b-a)} \left[\|x_j(a) - x(a)\| + K_1 \int_a^b \|u_j(r) - u(r)\| dr \right] \tag{A.3}$$

We can employ the assumptions given in the our proposition statement to bound the initial condition. Moreover, we can also bound the integral term by remembering that we are working under the assumption that $\|\xi_j - \xi\| < \epsilon$:

$$\|x_j(t) - x(t)\| \leq e^{K_1(b-a)} \Big[C + K_1(b - a) \Big] \epsilon \tag{A.4}$$

Letting $\bar{C} := e^{K_1(b-a)} \Big[C + K_1(b - a) \Big]$, we have the desired result. $\square$(*Step 1*)

Step 2: Let $\tau_k := [a, b] \in \tau$. If $\pi_j(t), \pi(t)$ are constant $\forall t \in \tau_k$, $|b - a| < C_1\epsilon$ for some $C_1 > 0$, and $\|x_j(a) - x(a)\| < C_2\epsilon$ for some $C_2 > 0$, then $\exists \bar{C} > 0$ such that $\|x_j(t) - x(t)\| < \bar{C}\epsilon$, $\forall t \in [a, b]$.

Proof of Step 2: We begin by considering the norm difference of trajectories for $t \in [a, b]$ under our assumptions, but also add and subtract a term to be able to immediately apply a Lipschitz argument in the next step:

$$\|x_j(t) - x(t)\| = \Big\| x_j(a) - x(a) +$$
$$+ \int_a^t \Big(f_{\pi_j(r)}(x_j(r), u_j(r)) - f_{\pi_j(r)}(x(r), u(r)) + f_{\pi_j(r)}(x(r), u(r)) - f_{\pi(r)}(x(r), u(r)) \Big) dr \Big\| \quad \text{(A.5)}$$

Next, we employ the Lipschitz continuity assumption (our Lipschitz constant is called K_1 here) as prescribed by Assumption 1 and the triangle inequality simultaneously:

$$\|x_j(t) - x(t)\| \leq \|x_j(a) - x(a)\| + K_1 \int_a^t \Big(\|x_j(r) - x(r)\| + \|u_j(r) - u(r)\| \Big) dr +$$
$$+ \int_a^t \Big\| f_{\pi_j(r)}(x(r), u(r)) - f_{\pi(r)}(x(r), u(r)) \Big\| \, dr \quad \text{(A.6)}$$

At this point, we want to employ an affine Lipschitz continuity argument. First, recall that we assume $\|u(t)\| \leq M, \forall t$. Second, recall that since $x(t)$ is the solution to a differential equation, it is continuous. Therefore over a compact domain, $[0, \mu_f]$, we know $\exists \bar{M} < \infty$ such that $\max_{t \in [0, \mu_f]} \|x(t)\| \leq \bar{M}$. Thus, if we define $K_3 = K_1(\bar{M} + M) + \|f_{\pi(t)}(0, 0)\|$, which is bounded by assumption, then we have:

$$\|f_{\pi(t)}(x(t), u(t))\| \leq K_3, \qquad \forall t \in \tau_k \quad \text{(A.7)}$$

An identical result holds for $f_{\pi_j(r)}(x(r), u(r))$, thus continuing with equation (A.6), we have:

$$\|x_j(t) - x(t)\| \leq \|x_j(a) - x(a)\| + 2K_3(b - a) + K_1 \int_a^t \Big(\|u_j(r) - u(r)\| + \|x_j(r) - x(r)\| \Big) dr \quad \text{(A.8)}$$

Next, we employ the Bellman-Gronwall Inequality to bound the integral term in the previous inequality.

$$\|x_j(t) - x(t)\| \leq e^{K_1(b-a)} \left[\|x_j(a) - x(a)\| + 2K_3(b - a) + K_1 \int_a^b \|u_j(r) - u(r)\| dr \right] \quad \text{(A.9)}$$

We can employ the assumptions given in the our proposition statement to bound the initial condition. Moreover, we can also bound the integral term by remembering that we are working under the assumption that $\|\xi_j - \xi\| < \epsilon$:

$$\|x_j(t) - x(t)\| \leq e^{K_1 C_1 \epsilon} \Big[C_2 + 2K_3 C_2 + C_2 \epsilon \Big] \epsilon \quad \text{(A.10)}$$

Letting $\bar{C} := e^{K_1 C_1 \epsilon} \Big[C_2 + 2K_2 C_2 + C_2 \epsilon \Big]$, we have the desired result. $\square$(*Step 2*)

Now we must prove that the intervals in our partition where the modes disagree have a bounded size.

Step 3: Let $\tau_k := [a, b] \in \tau$. If $\pi_j(t) \neq \pi(t), \forall t \in \tau_k$, then $|b - a| < 2\epsilon$.

Proof of Step 3: Recall that $|\xi_j(i) - \xi(i)| < \epsilon$, which implies that both $\sigma_j(i) = \sigma(i)$ and $|\mu_j(i) - \mu(i)| < \epsilon, \forall i$. If s only contained nonzero entries until it reached the end of its switching sequence (i.e. $\exists k$ such that $\forall i < k, s(i) > 0$, and $\forall i \geq k, s(i) = 0$), then $\mu_j(i) < \mu(i) + \epsilon \leq \mu(i) + s(i+1)/2 < \mu(i+1)$. This immediately gives our desired result.

For a general s, we do not have the bound $\mu_j(i) \leq \mu(i+1)$ since $s(i+1)/2$ maybe equal to zero. However, we do know that $|\mu_j(m(i)) - \mu(m(i))| < \epsilon$ and $|\mu_j(n(i)) - \mu(n(i))| < \epsilon$, which taken together imply that $|\mu_j(n(i)) - \mu_j(m(i))| < 2\epsilon$, or more clearly the length of any switch in the convergent sequence is bounded by 2ϵ, if $n(i) - m(i) > 0$. This result taken in tandem with the result in the previous case gives our required result. □(*Step 3*)

Given these three steps we can now apply our induction argument as follows. Since we assume a common initial condition, during the first interval in our partition, τ, we can employ step one to bound the norm difference in final condition at the end of the interval by ϵ times a constant, which then serves as the bound on the norm difference in initial condition for the next element in our partition.

For a general element in our partition, we assume that the norm difference in initial condition is bounded by ϵ times a constant. For this element, either the modes agree in which case we can apply step one to bound the norm difference in final condition, or the modes differ but the interval length is bounded by ϵ times a constant by step three, and we can immediately apply step two to bound the norm difference in final condition by ϵ times a constant. This again serves as the bound on the norm difference in initial condition for the next element in our partition.

Since the number of switches is finite by assumption, we can sum these various bounds and arrive at an ϵ times constant bound on the final condition. In fact, we can apply this same argument for any instant in time in our trajectory. Therefore, we have that $(x_j(t))_{j=1}^{\infty}$ converges pointwise to the trajectory $x(t)$ for $t \in [0, \infty)$, as required. □

In fact, we have a stronger condition on the convergence.

Proposition 2. *Let $(\xi_j := (\sigma_j, s_j, u_j))_{j=1}^{\infty}$ be a convergent sequence in our optimization space, $\mathcal{X}$, and let $\xi := (\sigma, s, u)$ be its limit. Let $(x_j(t))_{j=1}^{\infty}$ be the corresponding trajectories (defined using Equation 11) associated with each ξ_j, with common initial condition x_0. The sequence $(x_j)_{j=1}^{\infty}$ converges uniformly to the trajectory x associated with ξ, on $[0, \sum_{i=1}^{\infty} s(i)]$ with initial condition x_0.*

Proof. Fix an $\epsilon > 0$. Observe that $\exists j_0$ such that for all $j > j_0$, $\mu_{f,j} \leq \mu_f + \epsilon$. Without loss of generality assume that $j_0 = 1$. Let $D_\epsilon := [0, \mu_f + \epsilon]$. Consider $\eta_j(t) := \sup_{t \in D_\epsilon} \|x_j(t) - x(t)\|$ since D_ϵ is compact we in fact have $\eta_j = \max_{t \in D_\epsilon} \|x_j(t) - x(t)\|$. Since $\lim_{j \to \infty} \eta_j = 0$, we have that $x_j \to x$ uniformly on $[0, \sum_{i=1}^{\infty} s(i)]$ by Theorem 7.9 in [7]. □

Given Proposition 2, we can now check the continuity of the cost function.

Proposition 3. *The function J as defined in equation* (12) *is continuous.*

Proof. Let $(\xi_j := (\sigma_j, s_j, u_j))_{j=1}^{\infty}$ be a convergent sequence in our optimization space, $\mathcal{X}$, and let $\xi := (\sigma, s, u)$ be its limit. If we show that $\lim_{j \to \infty} J(\xi_j) = J(\xi)$, then since sequential continuity implies continuity in a metric space, we arrive at the required result. Let $(x_j(t))_{j=1}^{\infty}$ be the corresponding trajectories (defined using Equation 11) associated with each ξ_j, with common initial condition x_0, which converges to the trajectory $x(t)$, for each t in $[0, \infty]$ with initial condition x_0. Again we prove the result in steps.

Step 1: $\lim_{j \to \infty} \phi(x_j(\mu_{f,j})) = \phi(x(\mu_f))$

Proof of Step 1: First, we apply the triangle inequality:

$$\lim_{j \to \infty} \|\phi(x_j(\mu_{f,j})) - \phi(x(\mu_f))\| \leq \lim_{j \to \infty} \|\phi(x_j(\mu_{f,j})) - \phi(x(\mu_{f,j}))\| + \|\phi(x(\mu_{f,j})) - \phi(x(\mu_f))\| \quad \text{(A.11)}$$

Since $\lim_{j\to\infty}\mu_{f,j}=\mu_f$, $x_j\to x$ uniformly on $[0,\mu_f]$, and ϕ is continuous, we have

$$\lim_{j\to\infty}\|\phi(x_j(\mu_{f,j}))-\phi(x(\mu_{f,j}))\|=0. \tag{A.12}$$

Since ϕ, x are continuous, we have

$$\lim_{j\to\infty}\|\phi(x(\mu_{f,j}))-\phi(x(\mu_f))\|=0. \tag{A.13}$$

Combining these two steps, we have our result. □(*Step 1*)

Step 2: $\lim_{j\to\infty}\int_0^{\mu_{f,j}} L(x_j(t),u_j(t))dt=\int_0^{\mu_f} L(x(t),u(t))dt$

Proof of Step 2: First, fix an $\epsilon>0$. Recall by Proposition 2, $\exists j_0$ such that $\forall j>j_1$, $\|x_j(t)-x(t)\|<\epsilon$, $\forall t\in[0,\mu_f]$. Without loss of generality, assume that $j_0=1$. In order to prove this result, we want to apply to apply the Dominated Convergence Theorem. First, observe that each $L(x_j(t),u_j(t))$ belongs in $L^1,\forall j$. Second, observe that $L(x_j(t),u_j(t))\to L(x(t),u(t))$ almost everywhere, since L is continuous.

Finally, we construct a nonnegative $g:\mathbb{R}^1\to\mathbb{R}^1$, $g\in L^1$ such that $|L(x_j(t),u_j(t))|\le g(t)$, almost everywhere, $\forall j$:

$$g(t)=\max_{\substack{y\in\mathcal{B}^n(x(t),\epsilon)\\ v\in\mathcal{B}^m(0,M)}} L(y,v) \tag{A.14}$$

where $\mathcal{B}^n(x(t),\epsilon)$ denotes a ball in $\mathbb{R}^n$ of radius ϵ centered at $x(t)$, and $\mathcal{B}^m(0,M)$ denotes a ball in $\mathbb{R}^m$ of radius M centered at the origin. Since L is continuous and we are considering maximums over compact sets, g is well-defined, belongs in L^1, and dominates each $|L(x_j(t),u_j(t))|,\forall j$, as required. Thus, the Dominated Convergence Theorem immediately gives the desired result. □(*Step 2*)

Finally, combining the result of the previous two steps gives us sequential continuity which immediately implies continuity as desired. □

Finally, we must check that $\{\xi\in\mathcal{X}\mid\psi(\xi)\le 0\}$ is a closed set in order to apply Theorem 1. Since we are employing inequality constraints, showing that ψ is continuous gives us the required result.

Proposition 4. *The function ψ as defined in equation* (14) *is continuous.*

Proof. Using Lemma 5.6.7 together with Theorem 4.1.5 from [5] the result immediately follows. □

A.2 Optimality Function

In this section, we prove the convergence of Algorithm 1. Our algorithm works by inserting a new mode, $\hat{\alpha}$, in a small interval of length $\lambda\ge 0$ centered at a time, $\hat{t}$, with input $\hat{u}$. We begin by defining this type of insertion.

Definition 2. *Given $\xi=(\sigma,s,u)\in\mathcal{X}$ and $\eta=(\hat{\alpha},\hat{t},\hat{u})\in\mathcal{Q}\times[0,\mu_f]\times\mathcal{B}^m(0,M)$, we define the function $\rho^{(\eta)}:[0,\infty)\to\mathcal{X}$ as the perturbation of ξ after the insertion of mode $\hat{\alpha}$, at time $\hat{t}$ using $\hat{u}$ as the control, for a time interval of length λ. Let $\bar{\lambda}=\min_{\{i:|\mu(i)-\hat{t}|>0\}}\frac{1}{2}\left|\mu(i)-\hat{t}\right|$, then we write*

$\rho^{(\eta)}(\lambda) = \left(\rho_\sigma^{(\eta)}(\lambda), \rho_s^{(\eta)}(\lambda), \rho_u^{(\eta)}(\lambda)\right)$, *whenever* $\lambda \in [0, \bar{\lambda}]$,

$$\rho_\sigma^{(\eta)}(\lambda) = \begin{cases} (\hat{\alpha}, \sigma(1), \sigma(2), \dots) & \text{if } \hat{t} = 0 \\ (\sigma(1), \dots, \pi(\hat{t} - \frac{\bar{\lambda}}{2}), \hat{\alpha}, \dots) & \text{if } \hat{t} = \mu_f \\ (\sigma(1), \dots, \pi(\hat{t} - \frac{\bar{\lambda}}{2}), \hat{\alpha}, \pi(\hat{t} + \frac{\bar{\lambda}}{2}), \dots) & \text{if } \hat{t} \neq \mu(i) \\ (\sigma(1), \dots, \pi(\hat{t} - \frac{\bar{\lambda}}{2}), m(\kappa(\hat{t})) + 1, \dots, n(\kappa(\hat{t}), \hat{\alpha}, \pi(\hat{t} + \frac{\bar{\lambda}}{2}), \dots) & \text{if } \hat{t} = \mu(i) \end{cases} \tag{A.15}$$

$$\rho_s^{(\eta)}(\lambda) = \begin{cases} (\lambda, s(1) - \lambda, s(2), \dots) & \text{if } \hat{t} = 0 \\ (s(1), \dots, s(\kappa(\mu_f)) - \lambda, \lambda, 0, \dots) & \text{if } \hat{t} = \mu_f \\ (s(1), \dots, \hat{t} - \frac{\lambda}{2} - \mu(\kappa(\hat{t} - \frac{\bar{\lambda}}{2}) - 1), \lambda, \mu(\kappa(\hat{t} + \frac{\bar{\lambda}}{2})) - \hat{t} - \frac{\lambda}{2}, \dots) & \text{if } \hat{t} \neq \mu(i) \\ (s(1), \dots, \hat{t} - \frac{\lambda}{2} - \mu(\kappa(\hat{t} - \frac{\bar{\lambda}}{2}) - 1), s(m(i) + 1), \dots, & \\ \qquad \dots, s(n(i)), \lambda, \mu(\kappa(\hat{t} + \frac{\bar{\lambda}}{2})) - \hat{t} - \frac{\lambda}{2}, \dots) & \text{if } \hat{t} = \mu(i) \end{cases} \tag{A.16}$$

$$\rho_u^{(\eta)}(\lambda) = \begin{cases} u(t) + \big(\hat{u} - u(t)\big)\mathbb{1}_{[0,\lambda]}(t) & \text{if } \hat{t} = 0 \\ u(t) + \big(\hat{u} - u(t)\big)\mathbb{1}_{[\mu_f - \lambda, \mu_f]}(t) & \text{if } \hat{t} = \mu_f \\ u(t) + \big(\hat{u} - u(t)\big)\mathbb{1}_{[\hat{t} - \frac{\lambda}{2}, \hat{t} + \frac{\lambda}{2}]}(t) & \text{otherwise} \end{cases}, \tag{A.17}$$

and $\rho^{(\eta)}(\lambda) = \rho^{(\eta)}(\bar{\lambda})$ *whenever* $\lambda > \bar{\lambda}$.

Note that ρ, in addition to being a function of λ and η, is also a function of ξ, but we do not make this dependence explicit for notational convenience. If the dependence of ρ with respect to ξ is not clear, then we make the declaration explicit. Importantly, observe that $\rho_u^{(\eta)}$ is a needle variation (or strong variation) of the control $u(t)$ (as defined in Chapter 2, Section 13 of [6]). Figure 2 illustrates a pair (σ, s) after they are modified by the function $\rho^{(\eta)}$.

Proposition 5. *Given* $\eta \in \mathcal{Q} \times [0, \mu_f] \times \mathcal{B}^m(0, M)$*, the function* $\rho^{(\eta)}$ *is continuous.*

Proof. Note that $\rho_\sigma^{(\eta)}(\lambda) = \rho_\sigma^{(\eta)}(0)$ for each $\lambda > 0$, hence $\rho_\sigma^{(\eta)}$ is trivially continuous. Since the dependence on λ is affine for $\rho_s^{(\eta)}$, it is continuous. Finally, $\rho_u^{(\eta)}$ can be proven continuous by considering a convergent sequence $(\lambda_i)_{i=1}^\infty$, $\lambda_i \to \lambda$ as $i \to \infty$, and checking that $\rho_u^{(\eta)}(\lambda_i)$ converges to $\rho_u^{(\eta)}(\lambda)$ as $i \to \infty$. For values of λ such that $\lambda > \bar{\lambda}$ the function is continuous since it is constant. ∎

The previous proposition helps us understand the variation of the cost with respect to an insertion. We begin by studying the variation of the trajectory, $x^{(\rho(\lambda))}$, as λ changes. Note that $x^{(\xi)}(t) = x^{(\rho(0))}(t)$ for each $t \geq 0$, so a first order approximation of the trajectory is characterized by the directional derivative of $x^{(\rho(\lambda))}$ at $\lambda = 0$. To reduce the number of cases we need to consider in the future propositions, we define for a given $x : [0, \infty) \to \mathbb{R}^n$, $u : [0, \infty) \to \mathbb{R}^m$, and $\eta = (\hat{\alpha}, \hat{t}, \hat{u})$:

$$\Delta f(x, u, \eta) = \begin{cases} f_{\hat{\alpha}}(x(\hat{t}), \hat{u}) - f_{\pi(\hat{t} + \bar{\lambda})}(x(\hat{t}), u(\hat{t})) & \text{if } \hat{t} = 0 \\ f_{\hat{\alpha}}(x(\hat{t}), \hat{u}) - f_{\pi(\hat{t} - \bar{\lambda})}(x(\hat{t}), u(\hat{t})) & \text{if } \hat{t} = \mu_f \\ f_{\hat{\alpha}}(x(\hat{t}), \hat{u}) - \frac{1}{2}\left[f_{\pi(\hat{t} + \bar{\lambda})}(x(\hat{t}), u(\hat{t})) + f_{\pi(\hat{t} - \bar{\lambda})}(x(\hat{t}), u(\hat{t}))\right] & \text{otherwise} \end{cases} \tag{A.18}$$

where $\bar{\lambda} = \min_{\{i : |\mu(i) - \hat{t}| > 0\}} \frac{1}{2}\left|\mu(i) - \hat{t}\right|$. Observe that $\pi(\hat{t} + \bar{\lambda}) = \pi(\hat{t} - \bar{\lambda})$ whenever $\hat{t} \notin \{\mu(i)\}_{i \in \mathbb{N}}$. With an abuse of notation, we denote $x^{(\rho(\lambda))}$ by $x^{(\lambda)}$.

Proposition 6. *The directional derivative of $x^{(\lambda)}$ for λ positive, evaluated at zero, is:*

$$\left.\frac{dx^{(\lambda)}}{d\lambda}\right|_{\lambda=0}(t) = \begin{cases} \Phi(t,\hat{t})\Delta f(x^{(\xi)}, u, \eta) & \text{if } t \in [\hat{t}, \mu_f] \\ 0 & \text{otherwise} \end{cases}, \tag{A.19}$$

where $\Phi : [0,\infty) \times [0,\infty) \to \mathbb{R}^{n\times n}$ is the solution of the matrix differential equation:

$$\dot{X}(t,\hat{t}) = \frac{\partial f_{\pi(t)}}{\partial x}\big(x^{(\xi)}(t), u(t)\big)\; X(t,\hat{t}), \quad X(\hat{t},\hat{t}) = I. \tag{A.20}$$

Proof. The vector fields corresponding to ξ and $\rho^{(\eta)}(\lambda)$, $\lambda > 0$, are equal everywhere except in a small interval of length λ. Without loss of generality we assume that $\hat{t} \in (0, \mu_f)$, since the proof for the cases when $\hat{t} \in \{0, \mu_f\}$ is completely analogous and the case $\hat{t} > \mu_f$ is trivial given the fact that the flows are null. Let $\Delta x_\lambda(t) = x^{(\lambda)}(t) - x^{(0)}(t)$, and let $g(x(t), t)$ be its vector field. Recalling that $x^{(0)}(t) = x^{(\xi)}(t)$ for each $t \geq 0$:

$$g(x^{(\xi)}(t), t) = f_{\hat{\alpha}}(x^{(\xi)}(t), \hat{u}) - f_{\pi(t)}(x^{(\xi)}(t), u(t)) \tag{A.21}$$

whenever $t \in [\hat{t} - \frac{\lambda}{2}, \hat{t} + \frac{\lambda}{2}]$, and $g(x(t), t) = 0$ otherwise. Since $x^{(\lambda)}$ and $x^{(0)}$ have the same initial condition, $\Delta x_\lambda(t) = 0$ for each $t \in [0, \hat{t} - \frac{\lambda}{2})$. Therefore, as $\lambda \downarrow 0$ we have the desired result for $t \in [0, \hat{t})$.

For $t \geq \hat{t} - \frac{\lambda}{2}$, using equation (7) from [1] together with the conditions stated in Assumption 1, we know there exists $K > 0$ such that

$$\left\| \Delta x_\lambda(t) - \int_0^{\mu_f} \Phi(t,s) g\big(x^{(\xi)}(s), s\big) ds \right\| \leq K\lambda^2 \tag{A.22}$$

, or

$$\left\| \frac{\Delta x_\lambda(t)}{\lambda} - \int_{\mathbb{R}} \Phi(t,s) g\big(x^{(\xi)}(s), s\big) \frac{\mathbb{1}_{[\hat{t}-\frac{\lambda}{2}, \hat{t}+\frac{\lambda}{2}]}(s)}{\lambda} ds \right\| \leq K\lambda. \tag{A.23}$$

If we split the interval in two at $\hat{t}$, the integral in equation (A.23) can be written as the sum of two convolutions between $\Phi(t,s)g(x^{(\xi)}(s), s)$ and $\varphi_\lambda(s) = \frac{1}{\lambda}\mathbb{1}_{[0,\frac{1}{2}]}(\frac{s-\hat{t}}{\lambda})$. Since both $\Phi(t,\cdot)g(x^{(\xi)}(\cdot), \cdot)$ and φ_λ are functions in L^1, we can apply Theorem 8.14 from [2] and the triangular inequality to get

$$\lim_{\lambda \downarrow 0} \frac{\Delta x_\lambda}{\lambda}(t) = \Phi(t,\hat{t})\Delta f(x^{(\xi)}, u, \eta) \tag{A.24}$$

whenever $t \in [\hat{t}, \mu_f]$. In this last step, we used the fact that $\pi(\hat{t} \pm \frac{\lambda}{2}) = \pi(\hat{t} \pm \bar{\lambda})$ for each $\lambda \in (0, \bar{\lambda}]$. ◻

Given this variation of the state trajectory, we can now consider variations of the cost and constraint functions, which allows us to define our optimality function θ in a manner that guarantees if there are no feasible mode insertions which lower the cost then $\theta(\xi) = 0$.

Proposition 7. *Let J be the cost function defined in equation* (12)*. Then the directional derivative of $J(\rho^{(\eta)}(\lambda))$ evaluated at $\lambda = 0$ is*

$$\left.\frac{dJ(\rho^{(\eta)}(\lambda))}{d\lambda}\right|_{\lambda=0} = \Big(p^{(\xi)}(\hat{t})\Big)^T \Delta f(x^{(\xi)}, u, \eta) \; + \big[\hat{u} - u(\hat{t})\big]^T \frac{\partial L}{\partial u}\big(x^{(\xi)}(\hat{t}), u(\hat{t})\big) \tag{A.25}$$

where $p^{(\xi)}$, can be identified with the costate, and is the solution to the following differential equation

$$\begin{aligned}\dot{p}(t) &= -\frac{\partial f_{\pi(t)}}{\partial x}^T \big(x^{(\xi)}(t), u(t)\big)\; p(t) - \frac{\partial L}{\partial x}\big(x^{(\xi)}(t), u(t)\big) \\ p(\mu_f) &= \frac{\partial \phi}{\partial x}\big(x^{(\xi)}(\mu_f)\big).\end{aligned} \tag{A.26}$$

Proof. Without loss of generality, assume $\hat{t} \in (0, \mu_f)$. The case when $\hat{t} \in \{0, \mu_f\}$ is completely analogous. Let $\rho_u^{(\eta)}(\lambda)$ from equation (A.17) be denoted by u_λ. Also, denote $x^{(0)}$ by x. Consider

$$\begin{aligned}&\frac{J(\rho^{(\eta)}(\lambda)) - J(\rho^{(\eta)}(0))}{\lambda} = \\ &\qquad = \int_0^{\mu_f} \frac{1}{\lambda}\Big[L\big(x^{(\lambda)}(t), u_\lambda(t)\big) - L\big(x(t), u(t)\big)\Big]\, dt + \frac{1}{\lambda}\Big[\phi\big(x^{(\lambda)}(\mu_f)\big) - \phi\big(x(\mu_f)\big)\Big].\end{aligned} \tag{A.27}$$

Since L and ϕ are differentiable, we can apply the Mean Value Theorem; thus, there exists functions $s_x, s_u : [0, \mu_f] \to (0, 1)$ such that

$$L(x^{(\lambda)}(t), u_\lambda(t)) - L(x(t), u(t)) = g_{x,\lambda}(t)\Delta x_\lambda(t) + g_{u,\lambda}(t)\Delta u_\lambda(t), \tag{A.28}$$

and there exists a constant $s_y \in (0, 1)$ such that

$$\phi(x^{(\lambda)}(\mu_f)) - \phi(x(\mu_f)) = \frac{\partial \phi}{\partial x}^T\big(x(\mu_f) + s_y \Delta x_\lambda(\mu_f)\big)\Delta x_\lambda(\mu_f), \tag{A.29}$$

where $\Delta x_\lambda(t) = x_\lambda(t) - x(t)$, $\Delta u_\lambda(t) = u_\lambda(t) - u(t)$, and

$$g_{x,\lambda}(t) = \frac{\partial L}{\partial x}^T\big(x(t) + s_x(t)\Delta x_\lambda(t), u(t) + s_u(t)\Delta u_\lambda(t)\big) \tag{A.30}$$

$$g_{u,\lambda}(t) = \frac{\partial L}{\partial u}^T\big(x(t) + s_x(t)\Delta x_\lambda(t), u(t) + s_u(t)\Delta u_\lambda(t)\big). \tag{A.31}$$

Abusing notation, we denote $\frac{\partial L}{\partial x}^T(x(t), u(t))$ and $\frac{\partial L}{\partial u}^T(x(t), u(t))$ by $g_{x,0}(t)$ and $g_{u,0}(t)$, respectively.

Recall by Proposition 6, the limit of $\frac{\Delta x_\lambda(\mu_f)}{\lambda}$ exists as $\lambda \downarrow 0$. Also, by Proposition 1, $\Delta x_\lambda(t) \to 0$ as $\lambda \downarrow 0$ for each $t \in [0, \mu_f]$. Since $\frac{\partial \phi}{\partial x}$ is continuous, equation (A.29) becomes

$$\lim_{\lambda \downarrow 0} \frac{\phi(x^{(\lambda)}(\mu_f)) - \phi(x(\mu_f))}{\lambda} = \frac{\partial \phi}{\partial x}^T\big(x(\mu_f)\big)\Phi(\mu_f, \hat{t})\Delta f(x, u, \eta). \tag{A.32}$$

The computation of the limit of the integral term in equation (A.27) is carried out in two steps, one for each term in equation (A.28). First note that

$$\lim_{\lambda \downarrow 0} \int_0^{\mu_f} g_{u,0}(t)\frac{\Delta u_\lambda(t)}{\lambda} dt = \big[\hat{u} - u(\hat{t})\big]^T \frac{\partial L}{\partial u}\big(x(\hat{t}), u(\hat{t})\big) \tag{A.33}$$

which is the result of first writing the integral as a convolution between $g_{u,0}(t)$ and $\varphi_\lambda(t) = \frac{1}{\lambda}\mathbb{1}_{[-\frac{1}{2},\frac{1}{2}]}(\frac{t-\hat{t}}{\lambda})$ and then using Theorem 8.14 from [2]. Since $g_{u,\lambda}(t) \to g_{u,0}(t)$ for each $t \in [0, \mu_f]$,

$(g_{u,\lambda} - g_{u,0}) \in L^\infty$, and $\frac{\Delta u_\lambda}{\lambda} \in L^1$ for each $\lambda > 0$, we apply Hölder's inequality to obtain the following bound

$$\int_0^{\mu_f} \left\| [g_{u,\lambda}(t) - g_{u,0}(t)] \frac{\Delta u_\lambda(t)}{\lambda} \right\| dt \leq \| g_{u,\lambda} - g_{u,0} \|_\infty \left\| \frac{\Delta u_\lambda}{\lambda} \right\|_1 \to 0 \tag{A.34}$$

as $\lambda \downarrow 0$. Therefore, using the triangular inequality together with equations (A.33) and (A.34) we get the desired result for the second term in equation (A.28).

For the first term in equation (A.28), we employ a similar argument via Hölder's

$$\int_0^{\mu_f} \left\| g_{x,\lambda}(t) \frac{\Delta x_\lambda(t)}{\lambda} - g_{x,0}(t) \frac{dx_\lambda}{d\lambda}(t) \right\| dt \leq \int_0^{\mu_f} \left\| g_{x,\lambda}(t) \left[\frac{\Delta x_\lambda(t)}{\lambda} - \frac{dx_\lambda}{d\lambda}(t) \right] \right\| dt + \tag{A.35}$$

$$+ \int_0^{\mu_f} \left\| [g_{x,\lambda}(t) - g_{x,0}(t)] \frac{dx_\lambda}{d\lambda}(t) \right\| dt \tag{A.36}$$

$$\leq \| g_{x,\lambda} \|_\infty \left\| \frac{\Delta x_\lambda}{\lambda} - \frac{dx_\lambda}{d\lambda} \right\|_1 + \| g_{x,\lambda} - g_{x,0} \|_\infty \left\| \frac{dx_\lambda}{d\lambda} \right\|_1 \tag{A.37}$$

where $g_{x,\lambda} \in L^\infty$ for each $\lambda > 0$ (since $\frac{\partial L}{\partial x}$ is continuous), the space of acceptable controls is bounded, the trajectories are continuous, and the domain is compact. Also, $\frac{dx_\lambda}{d\lambda} \in L^1$ since it is a piecewise-continuous function over a compact domain, and $\frac{\Delta x_\lambda}{\lambda} \in L^1$ since both $x^{(\lambda)}$ and $x^{(0)}$ are continuous over a compact domain. Hence, by the pointwise convergence of $\frac{\Delta x_\lambda}{\lambda}$ to $\frac{dx_\lambda}{d\lambda}$, we have

$$\lim_{\lambda \downarrow 0} \int_0^{\mu_f} g_{x,\lambda}(t) \frac{\Delta x_\lambda(t)}{\lambda} dt = \int_0^{\mu_f} \frac{\partial L}{\partial x}^T \big(x(t), u(t) \big) \Phi(t, \hat{t}) dt \; \Delta f(x, u, \eta). \tag{A.38}$$

Finally, if we sum equations (A.32), (A.33), and (A.38), and note that the solution to the differential equation (A.26) is

$$p^{(\xi)}(t) = \Phi^T(\mu_f, \hat{t}) \frac{\partial \phi}{\partial x}(x(\mu_f)) + \int_0^{\mu_f} \Phi^T(t, \hat{t}) \frac{\partial L}{\partial x}(x(t), u(t)) dt \tag{A.39}$$

we get the desired result. □

In order to define our optimality function, we must also consider variations of the constraint function after the mode insertion procedure.

Proposition 8. *Let ψ be the constraint function defined in* (14). *The directional derivative of $\psi(\rho^{(\eta)}(\lambda))$ evaluated at $\lambda = 0$ is*

$$\left. \frac{d\psi(\rho^{(\eta)}(\lambda))}{d\lambda} \right|_{\lambda=0} = \max_{j \in \hat{\mathcal{J}}(\xi)} \max_{t \in \hat{\mathcal{T}}_j(\xi)} \frac{\partial h_j}{\partial x} \big(x^{(\xi)}(t) \big) \left. \frac{dx^{(\lambda)}}{d\lambda} \right|_{\lambda=0} (t) \tag{A.40}$$

where

$$\hat{\mathcal{J}}(\xi) = \left\{ j \in \{1, \ldots, R\} \;\middle|\; \max_{t \in \mathcal{T}_j} h_j \big(x^{(\xi)}(t) \big) = \psi(\xi) \right\} \tag{A.41}$$

$$\hat{\mathcal{T}}_j(\xi) = \left\{ t \in [0, \mu_f] \;\middle|\; h_j \big(x^{(\xi)}(t) \big) = \max_{\bar{t} \in [0, \mu_f]} h_j \big(x^{(\xi)}(\bar{t}) \big) \right\} \tag{A.42}$$

Proof. Using Theorem 5.4.6 and the proof of Theorem 5.4.7 in [5], we get that

$$\left.\frac{d\psi(\rho^{(\eta)}(\lambda))}{d\lambda}\right|_{\lambda=0} = \max_{j\in\hat{\mathcal{J}}(\xi)} \max_{t\in\hat{\mathcal{T}}_j(\xi)} \left.\frac{dh\big(x^{(\lambda)}\big)}{d\lambda}\right|_{\lambda=0}(t). \tag{A.43}$$

The result is clear from this equation by using the chain rule. □

Now we can prove that if $\theta(\xi)$ as defined in equation (19) is less than zero, then there exists a feasible mode insertion which reduces the overall cost (i.e. our optimality function captures the points of interest).

Theorem 2. *Consider the function θ defined in equation (19). Let $\xi \in \mathcal{X}$ and $\eta = (\hat{\alpha}, \hat{t}, \hat{u}) \in \mathcal{Q} \times [0, \mu_f] \times \mathcal{B}^m(0, M)$ be the argument which minimizes $\theta(\xi)$. If $\theta(\xi) < 0$, then there exists $\hat{\lambda} > 0$ such that, for each $\lambda \in (0, \hat{\lambda}]$, $J(\rho^{(\eta)}(\lambda)) \leq J(\xi)$ and $\psi(\rho^{(\eta)}(\lambda)) \leq 0$.*

Proof. First, we show that the functions $J \circ \rho^{(\eta)}$ and $\psi \circ \rho^{(\eta)}$, in addition to being differentiable at $\lambda = 0$ are also differentiable in a neighborhood around zero and their derivatives are upper semicontinuous in that same neighborhood. Consider $\lambda \geq 0$ sufficiently small (i.e. $\lambda < \bar{\lambda}$, with $\bar{\lambda}$ as in Definition 2), then using an argument analogous to the one presented in the proof of Proposition 7 we get

$$\frac{d(J\circ\rho^{(\eta)})}{d\lambda}(\lambda) = \Big(p^{(\lambda)}(\hat{t}+\lambda)\Big)^T \Delta f\big(x^{(\lambda)}, u, \eta_\lambda\big) + \big[\hat{u} - u(\hat{t}+\lambda)\big]^T \frac{\partial L}{\partial u}\big(x^{(\lambda)}(\hat{t}+\lambda), u(\hat{t}+\lambda)\big) \tag{A.44}$$

where $\eta_\lambda = (\hat{\alpha}, \hat{t}+\lambda, \hat{u})$. Note that the proof of Proposition 2 can also be applied to prove the uniform convergence of $p^{(\lambda)}$ to $p^{(\xi)}$ as $\lambda \downarrow 0$. Therefore, together with Assumption 3, we get that for any sequence $\{\lambda_i\}_{i=1}^{\infty}$ such that $\lambda_i \downarrow 0$, $\frac{d(J\circ\rho)}{d\lambda}(\lambda_i) \to \frac{d(J\circ\rho)}{d\lambda}(0)$, which proves the continuity of $\frac{d(J\circ\rho)}{d\lambda}$ at zero.

Similarly, using an argument analogous to the one presented in the proof of Proposition 8, we get that for λ sufficiently small,

$$\frac{d(\psi\circ\rho^{(\eta)})}{d\lambda}(\lambda) = \max_{j\in\hat{\mathcal{J}}(\rho^{(\eta)}(\lambda))} \max_{t\in\hat{\mathcal{T}}_j(\rho^{(\eta)}(\lambda))} \frac{\partial h_j}{\partial x}\big(x^{(\lambda)}(t)\big)\frac{dx^{(\lambda)}}{d\lambda}(t) \tag{A.45}$$

where $\hat{\mathcal{J}}$ and $\hat{\mathcal{T}}$ are defined in (A.41) and (A.42), respectively. Following the proof of Proposition 6, the directional derivative of $x^{(\lambda)}$ with respect to λ, for λ small enough, is

$$\frac{dx^{(\lambda)}}{d\lambda}(t) = \Phi(t, \hat{t}+\lambda)\Delta f\big(x^{(\lambda)}, u, \eta_\lambda\big) \tag{A.46}$$

if $t \in [\hat{t}+\lambda, \mu_f]$, and $\frac{dx^{(\lambda)}}{d\lambda}(t) = 0$ otherwise. Then $\frac{d(\psi\circ\rho^{(\eta)})}{d\lambda}$ is upper semicontinuous since the functions being maximized are continuous on λ and t, both set-valued maps $\hat{\mathcal{J}}$ and $\hat{\mathcal{T}}$ are outer semicontinuous, and for each $\xi \in \mathcal{X}$ the sets $\hat{\mathcal{J}}(\xi)$ and $\hat{\mathcal{T}}(\xi)$ are compact, hence satisfying the assumptions of Theorem 5.4.1 in [5].

Recall that for any η, $J(\rho^{(\eta)}(0)) = J(\xi)$ and $\psi(\rho^{(\eta)}(0)) = \psi(\xi)$. Then, by the Mean Value Theorem, given $\lambda > 0$ sufficiently small there exists $s_1, s_2 \in (0,1)$ such that

$$J(\rho^{(\eta)}(\lambda)) - J(\xi) = \frac{d(J\circ\rho^{(\eta)})}{d\lambda}(s_1\lambda)\cdot\lambda \tag{A.47}$$

$$\psi(\rho^{(\eta)}(\lambda)) - \psi(\xi) = \frac{d(\psi\circ\rho^{(\eta)})}{d\lambda}(s_2\lambda)\cdot\lambda. \tag{A.48}$$

As a consequence of the derivatives being upper semicontinuous, for each $\epsilon_1, \epsilon_2 > 0$ there exists a $\hat{\lambda} > 0$ such that for each $\lambda \in [0, \hat{\lambda})$,

$$\frac{d(J \circ \rho^{(\eta)})}{d\lambda}(\lambda) < \frac{d(J \circ \rho^{(\eta)})}{d\lambda}(0) + \epsilon_1 \tag{A.49}$$

$$\frac{d(\psi \circ \rho^{(\eta)})}{d\lambda}(\lambda) < \frac{d(\psi \circ \rho^{(\eta)})}{d\lambda}(0) + \epsilon_2 \tag{A.50}$$

It is immediate from the definition of θ that if $\theta(\xi) < 0$ then

$$\frac{d(J \circ \rho^{(\eta)})}{d\lambda}(0) < 0 \quad \text{and} \quad \psi(\xi) + \frac{d(\psi \circ \rho^{(\eta)})}{d\lambda}(0) < 0. \tag{A.51}$$

Using $\epsilon_1 = -\frac{1}{2}\frac{d(J \circ \rho^{(\eta)})}{d\lambda}(0)$ in equation (A.49), there exists a $\hat{\lambda}_1 > 0$ such that for each $\lambda \in [0, \hat{\lambda}_1)$ the following bound for equation (A.47) holds

$$J(\rho^{(\eta)}(\lambda)) - J(\xi) < \frac{1}{2}\frac{d(J \circ \rho^{(\eta)})}{d\lambda}(0) \cdot \lambda \leq 0. \tag{A.52}$$

In order to produce a similar bound for $(\psi \circ \rho^{(\eta)})$, we consider two cases. First, suppose $\psi(\xi) < 0$. Since both ψ and ρ are continuous functions, there exists $\hat{\lambda}_2 > 0$ such that for each $\lambda \in [0, \hat{\lambda}_2)$, $\psi(\rho^{(\eta)}(\lambda)) < 0$. Second, suppose $\psi(\xi) = 0$, then equation (A.51) implies that $\frac{d(\psi \circ \rho^{(\eta)})}{d\lambda}(0) < 0$. Choosing $\epsilon_2 = -\frac{1}{2}\frac{d(\psi \circ \rho^{(\eta)})}{d\lambda}(0)$ in equation (A.50), there exists a $\hat{\lambda}_2 > 0$ such that for each $\lambda \in [0, \hat{\lambda}_2)$ the following bound for equation (A.48) holds

$$\psi(\rho^{(\eta)}(\lambda)) < \frac{1}{2}\frac{d(\psi \circ \rho^{(\eta)})}{d\lambda}(0) \cdot \lambda \leq 0. \tag{A.53}$$

Finally, by choosing $\hat{\lambda} = \min\{\hat{\lambda}_1, \hat{\lambda}_2\}$, both inequalities (A.52) and (A.53) hold. □

This result proves that the vanishing points of our optimality function for Algorithm 1 contain solutions to our optimal control problem. We now address the validity of Assumption 4. First, recall that ρ is a needle variation; therefore, as a result of the previous theorem, if $\theta(\xi) < 0$ then we are not at a minimum in the sense of Pontryagin [4]. Unfortunately, numerical methods for optimization cannot implement these types of variations since that task would require the approximation of arbitrarily narrow discontinuous functions. This means that any practical algorithm using a numerical method would find minima that do not necessarily coincide with the minima prescribed by our θ function. If we were uninterested in constructing a practical algorithm, then Assumption 4 would be trivially satisfied by any of the theoretical algorithms proposed by Pontryagin.

However, we can construct a practical algorithm based on the following proposition.

Proposition 9. *If the vector fields $\{f_q\}_{q \in \mathcal{Q}}$ are affine with respect to the control and the running cost L is convex with respect to the control, then the optimality condition calculated via vector-space variations (variations that take the form of directional derivatives) and the optimality condition calculated via needle variations are equivalent.*

Proof. This proposition is proven in Section 4.2.6 in [5]. □

Under the hypotheses of the proposition above, there are many algorithms that satisfy Assumption 4, among them the algorithms described in Section 4.5 in [5]. Importantly, we can transform any nonlinear vector field into a new vector field that is affine with respect to its control using the following transformation:

$$\begin{pmatrix} \dot{x}(t) \\ \dot{z}(t) \end{pmatrix} = \begin{pmatrix} f_{\pi(t)}(x(t), z(t)) \\ v(t) \end{pmatrix}, \tag{A.54}$$

where $(x(t), z(t))^T$ in $\mathbb{R}^{m+n}$ becomes the new state variables, and $v(t) \in \mathbb{R}^m$ becomes the new control input. After the transformation those same algorithms would guarantee the validity of Assumption 4.

Finally, we can show that Algorithm 1 has the sufficient descent property with respect to our optimality function.

Theorem 3. *Algorithm $a : \mathcal{X} \to \mathcal{X}$, as defined in Algorithm 1 has the sufficient descent property with respect to the function $\theta : \mathcal{X} \to \mathbb{R}$.*

Proof. Note that given $\xi = (\sigma, s, u)$ and $\eta = (\hat{\alpha}, \hat{t}, \hat{u})$, the result of algorithm a when applied to ξ can be written as $a(\xi) = \big(\rho_\sigma, \hat{a}(\rho_s, \rho_u)\big)$, where $\hat{a}$ is the algorithm that solves Stage 1 in our Bi-Level Optimization Scheme and $(\rho_\sigma, \rho_s, \rho_u) \in \mathcal{X}$ is the element resulting from the function $\rho^{(\eta)}$ evaluated at $\lambda = 0$. Since $\mathcal{X}$ is a metric space, we can assume without loss of generality that the neighborhoods are of the form $\mathcal{B}_{\mathcal{X}}(\xi, \epsilon) = \{\xi' \in \mathcal{X} \mid d(\xi, \xi') < \epsilon\}$. Similar definitions follow for the neighborhoods of $\mathcal{S}$ and $\mathcal{U}$.

In order to prove the result, we must show that for each $\xi \in \mathcal{X}$ with $\theta(\xi) < 0$ there exists constants $\epsilon > 0$ and $\delta_\xi > 0$ such that for each $\xi' \in \mathcal{B}_{\mathcal{X}}(\xi, \epsilon)$ the following inequality holds

$$J\big(a(\xi')\big) - J(\xi') \leq -\delta. \tag{A.55}$$

Let $\xi \in \mathcal{X}$ such that $\theta(\xi) < 0$. By Assumption 4, we know that there exists $\epsilon_1 > 0$ and $\delta_1 > 0$ such that for each $s' \in \mathcal{B}_{\mathcal{S}}(s, \epsilon_1)$ and $u' \in \mathcal{B}_{\mathcal{U}}(u, \epsilon_1)$ the following inequality holds

$$J\big(\sigma, \hat{a}(s', u')\big) - J(\sigma, s', u') \leq -\delta_1. \tag{A.56}$$

Since the function $\rho^{(\eta)}$ evaluated at $\lambda = 0$ makes a modal insertion on an interval of time of length zero, if a pair $s_1, s_2 \in \mathcal{S}$ satisfies $\|s_1 - s_2\|_{l^1} < \epsilon$, then $\|\rho_{s_1} - \rho_{s_2}\|_{l^1} < \epsilon$. Similarly, if a pair $u_1, u_2 \in \mathcal{U}$ satisfies $\|u_1 - u_2\|_2 < \epsilon$, then $\|\rho_{u_1} - \rho_{u_2}\|_2 < \epsilon$.

Let $\epsilon > 0$ such that $\epsilon < \min\{\epsilon_1, 1\}$, and consider $\xi' \in \mathcal{B}_{\mathcal{X}}(\xi, \epsilon_1)$. Since $\epsilon < 1$, the modal sequences of ξ and ξ' are equal, hence denote ξ' as (σ, s', u'). By the inequality in equation (A.56), we have that

$$J\big(\rho_\sigma, \hat{a}(\rho_{s'}, \rho_{u'})\big) - J(\rho_\sigma, \rho_{s'}, \rho_{u'}) \leq -\delta_1 \tag{A.57}$$

Finally, note that $J(\rho_\sigma, \rho_{s'}, \rho_{u'}) = J(\sigma, s', u')$ since the modal insertion made by $\rho^{(\eta)}$ has no effect in the solution of the differential equation at $\lambda = 0$. Therefore we get the desired result for $\delta = \delta_1$. $\square$

Using this fact and Theorem 1, we have that our algorithm converges to points that satisfy our optimality condition as desired.

References

[1] M. Egerstedt, Y. Wardi, and H. Axelsson. Transition-Time Optimization for Switched-Mode Dynamical Systems. *IEEE Transactions on Automatic Control*, 51(1):110–115, 2006.

[2] G. B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Wiley, second edition, 1999.

[3] H. Gonzalez, R. Vasudevan, M. Kamgarpour, S. Sastry, R. Bajcsy, and C. Tomlin. A descent algorithm for the optimal control of constrained nonlinear switched dynamical systems. In *Proceedings of the 13th International Conference on Hybrid Systems: Computation and Control*, 2010.

[4] D. Mayne and E. Polak. First-Order Strong Variation Algorithms for Optimal Control. *Journal of Optimization Theory and Applications*, 16(3):277–301, 1975.

[5] E. Polak. *Optimization: Algorithms and Consistent Approximations*. Springer, 1997.

[6] L. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko. *The Mathematical Theory of Optimal Processes*. Interscience New York, 1962.

[7] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill New York, 1964.