Using max-plus algebra to compute reachability in dynamic systems

Gabe Flanagin

Advised by Rick Barnard

June 6, 2025

Contents

1 Introduction **1**

2 HJB equations **2**

2.1 General formulation . . . 2

2.2 Semigroup operators . . . 3

2.3 Linear quadratic control problems . . . 3

3 Quadratic under approximations **6**

3.1 Quadratic upper bound . . . 6

3.2 Quadratic lower bound: . . . 6

3.3 Efficient computation . . . 10

3.4 Converting quadratics to the desired form . . . 11

4 Max-plus algebra **11**

4.1 Intro and algebraic objects . . . 11

4.2 Additional notation . . . 12

4.3 Max-plus linearity of semigroup operators . . . 13

5 The algorithm **13**

5.1 Control problem associated to approximate Hamiltonian . . . 13

5.2 Time Discretization . . . 14

5.3 Semiconvex duality . . . 15

5.4 Steps for numerical implementation . . . 19

6 Appendix **23**

6.1 Theorem 8 proof . . . 23

7 References 25

1 Introduction

From indoor heating systems to financial risk management, optimal control has a wide range of practical applications in various fields of study. In general, optimal control is concerned with choosing a function of time u to influence a dynamic system $\dot{\xi} = f(\xi, u)$ while optimizing some running cost. The running cost associated to a choice of control u can be formulated as the functional

$$J(x, T; u) := \int_0^T \mathcal{L}(\xi_t, u_t)dt \qquad \xi_0 = x$$

where $\mathcal{L}$ is some problem dependent function. The subscripts represent the time variable, meaning the term u_t is equivalent to $u(t)$. We will use this notation throughout the report. If we wanted to maximize running cost, then we would define the optimal cost function(often called the value function)

$$V(x, T) := \sup_{u \in \mathcal{U}} J(x, T; u)$$

which gives the optimal cost of running the system for time T, starting from position x. The set $\mathcal{U}$ is a set of admissible control functions defined from $[0, \infty) \to \mathbb{R}^n$. We can describe $V(x, T)$ with the Hamilton Jacobi Bellman PDE, which takes the form

$$\frac{\partial V}{\partial T} = H(x, \nabla V) \qquad V(x, 0) = 0$$

where H is a function called the Hamiltonian, which will vary depending on the problem setup. We will discuss this in more detail in section 2. Efficient methods for approximating the solution to HJB PDEs have been developed for low dimensional problems(finite difference and finite element methods), but these methods often scale very poorly with dimension, suffering from the so called curse-of-dimensionality. Some common methods even scale at a rate of 200^n[2]. In this report, we explore a method developed by William McEneaney in [2][3][4] that avoids the curse-of-dimensionality by approximating nonlinear Hamiltonians with a pointwise max of quadratics and applying max-plus linear operators(see section 4) to progress an initial guess towards the steady state solution of the approximate HJB PDE. We begin by outlining how to derive the HJB equation.

2 HJB equations

2.1 General formulation

For this section we assume the following

$$J(x,T;u)=\int_0^T \mathcal{L}(\xi_t,u_t)dt \qquad V(x,T)=\inf_{u\in\mathcal{U}} J(x,T;u)$$

where $\dot{\xi}=f(\xi,u)$ and $\xi_0=x$. We assume the set of feasible controls $\mathcal{U}$ is chosen so that $V(x,T)$ exists for all $T\geq 0$. Consider some $w\in\mathcal{U}$, and let ξ^w be the trajectory that arises from the control choice w with $\xi_0^w=x\in\mathbb{R}^n$. For some $\tau\in[0,T]$, consider $J(x,\tau;w)+V(\xi_\tau^w,T-\tau)$, which describes the cost of traveling the trajectory ξ^w for τ seconds, plus the optimal cost for starting at ξ_τ^w and running the system for $T-\tau$ seconds. This interpretation naturally gives rise to the inequality

$$V(x,T)\leq J(x,\tau;w)+V(\xi_\tau^w,T-\tau)$$

where equality holds iff $w=\arg\min_{u\in\mathcal{U}} J(x,T;u)$ over the interval $[0,\tau]$. This is known as Bellman's principle of optimality[1]. If we think of ξ_τ^w as a new starting position, then for $\epsilon\in[0,T-\tau]$, applying the principle again gives

$$J(x,\tau;w)+V(\xi_\tau^w,T-\tau)\leq J(x,\tau;w)+J(\xi_\tau,\epsilon;w)+V(\xi_{\tau+\epsilon}^w,T-(\tau+\epsilon))$$

$$=J(x,\tau+\epsilon;w)+V(\xi_{\tau+\epsilon}^w,T-(\tau+\epsilon))$$

In other words $\mathcal{I}(x,\tau,T;w):=J(x,\tau;w)+V(\xi_\tau^w,T-\tau)$ is non-decreasing in τ. It follows that

$$0\leq\frac{d}{d\tau}\mathcal{I}(x,\tau,T;w)=\mathcal{L}(\xi_\tau^w,w_\tau)+\langle\nabla V(\xi_\tau^w,T-\tau),\dot{\xi}_\tau^w\rangle-\frac{\partial V}{\partial T}(\xi_\tau^w,T-\tau)$$

Furthermore, since $\mathcal{I}(x,0,T;w)=V(x,T)$ for all w, we can deduce that w is optimal iff $\frac{d}{d\tau}\mathcal{I}(x,\tau,T;w)=0$ for all $\tau\in[0,T]$. Setting $\tau=0$ and taking the inf over w gives the PDE

$$0=\inf_{w\in\mathcal{U}}\frac{d}{d\tau}\mathcal{I}(x,\tau,T;w)=\inf_{w\in\mathcal{U}}\left\{\mathcal{L}(x,w)+\langle\nabla V(x,T),\dot{\xi}_0^w\rangle-\frac{\partial V}{\partial T}(x,T)\right\}$$

$$=\inf_{w\in\mathcal{U}}\left\{\mathcal{L}(x,w)+\langle\nabla V(x,T),f(x,w)\rangle\right\}-\frac{\partial V}{\partial T}(x,T)$$

This is the Hamilton Jacobi Bellman equation for this problem. We express this as

$$\frac{\partial V}{\partial T}=H(x,\nabla V):=\inf_{w\in\mathcal{U}}\{\mathcal{L}(x,w)+\langle\nabla V,f(x,w)\rangle\} \qquad V(x,0)=0$$

where the initial condition comes from the definition of $J(x,0;u)$. We call H the Hamiltonian. If we instead wanted to solve $V^s(x,T) = \sup_{u\in\mathcal{U}} J(x,T;u)$, then with a few inequality flips the same argument could be used to find

$$\frac{\partial V^s}{\partial T} = H^s(x,\nabla V^s) := \sup_{u\in\mathcal{U}}\{\mathcal{L}(x,u) + \langle \nabla V^s, f(x,u)\rangle\} \qquad V(x,0) = 0$$

2.2 Semigroup operators

For this section we will assume

$$V(x,T) := \sup_{u\in\mathcal{U}} J(x,T;u) \text{ for some cost functional } J(x,T;u) := \int_0^T \mathcal{L}(\xi_t,u_t)dt$$

It is often useful to express the solution to this problem through an operator applied to $V(x,0)$. The operator is formulated using the Bellman optimality principle from the last section. That is, for some $\tau > 0$ and $\phi : \mathbb{R}^n \to \mathbb{R}$,

$$S_\tau[\phi](x) := \sup_{u\in\mathcal{U}} \Big\{ J(x,\tau;u) + \phi(\xi_\tau) \Big\}$$

where $\dot{\xi} = f(\xi,u)$ and $\xi_0 = x$. If we apply this operator to $V(x,T)$ for some $T > 0$, then

$$S_\tau[V(\cdot,T)](x) = \sup_{u\in\mathcal{U}} \Big\{ J(x,\tau;u) + V(\xi_\tau,T) \Big\} = \sup_{u\in\mathcal{U}} \Big\{ J(x,\tau;u) + \sup_{u^1\in\mathcal{U}} J(\xi_\tau,T;u^1) \Big\}$$

Letting $\xi_0^1 = \xi_\tau$, this is equal to

$$= \sup_{u\in\mathcal{U}} \sup_{u^1\in\mathcal{U}} \Big\{ \int_0^\tau \mathcal{L}(\xi_t,u_t)dt + \int_0^T \mathcal{L}(\xi_t^1,u_t^1)dt \Big\} = \sup_{u\in\mathcal{U}} \int_0^{T+\tau} \mathcal{L}(\xi_t,u_t)dt = V(x,T+\tau)$$

This means we can express $V(x,T)$ as

$$V(x,T) = S_T[V(\cdot,0)](x) = S_T[0](x)$$

A similar argument can be used to show $(S_{\tau_1} \circ S_{\tau_2})[\phi](x) = S_{\tau_1+\tau_2}[\phi](x)$ for all $\phi : \mathbb{R}^n \to \mathbb{R}$. This is why we call S_τ a semigroup operator, because $\big(\{S_\tau : \tau > 0\}, \circ\big)$ is a semigroup(no inverses).

2.3 Linear quadratic control problems

While many HJB PDEs involving nonlinear Hamiltonians(especially in higher dimensions) are very difficult to solve, problems involving quadratic Hamiltonians are relatively easy. Later in the report, we will use this to approximate nonlinear Hamiltonians with a pointwise max of quadratic Hamiltonians, greatly reducing the complexity of the PDE we want to solve. First however, we will give an interpretation of an HJB PDE with quadratic Hamiltonian. For $x, p \in \mathbb{R}^n$, consider a Hamiltonian

$$H^m(x,p) := \frac{1}{2}x^T D^m x + \frac{1}{2}p^T \Sigma^m p + p^T A^m x + (l_1^m)^T x + (l_2^m)^T p + \alpha^m$$

Where we assume $D^m \succ 0$(positive definite), $\Sigma^m = \frac{1}{\gamma^2}\sigma^m(\sigma^m)^T$ for some $\gamma > 0$ and $\sigma^m \in \mathbb{R}^{n\times n}$. We can work backwards to find that

$$H^m(x,p) = \sup_{u\in\mathcal{U}} \left\{x^T D^m x + p^T A^m x + (l_1^m)^T x + (l_2^m)^T p + \alpha^m + u^T\left[(\sigma^m)^T p - \frac{\gamma^2}{2}u\right]\right\}$$

$$= \sup_{u\in\mathcal{U}} \left\{ \left[x^T D^m x - \frac{\gamma^2}{2}|u|_2^2 + (l_1^m)^T x + \alpha^m\right] + \langle p, A^m x + l_2^m + \sigma^m u\rangle \right\}$$

By the derivation in section 2.1, we can think of this as coming from the control problem $V^m(x,T) = \sup_{u\in\mathcal{U}} J^m(x,T;u)$, where

$$J^m(x,T;u) := \int_0^T \mathcal{L}^m(\xi_t^m, u_t)dt \qquad \mathcal{L}^m(x,u) := x^T D^m x - \frac{\gamma^2}{2}|u|_2^2 + (l_1^m)^T x + \alpha^m$$

with $\dot{\xi}^m = A^m x + l_2^m + \sigma^m u$, $\xi_0^m = x$, and $\mathcal{U} := \{u : [0,\infty) \to \mathbb{R}^n \Big| \int_0^T |u_t|_2^2 dt < \infty \; \forall T < \infty\}$. Some specific assumptions are required to guarantee the existence of V^m. The following assumptions and solution existence results are discussed in more detail in [2][3][4]. Before stating these results we will need the following definition.

Definition 1. A function f is semiconvex if there exists a symmetric definite matrix β such that $f(x) + \frac{1}{2}x^T\beta x$ is convex over $\mathbb{R}^n$.

It is worth noting that the standard definition of semiconvexity forces β to be a scaling of the identity matrix. The definition we are using allows some future results to be slightly more general.

Assumptions:

1. $\exists c_A^m > 0$ such that $x^T A^m x \leq -c_A|x|_2^2$ for all $x \in \mathbb{R}^n$
2. Letting $c_\sigma^m = ||\sigma^m||$(operator norm), we require $\frac{\gamma^2}{(c_\sigma^m)^2} > \frac{\lambda_{max}(D^m)}{(c_A^m)^2}$

If these assumptions are satisfied, then there exists a steady state solution $V^m(x)$ satisfying

$$0 = H^m(x, \nabla V^m)$$

Furthermore, this solution is the unique continuous solution to $S_\tau^m[V^m] = V^m$ in the set

$$G_\delta^m := \left\{V : \mathbb{R}^n \to [0,\infty) \middle| V \text{ is semiconvex and } V(x) \leq \frac{c_A^m(\gamma-\delta)^2}{(c_\sigma^m)^2}|x|_2^2\right\}$$

for sufficiently small $\delta > 0$. It is also known that for any $\phi \in G_\delta^m$, $V^m(x) = \lim_{\tau\to\infty} S_\tau^m[\phi](x)$. This means that we can obtain an approximate this solution by starting with an initial guess in the set G_δ^m and applying S_τ^m a large number of times. Conveniently, when ϕ is a quadratic given by $\phi(x) = \frac{1}{2}(x-z)^T M(x-z) + r$, then

$$S_\tau^m[\phi](x) = \frac{1}{2}\left[(x - \Lambda_\tau^m z)^T P_\tau^m (x - \Lambda_\tau^m z) + z^T R_\tau^m z + 2(L_\tau^m)^T x + 2(K_\tau^m)^T z + b_\tau^m\right]$$

where

$$\begin{aligned}
\dot{P}^m &= (A^m)^T P^m + P^m A^m + D^m + P^m \Sigma^m P^m & P_0^m &= M \in \mathbb{R}^{n\times n}\\
\dot{\Lambda}^m &= \left[(P^m)^{-1} D^m + A^m\right]\Lambda^m & \Lambda_0^m &= I \in \mathbb{R}^{n\times n}\\
\dot{R}^m &= (\Lambda^m)^T D^m \Lambda^m & R_0^m &= 0 \in \mathbb{R}^{n\times n}\\
\dot{L}^m &= \left[P^m\Sigma^m + (A^m)^T\right]L^m + l_1^m + P^m l_2^m & L_0^m &= 0 \in \mathbb{R}^n\\
\dot{K}^m &= (\Lambda^m)^T\left[P^m l_2^m + P^m\Sigma^m L^m\right] & K_0^m &= 0 \in \mathbb{R}^n\\
\dot{b}^m &= \alpha^m + (L^m)^T \Sigma^m L^m + 2(L^m)^T l_2^m & b_0^m &= r \in \mathbb{R}
\end{aligned}$$

and the initial conditions are chosen to give $S_0^m[\phi](x) = \phi(x)$. See [2] and 8.2 in [6] for more details. This system of ODEs can be efficiently approximated with a Runge-Kutta algorithm.

As mentioned in the introduction, we will approximate Hamiltonians with a point-wise max of quadratics. These quadratics will take the form of the H^m's mentioned in this section. More specifically we will approximate a given H with

$$H(x,p) \approx \widetilde{H}(x,p) := \max_{m\in\mathcal{M}} H^m(x,p)$$

for some $\mathcal{M} := \{1, 2, ..., M\}$, $M \in \mathbb{N}$. We will also require $\widetilde{H} \leq H$, which will result in $\widetilde{V} \leq V$. If V models a problem of maximizing some cost, then compared to an over approximation, an under approximation of V will be closer to(or equal to) the cost resulting from some feasible choice of control. When we refer to the value functions as V,$\widetilde{V}$ or $V(x), \widetilde{V}(x)$, we are referring to the steady state solutions. Assumptions that guarentee the existence of $\widetilde{V}$ will be given in section 5. Before getting into this, we will give a method of creating quadratic functions H^m that guarantees $H^m \leq H$ on some convex compact subset of $\mathbb{R}^n$.

3 Quadratic under approximations

The goal of this section is to derive an efficient method for generating quadratic functions H^m which are under approximations of H. For notational ease we will write $H(z)$ where $z = [x, p]^T$. This is also to emphasize that the method in this section can be used on any function that satisfies the assumptions in theorems 3 and 4. For a given $z_m \in K$, we will first derive an analytic expression for a constant λ^m such that $T_2^m(z) + \frac{\lambda^m}{2}|z - z_m|_2^2 \leq H(z)$ for all $z \in K$, where T_2^m is the second order Taylor approximation of H centered at z_m. We then describe a way to efficiently compute a λ_1^m such that $\lambda_1^m \leq \lambda^m$, which allows us to generate the quadratics $H^m(z) = T_2^m(z) + \frac{\lambda_1^m}{2}|z - z_m|_2^2 \leq H(z)$. We also give an extension to Hamiltonian's that are not C^2 everywhere(under certain additional conditions). Section 3.2 and theorem 3 in section 3.1 come from [7].

3.1 Quadratic upper bound

Let $K \subset \mathbb{R}^n$ be compact and convex and consider $H \in C^2(K)$. Let $T_2^m(z)$ be the second degree Taylor approximation of H centered at $z_m \in K$. We will derive an analytic expression for a constant λ^m such that $T_2^m(z) + \frac{\lambda^m}{2}|z - z_m|_2^2 \leq H(z)$ for all $z \in K$. After this we will present a way to efficiently compute $\lambda_1^m \leq \lambda^m$ for any given z_m, which allows us to form the quadratics $T_2^m(z) + \frac{\lambda_1^m}{2}|z - z_m|_2^2 \leq H(z)$. We also give an extension to functions that are not C^2 everywhere(under certain additional conditions). Section 3.2 and theorem 3 in section 3.1 come from [7].

3.2 Quadratic lower bound:

First we will setup some objects. Denote the vector space of real symmetric $n \times n$ matrices as S_n, and let $\lambda_{min} : S_n \to \mathbb{R}$ give the minimal eigenvalue of a matrix in S_n. Similarly λ_{max} is defined to give the max eigenvalue of a matrix in S_n

Lemma 2. *Properties of the λ_{min} function.*

As an unlabeled note, it is easy to show that for all $A \in S_n$ and $c \geq 0$,

$$\lambda_{min}(cA) = c\lambda_{min}(A) = -c\lambda_{max}(-A) = -\lambda_{max}(-cA)$$

Property (a): $\forall A, B \in S_n$, $\lambda_{min}(A + B) \geq \lambda_{min}(A) + \lambda_{min}(B)$

Proof: Recall that $\forall C \in S_n$ and $u \in \mathbb{R}^n \backslash \{0\}$, $\lambda_{min}(C) \leq \frac{u^T C u}{u^T u}$

Now let $u \in \mathbb{R}^n \backslash \{0\}$ be such that $(A+B)u = \lambda_{min}(A+B)u$, we have

$$\lambda_{min}(A+B) = \frac{u^T(A+B)u}{u^T u} = \frac{u^T A u}{u^T u} + \frac{u^T B u}{u^T u} \geq \lambda_{min}(A) + \lambda_{min}(B) \qquad \square$$

Property (b): λ_{min} is continuous from $(S_n, |\cdot|_\infty)$ to $(\mathbb{R}, |\cdot|)$, where $|A|_\infty := \max\limits_{1 \leq i,j \leq n}\{|A_{ij}|\}$

Proof: Let $\epsilon > 0$ and assume $A, B \in S_n$ with $|A - B|_\infty < \frac{\epsilon}{n}$

By (a) we have

$$\lambda_{min}(A) - \lambda_{min}(B) = \lambda_{min}(A) - \lambda_{min}(B - A + A) \leq -\lambda_{min}(B - A) = \lambda_{max}(A - B)$$

$$\lambda_{min}(B) - \lambda_{min}(A) = \lambda_{min}(B) - \lambda_{min}(A - B + B) \leq -\lambda_{min}(A - B) = \lambda_{max}(B - A)$$

Hence $|\lambda_{min}(A) - \lambda_{min}(B)| \leq \max\{\lambda_{max}(A-B), \lambda_{max}(B-A)\}$

Let u be a unit eigenvector of $A - B$ associated to eigenvalue $\lambda_{max}(A - B)$

For $i = \arg\max\limits_{1 \leq j \leq n}\{|u_j|\}$ we have

$$\lambda_{max}(A-B) = \sum_{j=1}^{n} \frac{[A-B]_{ij}u_j}{u_i} \leq \sum_{j=1}^{n} |\frac{[A-B]_{ij}u_j}{u_i}| \leq \sum_{j=1}^{n} |[A-B]_{ij}| < \sum_{j=1}^{n} \frac{\epsilon}{n} = \epsilon$$

The same argument applies to $\lambda_{max}(B - A)$, therefore $|\lambda_{min}(A) - \lambda_{min}(B)| < \epsilon$ and λ_{min} is continuous $\quad \square$

Now we will begin deriving the quadratic lower bound. Define $\lambda^m := \min\limits_{z \in K}\{\lambda_{min}(D^2H(z) - D^2H(z_m))\}$, this is well defined because of Lemma 1(b) and the fact that H is C^2 and K is compact.

Theorem 3. *If H is a C^2 function from $K \to \mathbb{R}$ and $\lambda_1^m \leq \lambda^m$, then the quadratic function*

$$H^m(z) = H(z_m) + \nabla H(z_m)^T(z - z_m) + \frac{1}{2}(z - z_m)^T(D^2H(z_m) + \lambda_1^m I)(z - z_m)$$

is an under approximation of H on K

Proof: Let $z \in \mathbb{R}^n$ and consider the first degree Taylor expansion of H centered at z_m,

$$T_1(z) = H(z_m) + \nabla H(z_m)(z - z_m)$$

We know that for some $\xi \in \{z + t(z_m - z) : t \in [0,1]\} \subset K$, this has remainder

$$R_1(z) = \frac{1}{2}(z - z_m)^T D^2H(\xi)(z - z_m)$$

By the definition of λ^m and the fact that $D^2H(\xi) - D^2H(z_m)$ is symmetric, we know that $\forall z \in K$,

$$\lambda^m |z - z_m|_2^2 \le \lambda_{min}(D^2H(\xi) - D^2H(z_m))|z - z_m|_2^2 \le (z - z_m)^T (D^2H(\xi) - D^2H(z_m))(z - z_m)$$

Hence for $\lambda_1^m \le \lambda^m$, we have

$$\begin{aligned} H(z) &= T_1(z) + R_1(z) \\ &= T_1(z) + \frac{1}{2}(z - z_m)^T D^2H(z_m)(z - z_m) + \frac{1}{2}(z - z_m)^T (D^2H(\xi) - D^2H(z_m))(z - z_m) \\ &\ge T_1(z) + \frac{1}{2}(z - z_m)^T D^2H(z_m)(z - z_m) + \frac{1}{2}\lambda_1^m |z - z_m|_2^2 = H^m(z) \end{aligned}$$

This holds for all $z \in K$, therefore we have proved the result. $\square$

The following theorem gives an extension to functions that are not C^2 on all of K. This is not given in [7] and to our knowledge is a new extension.

Theorem 4. *If H is continuous, C^1 everywhere but a finite or empty set, convex in some ϵ ball centered at each point it is not C^1, C^2 everywhere but a countable or empty set, and twice differentiable at $z_m \in K$, then $H^m(z) \le H(z)$ for all $z \in K$.*

Proof: The proof will be split into three sections. (i) Will construct the functions we need for the proof, (ii) will give the technical parts of the proof, and (iii) will conclude that $H(y) \ge H^m(y)$ for all $y \in K$.

(i) Let H be C^2 at z_m and consider $y \in K$. If $y = z_m$, then by definition $H(y) = H^m(y)$. Otherwise, set $t_y := |y - z_m|_2$ and define the following for $t \in [0, t_y]$,

$$\gamma(t) := \frac{y - z_m}{|y - z_m|_2} t + z_m \qquad h(t) := H(\gamma(t)) \qquad \rho(t) = H^m(\gamma(t))$$

By construction we have that

$$h(t_y) = H(y) \qquad \rho(t_y) = H^m(y)$$

$$h(0) = \rho(0) \qquad h'(0) := \rho'(0)$$

Furthermore, for all t where $H(z)$ is C^2 at $z = \gamma(t)$,

$$h''(t) = h''(0) + (\gamma(t) - z_m)^T (D^2H(\gamma(t)) - D^2H(z_m))(\gamma(t) - z_m) \ge h''(0) + \lambda^m = \rho''(t)$$

(ii) This part of the proof will show that for all $T \in (0, t_y)$ where h is C^1, $h'(T) \ge \rho'(T)$. Let $T \in (0, t_y)$ be a point where h is C^1, define $S = \{t \in (0, T) : h \text{ is not } C^1 \text{ at } t\}$.

If $S = \emptyset$, recall that $h''(t) \geq \rho''(t)$ for all t where $H(z)$ is C^2 at $z = \gamma(t)$(all but a countable set or empty). Since $h'(0) = \rho'(0)$, one can use the FTC to prove $h'(T) \geq \rho'(T)$.

If $S \neq \emptyset$, then by assumption S is finite. Write $S = \{t_1, ..., t_n\}$ where $0 < t_1 < ... < t_n < T$.

Set $t_0 = 0$ and $t_{n+1} = T$, let $\zeta_i : [t_i, t_{i+1}] \to \mathbb{R}$ be defined by $\zeta_i(t) = \begin{cases} \lim\limits_{t \downarrow t_i} h'(t) & \text{if } t = t_i \\ h'(t) & \text{if } t \in (t_i, t_{i+1}) \\ \lim\limits_{t \uparrow t_{i+1}} h'(t) & \text{if } t = t_{i+1} \end{cases}$

By construction each ζ_i is continuous, and C^1 everywhere but a countable set, hence

$$h'(T) = \zeta_n(t_{n+1}) = \zeta_0(t_1) + \sum_{i=1}^{n} \Big[\zeta_i(t_{i+1}) - \zeta_{i-1}(t_i)\Big]$$

$$= \zeta_0(0) + \zeta_0(t_1) - \zeta_0(0) + \sum_{i=1}^{n} \Big[\zeta_i(t_{i+1}) - \zeta_i(t_i) + \zeta_i(t_i) - \zeta_{i-1}(t_i)\Big]$$

$$= \zeta_0(0) + \int_0^{t_1} \zeta_0'(t)dt + \sum_{i=1}^{n} \Big[\int_{t_i}^{t_{i+1}} \zeta_i'(t)dt + \zeta_i(t_i) - \zeta_{i-1}(t_i)\Big]$$

By the assumption that H is locally convex everywhere it is not C^1(a finite superset of $\gamma(S)$), we have that for all $1 \leq i \leq n$,

$$\zeta_i(t_i) = \lim_{t \downarrow t_i} h'(t) \geq \lim_{t \uparrow t_i} h'(t) = \zeta_{i-1}(t_i)$$

Hence,

$$h'(T) \geq \zeta_0(0) + \int_0^{t_0} \zeta_0'(t)dt + \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} \zeta_i'(t)dt$$

$$= h'(0) + \int_0^{t_0} h''(t)dt + \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} h''(t)dt$$

$$= h'(0) + \int_0^{T} h''(t)dt \geq \rho'(0) + \int_0^{T} \rho''(t)dt = \rho'(T)$$

(iii) Now that we have proved $h'(T) \geq \rho'(T)$ for all $T \in (0, t_y)$ where h is C^1(everywhere but a finite or empty set), we can simply use the FTC to find

$$H(y) = h(t_y) = h(0) + \int_0^{t_y} h'(t)dt \geq \rho(0) + \int_0^{t_y} \rho'(t)dt = \rho(t_y) = H^m(y)$$

We chose y to be arbitrary in (i), therefore $H(y) \geq H^m(y)$ for all $y \in K$ $\square$

3.3 Efficient computation

For an $n \times n$ interval matrix A with compact entries $A_{ij} = A_{ji} = [a_{ij}, b_{ij}] \subset \mathbb{R}$, define $\mathcal{R}(A) := \{B \in S_n : B_{ij} \in A_{ij}\}$, we note that this set is compact. We will slightly abuse notation and let $\lambda_{min}(A) := \min_{B \in \mathcal{R}(A)} \{\lambda_{min}(B)\}$, which exists due to the continuity of λ_{min} and compactness of entries in A.

Now define the interval matrix $\widetilde{D^2H(z_m)}$ where $[\widetilde{D^2H(z_m)}]_{ij} := [\underline{d_{ij}} - d_{ij}, \overline{d_{ij}} - d_{ij}]$ and

$$\overline{d_{ij}} \geq \sup_{z \in K} \left\{[D^2H(z)]_{ij}\right\} \qquad \underline{d_{ij}} \leq \inf_{z \in K} \left\{[D^2H(z)]_{ij}\right\} \qquad d_{ij} := [D^2H(z_m)]_{ij}$$

Ideally these inequalities should be close estimates, as it will give us a better approximation of H. Notice $\forall z \in K$, $D^2H(z) - D^2H(z_m) \in \mathcal{R}(\widetilde{D^2H(z_m)})$, which implies $\lambda_{min}(\widetilde{D^2H(z_m)}) \leq \lambda^m$ We will now express the matrix $\widetilde{D^2H(z_m)}$ as $\frac{X+Y}{2}$ with

$$X_{i,j} := [\overline{d_{i,j}} + \underline{d_{i,j}} - 2d_{ij}, \overline{d_{i,j}} + \underline{d_{i,j}} - 2d_{ij}] \qquad Y_{i,j} := [\underline{d_{i,j}} - \overline{d_{i,j}}, \overline{d_{i,j}} - \underline{d_{i,j}}]$$

Choose $M \in \mathcal{R}(X+Y)$ such that $\lambda_{min}(X+Y) = \lambda_{min}(M)$, we know $\exists M_1 \in \mathcal{R}(X)$ and $M_2 \in \mathcal{R}(Y)$ such that $M = M_1 + M_2$, hence by lemma 1(a)

$$\lambda_{min}(X+Y) = \lambda_{min}(M) \geq \lambda_{min}(M_1) + \lambda_{min}(M_2) \geq \lambda_{min}(X) + \lambda_{min}(Y)$$

With a nearly identical argument to what was used in the proof of lemma 1(b), one can show

$$\lambda_{min}(Y) \geq -\max_{1 \leq i \leq n} \{\sum_{j=1}^{n} \max_{y \in Y_{ij}} |y|\} = -\max_{1 \leq i \leq n} \{\sum_{j=1}^{n} (\overline{d_{ij}} - \underline{d_{ij}})\}$$

Putting these results together we have

$$\lambda^m \geq \lambda_{min}(\widetilde{D^2H(z_m)}) = \frac{1}{2}\lambda_{min}(X+Y) \geq \frac{1}{2}\Big[\lambda_{min}(X) - \max_{1 \leq i \leq n} \Big\{ \sum_{j=1}^{n} (\overline{d_{i,j}} - \underline{d_{i,j}})\Big\}\Big]$$

So, letting $\lambda_2^m = \frac{1}{2}\Big[\lambda_{min}(X) - \max_{1 \leq i \leq n} \Big\{ \sum_{j=1}^{n} (\overline{d_{i,j}} - \underline{d_{i,j}})\Big\}\Big]$, we obtain $\lambda_2^m \leq \lambda^m$. This can be easily computed by inputting matrices containing coarse lower and upper bounds for the entries in D^2H, call these $\underline{D^2H}$ and $\overline{D^2H}$. Then

$$\lambda_{min}(X) = \lambda_{min}(\overline{D^2H} + \underline{D^2H} - 2D^2H(z_m))$$

and $\max_{1 \leq i \leq n} \Big\{ \sum_{j=1}^{n} (\overline{d_{i,j}} - \underline{d_{i,j}})\Big\}$ is the max row sums of $\overline{D^2H} - \underline{D^2H}$.

3.4 Converting quadratics to the desired form

As discussed in Section 2.3, we would like our quadratic Hamiltonians to be of the form

$$H^m(x,p) = \frac{1}{2}x^T D^m x + \frac{1}{2}p^T \Sigma^m p + p^T A^m x + (l_1^m)^T x + (l_2^m)^T p + \alpha^m$$

The quadratics generated from the method in section 3 are of the form

$$H^m(z) = H(z_m) + \nabla H(z_m)^T (z - z_m) + \frac{1}{2}(z - z_m)^T (D^2 H(z_m) + \lambda_2^m I)(z - z_m)$$

Where $z = \begin{bmatrix} x \\ p \end{bmatrix}$ and $z_m = \begin{bmatrix} x_m \\ p_m \end{bmatrix}$. To transform to the desired form, write

$$D^2 H(z_m) + \lambda_2^m I = \begin{bmatrix} Q_1^m & Q_2^m \\ Q_3^m & Q_4^m \end{bmatrix} \qquad \nabla H(z_m) = \begin{bmatrix} G_1^m \\ G_2^m \end{bmatrix}$$

Grouping the various quadratic, linear and constant terms, we find

$$D^m = Q_1^m \qquad \Sigma^m = Q_4^m \qquad A^m = Q_3^m$$

$$l_1^m = G_1^m - Q_1^m x_m - Q_2^m p_m \qquad l_2^m = G_2^m - Q_4^m p_m - Q_3^m x_m$$

$$\alpha^m = H(z_m) - \nabla H(z_m)^T z_m + \frac{1}{2} z_m^T (D^2 H(z_m) + \lambda_2^m I) z_m$$

4 Max-plus algebra

4.1 Intro and algebraic objects

Let $\mathbb{R}^- := \mathbb{R} \cup \{-\infty\}$, we will define the following operations $\oplus, \otimes : \mathbb{R}^- \times \mathbb{R}^- \to \mathbb{R}^-$, where

$$x \oplus y = \max\{x, y\} \qquad x \otimes y = x + y$$

It is relatively straightforward to show that $(\mathbb{R}^-, \oplus, \otimes)$ is a commutative semiring with unity 0(and additive identity $-\infty$). This is a semiring and not a ring because real numbers do not have inverses under $\oplus$. It is also idempotent, meaning $a \oplus a = a$ for all $a \in \mathbb{R}^-$. Due to the similarities between the structures of $(\mathbb{R}^-, \oplus, \otimes)$ and $(\mathbb{R}, +, \times)$, many algebraic results from standard arithmetic have analogous results in max-plus algebra. An important result that we will use in section 5.3 is that, for symmetric definite $\beta \in \mathbb{R}^{n \times n}$,

$$\mathcal{S}_\beta := \{f : \mathbb{R}^n \to \mathbb{R} \big| f(x) + \frac{1}{2}x^T \beta x \text{ convex over } \mathbb{R}^n\}$$

is a max-plus vector space. This means that

$$a \otimes f \oplus b \otimes g \in \mathcal{S}_\beta \qquad \forall a, b \in \mathbb{R}^- \text{ and } f, g \in \mathcal{S}_\beta$$

4.2 Additional notation

For some $m \in \mathbb{N}$ and $\{x_1, ..., x_m\} \subset \mathbb{R}^-$, we define

$$\bigoplus_{i=1}^{m} x_i = x_1 \oplus ... \oplus x_m = \max\{x_1, ..., x_m\}$$

This is of course the analogous max-plus summation notation, and it follows from distributivity that $\forall \lambda \in \mathbb{R}^-$, $\bigoplus_{i=1}^{m} \left[\lambda \otimes x_i\right] = \lambda \otimes \bigoplus_{i=1}^{m} x_i$

Similarly we can define the analogous max-plus integral over some $D \subset \mathbb{R}^n$. We will not go into the details of max-plus integral existence, since it will not be needed. We will need the fact that if f is continuous and bounded above, then $\int_D^{\oplus} f(x)dx$ exists over any convex $D \subset \mathbb{R}^n$ and

$$\int_D^{\oplus} f(x)dx = \sup_{x \in D} f(x)$$

In the simple case of continuous $f : [0,1] \to \mathbb{R}$, the analogy to standard algebra can be seen by formulating the Riemann integral with max-plus operations, that is

$$\int_{[0,1]}^{\oplus} f(x)dx = \lim_{k\to\infty} \left[\frac{1}{k} \otimes \bigoplus_{i=1}^{k} f(\frac{i}{k})\right] = \lim_{k\to\infty} \left[\frac{1}{k} + \max\{f(\frac{1}{k}), f(\frac{2}{k}), ..., f(1)\}\right]$$

$$= \lim_{k\to\infty} \max\{f(\frac{1}{k}), f(\frac{2}{k}), ..., f(1)\}$$

Which by the continuity of f is equal to $\sup_{x\in[0,1]} f(x)$.

The max-plus integral is also max-plus linear, so if $f_i : D \to \mathbb{R}^-$ and $\lambda_i \in \mathbb{R}^-$ for $1 \leq i \leq k$, then

$$\int_D^{\oplus} \bigoplus_{i=1}^{k} \lambda_i \otimes f_i(x)dx = \bigoplus_{i=1}^{k} \lambda_i \otimes \int_D^{\oplus} f_i(x)dx$$

The last piece of notation we will use is the analogous max-plus convolution, given as

$$f \odot g = \int_{\mathbb{R}^n}^{\oplus} f(x) \otimes g(x)dx$$

Being defined by an integral, it is clear that the max-plus convolution is max-plus linear in both arguments. It is also important to note that we can express the Legendre-Fenchel transform as a max-plus convolution. Recall that the Legendre-Fenchel transform of a function $f(x)$ is defined as $f^*(y) := \sup_{x\in\mathbb{R}^n} \{\langle x, y\rangle - f(x)\}$. If we let $\psi(x,y) = \langle x, y\rangle$, then

$$f^*(y) = \int_{\mathbb{R}^n}^{\oplus} \psi(x,y) \otimes [-f(x)]dx = \psi(\cdot, y) \odot [-f(\cdot)]$$

4.3 Max-plus linearity of semigroup operators

An important property of the semigroup operators described in section 2.2 is that they are max-plus linear. As a specific example consider the semigroup operator associated to the linear quadratic control problem in section 2.3, if $\phi(x) = \bigoplus_{i=1}^{N} \lambda_i \otimes \phi_i(x)$ for some $N \in \mathbb{N}$ and $\lambda_i \in \mathbb{R}^-$, then

$$S_\tau^m[\phi](x) = \sup_{u \in \mathcal{U}} \left[J^m(x, \tau; u) \otimes \bigoplus_{i=1}^{N} \lambda_i \otimes \phi_i(\xi_\tau^m) \right]$$

$$= \bigoplus_{i=1}^{N} \lambda_i \otimes \sup_{u \in \mathcal{U}} \left[J^m(x, \tau; u) \otimes \phi_i(\xi_\tau^m) \right] = \bigoplus_{i=1}^{N} \lambda_i \otimes S_\tau^m[\phi_i](x)$$

5 The algorithm

First we will formulate the control problem associated to the approximate Hamiltonian and its HJB PDE, which is

$$\frac{\partial \widetilde{V}}{\partial T} = \widetilde{H}(x, \nabla \widetilde{V}) := \bigoplus_{m \in \mathcal{M}} H^m(x, \nabla \widetilde{V}) \qquad \widetilde{V}(x, 0) = 0$$

where $\mathcal{M} := \{1, ..., M\}$ for some $M \in \mathbb{N}$. We will then discretize this problems associated semigroup operator and use this discretized operator to progress an initial solution guess towards the steady state. This would generally involve solving an exponential amount of differential equations, but we will avoid this using semiconvex duality and the max-plus linearity of semigroup operators. Much of this section comes from [2], however we offer relaxed assumptions for the hypothesis of theorem 8 to hold.

5.1 Control problem associated to approximate Hamiltonian

Combining the results in sections 2.1 and 2.3, we can think of $\widetilde{V}$ as

$$\widetilde{V}(x, T) = \sup_{w \in \mathcal{W}} \sup_{u \in \mathcal{U}} \int_0^T \mathcal{L}^{w_t}(\xi_t^{w_t}, u_t) dt \qquad \mathcal{W} := \{w : [0, \infty) \to \mathcal{M} \big| w \text{ measurable}\}$$

Where $\mathcal{U}$ is the same set given in section 2.3, and the control w determines which choice of $m \in \mathcal{M}$ is used as the system dynamics and running cost at a given point in time. ξ^w is the trajectory the systems takes when using the controls w and u with $\xi_0^w = x$. When ξ^w is being integrated we will refer to it as $\xi_t^{w_t}$ to emphasize that if $w_t = m$, then the dynamics are those of ξ_t^m from 2.3. Otherwise we will write ξ_t^w to refer to the state of ξ^w at time t. To guarantee the existence of $\widetilde{V}$ we will need to make a few more assumptions. We will not go into detail on why these assumptions are required as it is out of scope. In this regard, we refer the reader to [2], [3], [4].

Assumptions:

1. Assume the assumptions from 2.3 hold for all $m \in \mathcal{M}$, note that γ should be the same for each H^m, this can be achieved without changing the Hamiltonians by rescaling each σ^m.

2. Assume $\frac{\gamma^2}{c_\sigma^2} > \frac{c_D}{c_A}$, where

$$c_A := \min_{m\in\mathcal{M}} c_A^m \qquad c_D := \max_{m\in\mathcal{M}} \lambda_{min}(D^m) \qquad c_\sigma := \max_{m\in\mathcal{M}} = c_\sigma^m$$

Analogous to the result in 2.3, if these assumptions are satisfied then there exists a unique steady state viscosity solution in

$$G_\delta := \left\{V : \mathbb{R}^n \to [0,\infty) \middle| V \text{ is semiconvex and } V(x) \le \frac{c_A(\gamma-\delta)^2}{c_\sigma^2}|x|_2^2\right\}$$

for sufficiently small $\delta > 0$. Furthermore, for any $\phi \in G_\delta$, this steady state is equal to $\lim_{\tau\to\infty} \widetilde{S}_\tau[\phi](x)$ where

$$\widetilde{S}_\tau[\phi](x) := \sup_{w\in\mathcal{W}} \sup_{u\in\mathcal{U}} \left\{ \int_0^\tau \mathcal{L}^{w_t}(\xi_t^{w_t}, u_t)dt + \phi(\xi_\tau^w)\right\}$$

Computing this operator applied to some function ϕ is difficult because a control w can rapidly switch the system dynamics and running cost. To make computing this feasible, we will approximate the semigroup operator by discretizing the variable w in time. The next section will focus on this topic.

5.2 Time Discretization

Fix $\tau > 0$ and define

$$\Omega_\tau := \{\omega : [0,\infty) \to \mathcal{M} | \forall n \in \mathbb{N}, \omega_t \text{ is constant on } [(n-1)\tau, n\tau]\}$$

For a given $j \in \mathbb{N}$, we approximate $\widetilde{S}_{j\tau}[\phi](x)$ with

$$\overline{S}_{\tau,j}[\phi](x) := \sup_{\omega\in\Omega_\tau} \sup_{u\in\mathcal{U}} \left[\int_0^{j\tau} \mathcal{L}^{\omega_t}(\xi_t^{\omega_t}, u_t)dt + \phi(\xi_{j\tau}^\omega)\right]$$

This means that over each time interval of length τ, we force the system dynamics and running cost to follow those of one specific constituent linear quadratic control problem. With a similar approach to how we proved $S_\tau^m[V^m(\cdot,T)](x) = V^m(x,T+\tau)$(we showed this for a more general problem in 2.2), one can show that parameterized by j, the operators $\overline{S}_{\tau,j}$ form a semigroup, in other words

$$\overline{S}_{\tau,j_1} \circ \overline{S}_{\tau,j_2} = \overline{S}_{\tau,j_1+j_2} \qquad \forall j_1, j_2 \in \mathbb{N}$$

Furthermore, by definition we have

$$\overline{S}_{\tau,1}[\phi](x) = \max_{m\in\mathcal{M}} \sup_{u\in\mathcal{U}} \Big[\int_0^\tau \mathcal{L}^m(\xi_t^m, u_t)dt + \phi(\xi_\tau^m) \Big] = \bigoplus_{m\in\mathcal{M}} S_\tau^m[\phi](x)$$

From 4.3 we know each S_τ^m is max-plus linear, therefore the above implies $\overline{S}_{\tau,1}$ is max-plus linear and

$$\overline{S}_{\tau,2}[\phi](x) = (\overline{S}_{\tau,1} \circ \overline{S}_{\tau,1})[\phi](x) = \bigoplus_{\underline{m}\in\mathcal{M}^2} (S_\tau^{\underline{m}_2} \circ S_\tau^{\underline{m}_1})[\phi](x)$$

More generally, if $j \in \mathbb{N}$ and $\phi(x) = \bigoplus_{k=1}^{K} \phi_k(x)$, then

$$\overline{S}_{\tau,j}[\phi](x) = \Big(\prod_{i=1}^{j} \overline{S}_{\tau,1}\Big)[\phi](x) = \bigoplus_{\underline{m}\in\mathcal{M}^j} \Big(\prod_{i=1}^{j} S_\tau^{\underline{m}_i}\Big)[\phi](x) = \bigoplus_{k=1}^{K} \bigoplus_{\underline{m}\in\mathcal{M}^j} \Big(\prod_{i=1}^{j} S_\tau^{\underline{m}_i}\Big)[\phi_k](x)$$

where $\prod$ represents the composition of operators.

With an initial quadratic guess $\phi \in G_\delta$ and large $J \in \mathbb{N}$, we could approximate the steady state solution $\widetilde{V}(x)$ with this form of $\overline{S}_{\tau,J}[\phi](x)$. However, this would involve solving $\sum_{i=1}^{J} M^i$ ODEs up to time τ, which is certainly not efficient for large J and $\#\mathcal{M}$. We will avoid this using semiconvex duality, the subject of the next section. We will note that by construction $\Omega_\tau \subset \mathcal{W}$, which implies that $\overline{S}_{\tau,j}[\phi](x) \leq \widetilde{S}_{j\tau}[\phi](x)$ for all ϕ, x. Furthermore, it is shown in [2] that our discretized problem has a steady state $\overline{V}(x) \in G_\delta$ such that for any $\phi \in G_\delta$, $\overline{V}(x) = \lim_{j\to\infty} \overline{S}_{\tau,j}[\phi](x)$. Recalling that the steady state solution $\widetilde{V}(x)$ is equal to $\lim_{j\to\infty} \widetilde{S}_{j\tau}[\phi](x)$ for any $\phi \in G_\delta$, it follows that $\overline{V}(x) \leq \widetilde{V}(x)$. This means that in using the discretized semigroup operator, we are not losing the property that the solution to our approximate problem is less than or equal the solution to the original problem.

5.3 Semiconvex duality

Theorem 5. *Let C be a symmetric definite matrix and define*

$$\psi(x,y) := \frac{1}{2}(x-y)^T C(x-y)$$

If $\exists$ symmetric definite β such that $\phi \in \mathcal{S}_\beta$ and $\beta + C \prec 0$, then $\forall x \in \mathbb{R}^n$,

$$\phi(x) = \int_{\mathbb{R}^n}^{\oplus} \psi(x,y) \otimes a(y) dy = \psi(x,\cdot) \odot a(\cdot)$$

where

$$a(y) := -\int_{\mathbb{R}^n}^{\oplus} \psi(x,y) \otimes [-\phi(x)] dx = -\big[\psi(\cdot,y) \odot [-\phi(\cdot)]\big]$$

We call a the semiconvex dual of ϕ with respect to ψ. For us, all semiconvex duals will be with respect to ψ so we will simply refer to them as the semiconvex dual or the dual. The proof of this theorem is quite long, so as to not distract from the algorithm, we assume that convex duality through the Legendre-Fenchel transform is established. That is, for any function f, $f^*(y) := \sup_{x\in\mathbb{R}^n} \{\langle y, x\rangle - f(x)\}$ is convex, and $f^{**} = f$ iff f is convex. See [5] for more detail.

Proof: Assume $\phi \in \mathcal{S}_\beta$ with $\beta + C \prec 0$. Define $f(x) := \phi(x,y) - \frac{1}{2}x^T Cx$ and notice

$$f(x) = [\phi(x,y) + \frac{1}{2}x^T\beta x] - [\frac{1}{2}x^T\beta x + \frac{1}{2}x^T Cx]$$

is the sum of a convex function and a strictly convex function, and is therefore strictly convex(in x). Hence $f^{**} = f$ and,

$$a(y) = -\big[\psi(\cdot,y) \odot [-\phi(\cdot)]\big] = -\sup_{x\in\mathbb{R}^n}\{\psi(x,y) - \phi(x)\}$$

$$= -\sup_{x\in\mathbb{R}^n}\{\frac{1}{2}x^T Cx + \frac{1}{2}y^T Cy - y^T Cx - \phi(x)\}$$

$$= -\frac{1}{2}y^T Cy - \sup_{x\in\mathbb{R}^n}\{\langle -Cy, x\rangle - f(x)\} = -\frac{1}{2}y^T Cy - f^*(-Cy)$$

And

$$\psi(x,\cdot) \odot a(\cdot) = \sup_{y\in\mathbb{R}^n}\{\psi(x,y) - \frac{1}{2}y^T Cy - f^*(-Cy)\}$$

$$= \frac{1}{2}x^T Cx + \sup_{y\in\mathbb{R}^n}\{\langle x, -Cy\rangle - f^*(-Cy)\}$$

Since C is definite, it is full rank and this is equivalent to

$$\frac{1}{2}x^T Cx + \sup_{z\in\mathbb{R}^n}\{\langle x, z\rangle - f^*(z)\} = \frac{1}{2}x^T Cx + f^{**}(x) = \frac{1}{2}x^T Cx + f(x) = \phi(x) \quad \Box$$

It can also be shown that $-a \in \mathcal{S}_d$ for some d with $d + C \prec 0$, see [2] for proof.

Lemma 6. *If $\phi \in \mathcal{S}_\beta$ and $-a \in \mathcal{S}_d$ with $\beta + C \prec 0$ and $d + C \prec 0$, then $\phi(x) = \psi(x,\cdot) \odot a(\cdot)$ if and only if $a(y) = -\big[\psi(\cdot,y) \odot [-\phi(\cdot)]\big]$*

Proof: The if direction follows from the previous theorem. To prove the only if direction, assume $\phi(x) = \psi(x, \cdot) \odot a(\cdot)$ and consider the dual of $-a(y)$, which by the previous theorem is $-\big[\psi(x, \cdot) \odot a(\cdot)\big]$. By assumption this is equal to $-\phi$, therefore $-a(y) = \psi(\cdot, y) \odot [-\phi(\cdot)]$ $\square$
This establishes that the dual operation gives a one to one relationship between this class of semiconvex functions and semiconcave functions.

It can be proved(with a similar method to the theorem 8 proof in the appendix), that there exists a symmetric definite C such that $\overline{V}$,$\widetilde{V}$, and each V^m, are in some $\mathcal{S}_\beta$ with $\beta + C \prec 0$. In other words we can choose C so that the steady state solutions have a dual.

Lemma 7. *If $\phi_1, ...\phi_k \in \mathcal{S}_\beta$ have semiconvex duals $a_1, ..., a_k$, then $\bigoplus_{i=1}^{k} a_i$ is the dual of $\bigoplus_{i=1}^{k} \phi_i$*

Proof: Since $\mathcal{S}_\beta$ is a max-plus vector space, $\bigoplus_{i=1}^{k} \phi_i \in \mathcal{S}_\beta$. Therefore by the previous lemma, the following proves the result.

$$\bigoplus_{i=1}^{k} \phi_i(x) = \bigoplus_{i=1}^{k} \int_{\mathbb{R}^n}^{\oplus} \psi(x,y) \otimes a_i(y)dy = \int_{\mathbb{R}^n}^{\oplus} \psi(x,y) \otimes \big[\bigoplus_{i=1}^{k} a_i(y)\big]dy \quad \square$$

Theorem 8. *If $\phi \in \mathcal{S}_{-C}$ and $\min_{m\in\mathcal{M}} \lambda_{min}(D^m + (A^m)^T C + CA^m) > 0$, then for all $\tau > 0$ and $m \in \mathcal{M}$, $\exists s > 0$ such that $S_\tau^m[\phi] \in \mathcal{S}_{-(C+sI)}$ and $S_\tau^m[\phi]$ has a dual.*

Proof: The proof is quite long, so we include it in the appendix as an optional read. We note that the assumption required for this theorem to hold is slightly relaxed compared to what is given in [2]. Furthermore, we prove the statement for all $\tau > 0$, compared to [2] where it is proved for all τ in some non-constructive interval $(0, \overline{\tau})$.

If C is chosen so that the previous theorem is satisfied, then since $\psi(x, y) = \frac{1}{2}(x-y)^T C(x-y)$ is in $\mathcal{S}_{-C}$, we know that $S_\tau^m[\psi(\cdot, y)](x)$ has a unique semiconvex dual, which we denote

$$\mathcal{B}_\tau^m(y, z) := -\big[\psi(\cdot, z) \odot [-S_\tau^m[\psi(\cdot, y)](\cdot)]\big]$$

Theorem 9. *Define $\hat{\mathcal{B}}_\tau^m[a](z) := \mathcal{B}_\tau^m(\cdot, z) \odot a(\cdot)$, if $\phi \in \mathcal{S}_\beta$ has semiconvex dual a, then*

$$S_\tau^m[\phi](x) = \psi(x, \cdot) \odot \hat{\mathcal{B}}_\tau^m[a](\cdot)$$

Proof: Let $\phi(x) = \psi(x, \cdot) \odot a(\cdot)$, then

$$S_\tau^m[\phi](x) = \sup_{u\in\mathcal{U}} \Big\{ J^m(x, \tau; u) + \phi(\xi_\tau^m) \Big\} = \sup_{u\in\mathcal{U}} \Big\{ J^m(x, \tau; u) \otimes \int_{\mathbb{R}^n}^{\oplus} \psi(\xi_\tau^m, y) \otimes a(y)dy \Big\}$$

$$= \int_{\mathbb{R}^n}^{\oplus} \sup_{u\in\mathcal{U}} \left\{ J^m(x,\tau;u) \otimes \psi(\xi_\tau^m, y) \right\} \otimes a(y)dy = \int_{\mathbb{R}^n}^{\oplus} S_\tau^m[\psi(\cdot,y)](x) \otimes a(y)dy$$

$$= \int_{\mathbb{R}^n}^{\oplus} \int_{\mathbb{R}^n}^{\oplus} \psi(x,z) \otimes \mathcal{B}_\tau^m(y,z)dz \otimes a(y)dy$$

$$= \int_{\mathbb{R}^n}^{\oplus} \psi(x,z) \otimes \int_{\mathbb{R}^n}^{\oplus} \mathcal{B}_\tau^m(y,z) \otimes a(y)dydz = \int_{\mathbb{R}^n}^{\oplus} \psi(x,z) \otimes \hat{\mathcal{B}}_\tau^m[a](z)dz \quad \square$$

We note that this does not prove that $\hat{\mathcal{B}}_\tau^m[a]$ is the dual of $S_\tau^m[\phi]$(which we know exists by theorem 8). In the case of quadratic $\hat{\mathcal{B}}_\tau^m[a](\cdot)$, which follows from quadratic ϕ, it is not too difficult to show that $\hat{\mathcal{B}}_\tau^m[a]$ exists and is the dual of $S_\tau^m[\phi]$. Since we will be working with quadratics, this relationship will be assumed for the following statements and the remainder of the report. Since we are assuming that for all m, $\hat{\mathcal{B}}_\tau^m[a]$ is the dual of $S_\tau^m[\phi]$, theorem 9 implies that for all $m_1, m_2 \in \mathcal{M}$,

$$(\hat{\mathcal{B}}_\tau^{m_2} \circ \hat{\mathcal{B}}_\tau^{m_1})[a](z) \text{ is the semiconvex dual of } (S_\tau^{m_2} \circ S_\tau^{m_1})[\phi](x)$$

And simple induction argument then shows that for all $\underline{m} \in \mathcal{M}^j$,

$$\left(\prod_{i=1}^{j} \hat{\mathcal{B}}_\tau^{\underline{m}_i}\right)[a](z) \text{ is the semiconvex dual of } \left(\prod_{i=1}^{j} S_\tau^{\underline{m}_i}\right)[\phi](x)$$

By lemma 7, this implies

$$\bigoplus_{m\in\mathcal{M}^j} \left(\prod_{i=1}^{j} \hat{\mathcal{B}}_\tau^{\underline{m}_i}\right)[a](z) \text{ is the semiconvex dual of } \bigoplus_{\underline{m}\in\mathcal{M}^j} \left(\prod_{i=1}^{j} S_\tau^{\underline{m}_i}\right)[\phi](x) = \overline{S}_{\tau,j}[\phi](x)$$

This is beneficial because each $\mathcal{B}_\tau^m$ is a quadratic, and if ϕ is quadratic, then a is quadratic, which means solving for $\hat{\mathcal{B}}_\tau^m[a](z) = \max_{y\in\mathbb{R}^n}\{\mathcal{B}_\tau^m(y,z) + a(y)\}$ will amount to taking the max of a concave quadratic. The resulting function $\hat{\mathcal{B}}_\tau^m[a]$ will also be quadratic and successive applications of the dual operators can each be calculated through a set of matrix multiplications and two matrix inversions. These equations will be derived in the next section. Before that we will give a way of choosing of $C = cI$ so that theorem 8 is satisfied. The choice interval we give for c is a bit larger than what is given in [2].

Remark 10. If one chooses $C = cI$ with $c \neq 0$ and

$$-\min_{m\in\mathcal{M}} \frac{|\lambda_{min}(D^m)|}{|\lambda_{max}((A^m)^T + A^m)|} < c < \min_{m\in\mathcal{M}} \frac{|\lambda_{min}(D^m)|}{|\lambda_{min}((A^m)^T + A^m)|}$$

then $\min_{m\in\mathcal{M}} \lambda_{min}((A^m)^T C + CA^m + D^m) > 0$. We exclude the case of the min/max eigenvalue of $(A^m)^T + A^m$ equaling 0, because the assumption that $\lambda_{min}(D^m) > 0$ makes the desired result trivial.

Proof: Let $m \in \mathcal{M}$, using Lemma 1 from section 3.1 we have

$$\lambda_{min}(D^m + (A^m)^T C + CA^m) \geq \lambda_{min}(((A^m)^T + A^m)c) + \lambda_{min}(D^m)$$
$$\geq \lambda_{min}(D^m) - |\lambda_{min}(((A^m)^T + A^m)c)|$$

If $c > 0$ then the above is equal to

$$\lambda_{min}(D^m) - |\lambda_{min}(((A^m)^T + A^m)|c > \lambda_{min}(D^m) - |\lambda_{min}(D^m)| = 0$$

If $c < 0$ then it is equal to

$$\lambda_{min}(D^m) - |\lambda_{max}((A^m)^T + A^m)||c| > \lambda_{min}(D^m) - |\lambda_{min}(D^m)| = 0 \quad \square$$

We note that it is useful to choose $c < 0$, as it makes $\mathcal{S}_{-C}$ a larger space.

5.4 Steps for numerical implementation

1. Use the method described in sections 3.2 and 3.3 to obtain an approximation

$$H(x,p) \approx \widetilde{H}(x,p) = \bigoplus_{m \in \mathcal{M}} H^m(x,p)$$

   This is stored as the set of ordered sextets $\{(D^m, \Sigma^m, A^m, l_1^m, l_2^m, \alpha^m) : m \in \mathcal{M}\}$

2. Verify that the assumptions in 5.1 are satisfied. Choose $\delta > 0$ so that the steady state solutions $\widetilde{V}, \overline{V}$ are in G_δ, this should hold for each V^m as well.

3. Choose C such that the following assumptions are satisfied

   (a) $\min_{m \in \mathcal{M}} \lambda_{min}(D^m + (A^m)^T C + CA) > 0$. This is for theorem 8 to hold. Remark 10 can be used to achieve this.

   (b) There exists β with $\beta + C \prec 0$ such that $\overline{V} \in \mathcal{S}_\beta$, and each $V^m \in \mathcal{S}_\beta$.

   Also choose a time step $\tau > 0$ and a number of steps J.

4. Build an initial solution guess $\overline{V}^0$ defined as a max of quadratics in $G_\delta \cap \mathcal{S}_\beta$, denoted

$$\overline{V}^0(x) = \bigoplus_{k=1}^{K} \hat{V}_k^0(x) \quad \text{where} \quad \hat{V}_k^0(x) := \frac{1}{2}(x - x_k^0)^T G_k^0 (x - x_k^0) + g_k^0$$

   We store this as the set of triples $\mathcal{G}^0 := \{(G_k^0, x_k^0, g_k^0) : 1 \leq k \leq K\}$. The superscripts will denote the number of applications of $\overline{S}_{\tau,1}$ we have completed, in other words

$$\overline{V}^j(x) := \overline{S}_{\tau,j}[\overline{V}^0](x) = \bigoplus_{k=1}^{K} \bigoplus_{\underline{m} \in \mathcal{M}^j} \Big(\prod_{i=1}^{j} S_\tau^{m_i}\Big)[\hat{V}_k^j](x)$$

5. Compute the dual of $\overline{V}^0$, denoted $\overline{a}^0$. By lemma 7 this is given as

$$\overline{a}^0(z) = \bigoplus_{k=1}^{K} \hat{a}_k^0(z) \quad \text{where} \quad \hat{a}_k^0(z) = -\big[\psi(\cdot, z) \odot [-\hat{V}_k^0(\cdot)]\big]$$

This is easily computable through

$$\begin{aligned} \hat{a}_k^0(a) &= -\max_{x\in\mathbb{R}^n}\{\psi(x,z) - \hat{V}_k^0(x)\} \\ &= \frac{1}{2}\min_{x\in\mathbb{R}^n}\{(x-x_k^0)^T G_k^0(x-x_k^0) + 2g_k^0 - (x-z)^T C(x-z)\} \\ &= \frac{1}{2}(z-z_k^0)^T Q_k^0(z-z_k^0) + q_k^0 \end{aligned}$$

Where letting $F = (G_k^0 - C)^{-1}$,

$$Q_k^0 = -CFG_k^0 \qquad z_k^0 = -(Q_k^0)^{-1}CFG_k^0 x_k^0 \qquad q_k^0 = g_k^0 - \frac{1}{2}(z_k^0)^T Q_k^0 z_k^0 - \frac{1}{2}x_k^0 CFG x_k^0$$

We store the $\hat{a}_k^0$'s as the set of triples $\mathcal{Q}^0 := \{(Q_k^0, z_k^0, q_k^0) : 1 \le k \le K\}$. We will note a few tricks used to reduce the complexity of these formulas, as they were also used when deriving the next few formulas. First note that

$$\overline{x} := \arg\min_{x\in\mathbb{R}^n}\{\hat{V}_k^0(x) - \psi(x,z)\} = F(Gx_k^0 - Cz)$$

It is useful to re-arrange the terms in $\hat{V}_k^0(x) - \psi(x,z)$ before substituting $\overline{x}$ in. If we distribute and combine the quadratic x terms we find

$$\hat{V}_k^0(x) - \psi(x,z) = \frac{1}{2}\big[x^T F^{-1}x - z^T Cz - 2z^T Cx - 2(x_k^0)^T G_k^0 x + (x_k^0)^T G_k^0 x_k^0 + g_k^0\big]$$

This gives some immediate cancellation when we plug in $\overline{x}$. When grouping terms to find (Q_k^0, z_k^0, r_k^0), we often perform manipulations like $C + CFC = CF(F^{-1}+C) = CFG$ and $G - GFG = GF(F^{-1} - G) = GFC$.

6. For each $m \in \mathcal{M}$, compute the matrices describing $S_\tau^m[\psi(\cdot,y)](x)$ by running the system of differential equations given in section 2.2 up to time τ. Recall that with these matrices we have

$$S_\tau^m[\psi(\cdot,z)](x) = \frac{1}{2}\big[(x-\Lambda_\tau^m z)^T P_\tau^m (x - \Lambda_\tau^m z) + z^T R_\tau^m z + 2(L_\tau^m)^T x + 2(K_\tau^m)^T z + b_\tau^m\big]$$

Now compute each $\mathcal{B}_\tau^m(y,z)$. Using similar tricks to those described in the previous step we can find

$$\mathcal{B}_\tau^m(y,z) = -\big[\psi(\cdot,z) \odot [-S_\tau^m[\psi(\cdot,y)](\cdot)\big] = \min_{x\in\mathbb{R}^n}\{S_\tau^m[\psi(\cdot,y)](x) - \psi(x,z)\}$$

$$= \frac{1}{2}\left[y^T M_1^m y + 2y^T M_2^m z + z^T M_3^m z + 2(\rho_1^m)^T y + 2(\rho_2^m)^T z + \beta^m\right]$$

Where, denoting $E := (P_\tau^m - C)^{-1}$,

$$\begin{aligned} M_1^m &= -CEP_\tau^m & \rho_1^m &= -CEL_\tau^m \\ M_2^m &= CEP_\tau^m \Lambda_\tau^m & \rho_2^m &= \Lambda_\tau^m P_\tau^m E L_\tau^m + K_\tau^m \\ M_3^m &= R_\tau^m - (\Lambda_\tau^m)^T CEP_\tau^m \Lambda_\tau^m & \beta^m &= b_\tau^m - (L_\tau^m)^T F L_\tau^m \end{aligned}$$

7. Compute $\overline{a}^J$ by starting at $j = 0$ and performing the following update J times.

   Given $\overline{a}^j$(the dual of $\overline{V}^j$), compute $\overline{a}^{j+1}$(the dual of $\overline{V}^{j+1}$). At any $j \in \mathbb{N} \cup \{0\}$, $\overline{a}^j$ will be given as

$$\overline{a}^j(z) = \bigoplus_{k=1}^{K} \bigoplus_{\underline{m} \in \mathcal{M}^j} \hat{a}^j_{\underline{m},k}(z) \qquad \text{where} \qquad \hat{a}^j_{\underline{m},k}(z) := \left(\prod_{i=1}^{j} \hat{\mathcal{B}}_\tau^{\underline{m}_i}\right)[\hat{a}_k^0](z)$$

   Since each $\hat{a}_k^0$ is quadratic, each $\hat{a}^j_{\underline{m},k}$ will be quadratic. We write this as

$$\hat{a}^j_{\underline{m},k}(z) = \frac{1}{2}(z - z^j_{\underline{m},k})^T Q^j_{\underline{m},k}(z - z^j_{\underline{m},k}) + q^j_{\underline{m},k}$$

   Which is consistent with the form of each $\hat{a}_k^0$. We store this as the set of triplets

$$\mathcal{Q}^j = \{(Q^j_{\underline{m},k}, z^j_{\underline{m},k}, q^j_{\underline{m},k}) : 1 \le k \le K, \underline{m} \in \mathcal{M}^j\}$$

   To go from $\overline{a}^j$ to $\overline{a}^{j+1}$, we simply apply each $\hat{\mathcal{B}}_\tau^m$ to each $\hat{a}^j_{\underline{m},k}$. That is, for all $m \in \mathcal{M}$ and $\underline{m} \in \mathcal{M}^j$, let $\underline{M} = (\underline{m}, m) \in \mathcal{M}^{j+1}$ and compute

$$\hat{a}^{j+1}_{\underline{M},k}(z) := \hat{\mathcal{B}}_\tau^m[\hat{a}^j_{\underline{m},k}](z) = \mathcal{B}_\tau^m(\cdot, z) \odot \hat{a}^j_{\underline{m},k}(\cdot)$$

$$\begin{aligned} = \max_{y \in \mathbb{R}^n} \Big\{ &\frac{1}{2}\left[y^T M_1^m y + 2y^T M_2^m z + z^T M_3^m z + 2(\rho_1^m)^T y + 2(\rho_2^m)^T z + \beta^m\right] + \\ &+ \frac{1}{2}(y - z^j_{\underline{m},k})^T Q^j_{\underline{m},k}(y - z^j_{\underline{m},k}) + q^j_{\underline{m},k} \Big\} \\ &= \frac{1}{2}(z - z^{j+1}_{\underline{M},k})^T Q^{j+1}_{\underline{M},k}(z - z^{j+1}_{\underline{M},k}) + q^{j+1}_{\underline{M},k} \end{aligned}$$

   Where letting $D := (M_3^m + Q^j_{\underline{m},k})^{-1}$ and $E := D(Q^j_{\underline{m},k} z^j_{\underline{m},k} - \rho_2^m)$,

   $Q^{j+1}_{\underline{M},k} = M_1^m - M_2^m D (M_2^m)^T$

   $z^{j+1}_{\underline{M},k} = -(Q^{j+1}_{\underline{M},k})^{-1}\left[M_2^m E + \rho_m^1\right]$

$q^{j+1}_{\underline{M},k} = q^j_{\underline{m},k} + \frac{1}{2}\beta^m - \frac{1}{2}(z^{j+1}_{\underline{M},k})^T Q^{j+1}_{\underline{M},k} z^{j+1}_{\underline{M},k} + \frac{1}{2}(z^j_{\underline{m},k})^T Q^j_{\underline{m},k}(z^j_{\underline{m},k} - E)$

We then have

$$\overline{a}^{j+1}(z) = \bigoplus_{k=1}^{K} \bigoplus_{\underline{M}\in\mathcal{M}^{j+1}} \hat{a}^{j+1}_{\underline{M},k}(z)$$

Which is stored as

$$\mathcal{Q}^{j+1} = \{(Q^{j+1}_{\underline{M},k}, z^{j+1}_{\underline{M},k}, q^{j+1}_{\underline{M},k}) : 1 \le k \le K, \underline{M} \in \mathcal{M}^{j+1}\}$$

8. Recover the approximate solution $\overline{V}^J$. By lemma 7 this is given as

$$\overline{V}^j(x) = \bigoplus_{k=1}^{K} \bigoplus_{\underline{m}\in\mathcal{M}^J} \hat{V}^J_{\underline{m},k} \qquad \text{where} \qquad \hat{V}^J_{\underline{m},k} = \psi(x,\cdot) \odot \hat{a}^J_{\underline{m},k}(\cdot)$$

We can easily solve for each $\hat{V}^J_{\underline{m},k}$ with

$$\hat{V}^J_{\underline{m},k} = \max_{z\in\mathbb{R}^n}\{\frac{1}{2}(x-z)^T C(x-z) + \frac{1}{2}(z - z^J_{\underline{m},k})^T Q^J_{\underline{m},k}(z - z^J_{\underline{m},k}) + q^J_{\underline{m},k}\}$$
$$= \frac{1}{2}(x - x^J_{\underline{m},k})^T G^J_{\underline{m},k}(x - x^J_{\underline{m},k}) + g^J_{\underline{m},k}$$

Where letting $N = (Q^J_{\underline{m},k} + C)^{-1}$,

$G^J_{\underline{m},k} = CNQ^J_{\underline{m},k}$

$x^J_{\underline{m},k} = (G^J_{\underline{m},k})^{-1} CNQ^J_{\underline{m},k} z^J_{\underline{m},k}$

$g^J_{\underline{m},k} = q^J_{\underline{m},k} - \frac{1}{2}(x^J_{\underline{m},k})^T G^J_{\underline{m},k} x^J_{\underline{m},k} + \frac{1}{2} z^J_{\underline{m},k} CNQ^J_{\underline{m},k} z^J_{\underline{m},k}$

While this algorithm scales well with dimension if $\#\mathcal{M}$ is fixed, we often need more quadratics to accurately approximate the Hamiltonian as dimension grows. Furthermore, we are storing $K\mathcal{M}^j$ quadratics at the jth step, which grows fast if $\mathcal{M}$ is a large set. One potential way to decrease this growth is by pruning certain quadratics at each step in the iteration. It is relatively easy to show that if $\hat{a}^j_{\underline{m},k}(z) < \overline{a}^j(z)$ for all z, then for all $p \in \mathbb{N}$,$\underline{n} \in \mathcal{M}^p$ and $z \in \mathbb{R}^n$ we will have $\hat{a}^{j+p}_{(\underline{m},\underline{n}),k}(z) < \overline{a}^{j+p}(z)$. Furthermore, from the definition of the dual operation this will imply $\hat{V}^{j+p}_{(\underline{m},\underline{n}),k}(x) < \overline{V}^{j+p}(x)$ for all $x \in \mathbb{R}^n$. In other words, if a quadratic $\hat{a}^j_{\underline{m},k}$ plays no part in the max that defines $\overline{a}^j$, then at all future iterations $j+p$, the quadratics obtained by applying the operators $\hat{\mathcal{B}}^m_\tau$ to $\hat{a}^j_{\underline{m},k}$ will play no part in the representation of $\overline{a}^{j+p}$. Consequently, the dual of any of these aforementioned quadratics will play no part in the representation of $\overline{V}^{j+p}$. This means that at any iteration, we can drop any $\hat{a}^j_{\underline{m},k}$ with this property and not affect our approximate solution. To our knowledge, not many efficient methods of determining whether a quadratic $\hat{a}^j_{\underline{m},k}$ has this property have been developed, so this is still an open area of research. There is also potential for pruning quadratics that don't have this property and play only a small role in the max that defines $\overline{a}^j$. If we "over prune" at the jth step, then function $\overline{V}^j$ will still be a max of quadratics in G_δ, meaning it will continue to converge to the steady state $\overline{V}$.

6 Appendix

6.1 Theorem 8 proof

Assume $\phi \in \mathcal{S}_{-C}$ for a symmetric definite C such that

$$\min_{m \in \mathcal{M}} \lambda_{min}((A^m)^T C + CA^m + D^m) > 0$$

Fix $m \in \mathcal{M}$ and $\tau > 0$, recall that a continuous function f is convex iff $f(x + hv) - 2f(x) + 2f(x + hv) \geq 0$ for all $h > 0$, $x, v \in \mathbb{R}^n$ with $v \neq 0$. One can relatively easily use the properties of ξ^m and $\mathcal{L}^m$ to show that S^m_τ preserves continuity, so since ϕ is a real valued semiconvex function(and hence continuous), we have $S^m_\tau[\phi] \in \mathcal{S}_{-(C+sI)}$ iff

$$\begin{aligned} &S^m_\tau[\phi](x + hv) - 2S^m_\tau[\phi](x) + S^m_\tau[\phi](x - hv) \\ \geq &\frac{1}{2}\left[(x + hv)^T (C + sI)(x + hv) - 2x^T (C + sI)x + (x - hv)^T (C + sI)(x - hv)\right] \\ &= h^2 v^T (C + sI)v \end{aligned}$$

$\forall h > 0$ and $x, v \in \mathbb{R}^n, v \neq 0$. We will prove this assertion. First recall that

$$S^m_\tau[\phi](x) = \sup_{u \in \mathcal{U}} \left\{ \int_0^\tau \mathcal{L}^m(\xi^m_t, u_t)dt + \phi(\xi^m_\tau) \right\} = \sup_{u \in \mathcal{U}} \left\{ J^m(x, \tau; u) + \phi(\xi^m_\tau) \right\}$$

Now let $h > 0$ and $x, v \in \mathbb{R}^n$ with $v \neq 0$. By definition of the sup, we know that for any given $\epsilon > 0$, there exists $u^\epsilon \in \mathcal{U}$ and associated trajectory ξ^0_t(which follows the dynamics of ξ^m with control u^ϵ and initial condition $\xi^0_0 = x$) such that

$$S^m_\tau[\phi](x) \leq J^m(x, \tau; u^\epsilon) + \phi(\xi^0_\tau) + \epsilon$$

Using the notation in [2], let ξ^h_t, ξ^{-h}_t follow the dynamics of ξ^m with control u^ϵ and initial conditions $\xi^h_0 = x + hv$, $\xi^{-h}_0 = x - hv$. By the definition of $S^m_\tau[\phi]$ and the previous inequality,

$$S^m_\tau[\phi](x + hv) - 2S^m_\tau[\phi](x) + S^m_\tau[\phi](x - hv)$$

$$\geq \big[J^m(x+hv,\tau;u^\epsilon)+\phi(\xi_\tau^h)\big]-2\big[J^m(x,\tau;u^\epsilon)+\phi(\xi_\tau^0)+\epsilon\big]+\big[J^m(x-hv,\tau;u)+\phi(\xi_\tau^{-h})\big]$$
$$=\int_0^\tau\Big[\mathcal{L}^m(\xi_t^h,u_t^\epsilon)-2\mathcal{L}^m(\xi_t^0,u_t^\epsilon)+\mathcal{L}^m(\xi_t^{-h},u_t^\epsilon)\Big]dt+\phi(\xi_\tau^h)-2\phi(\xi_\tau^0)+\phi(\xi_\tau^{-h})-2\epsilon$$

Recall that $\dot{\xi}^m = A^m\xi^m + l_2^m + \sigma^m u$, which implies

$$\dot{\xi}_t^h-\dot{\xi}_t^0 = A^m\big[\xi_t^h-\xi_t^0\big] \quad \text{and} \quad \dot{\xi}_t^0-\dot{\xi}_t^{-h} = A^m\big[\xi_t^0-\xi_t^{-h}\big]$$

Since $\xi_0^h-\xi_0^0=\xi_0^0-\xi_0^{-h}=hv$, it is clear that $\xi_t^h-\xi_t^0=\xi_t^0-\xi_t^{-h}$ for all $t>0$. Define this as

$$\triangle\xi_t := \xi_t^h-\xi_t^0 = e^{(A^m)t}hv$$

Recalling that $\mathcal{L}^m(x,u)=\frac{1}{2}x^TD^mx-\frac{\gamma^2}{2}|u|_2^2+(l_1^m)^Tx+\alpha^m$, we have

$$\mathcal{L}^m(\xi_t^h,u_t^\epsilon)-2\mathcal{L}^m(\xi_t^0,u_t^\epsilon)+\mathcal{L}^m(\xi_t^{-h},u_t^\epsilon)$$
$$=\frac{1}{2}\Big[(\xi_t^h)^TD^m\xi_t^h-2(\xi_t^0)^TD^m\xi_t^0+(\xi_t^{-h})^TD^m\xi_t^{-h}\Big]=(\triangle\xi_t)^TD^m\triangle\xi_t$$

Furthermore, since $\phi\in\mathcal{S}_{-C}$, we know $\phi(x)-\frac{1}{2}x^TCx$ is convex and

$$\phi(\xi_\tau^h)-2\phi(\xi_\tau^0)+\phi(\xi_\tau^{-h})\geq\frac{1}{2}\Big[(\xi_\tau^h)^TC\xi_\tau^h-2(\xi_\tau^0)^TC\xi_\tau^0+(\xi_\tau^{-h})^TC\xi_\tau^{-h}\Big]=(\triangle\xi_\tau)^TC\triangle\xi_\tau$$

Since $\frac{d}{dt}\triangle\xi_t=A^m\triangle\xi_t$, we can write

$$(\triangle\xi_\tau)^TC\triangle\xi_\tau=\int_0^\tau\frac{d}{dt}\big[(\triangle\xi_t)^TC\triangle\xi_t\big]dt+(\triangle\xi_0)^TC\triangle\xi_0$$
$$=\int_0^\tau\big[(\triangle\xi_t)^T(A^m)^TC\triangle\xi_t+(\triangle\xi_t)^TCA^m\triangle\xi_t\big]dt+h^2v^TCv$$
$$=\int_0^\tau(\triangle\xi_t)^T\big[(A^m)^TC+CA^m\big]\triangle\xi_tdt+h^2v^TCv$$

Combining the previous few results and setting $B^m:=D^m+(A^m)^TC+CA^m$ we have

$$S_\tau^m[\phi](x+hv)-2S_\tau^m[\phi](x)+S_\tau^m[\phi](x-hv)$$
$$\geq\int_0^\tau(\triangle\xi_t)^TD^m\triangle\xi_tdt+(\triangle\xi_\tau)^TC\triangle\xi_\tau-2\epsilon$$
$$=\int_0^\tau(\triangle\xi_t)^TD^m\triangle\xi_tdt+\int_0^\tau(\triangle\xi_t)^T\big[(A^m)^TC+CA^m\big]\triangle\xi_tdt+h^2v^TCv-2\epsilon$$
$$=\int_0^\tau(\triangle\xi_t)^TB^m\triangle\xi_tdt+h^2v^TCv-2\epsilon\geq\int_0^\tau\lambda_{min}(B^m)|\triangle\xi_t|_2^2dt+h^2v^TCv-2\epsilon$$

We will now analyze the term $\triangle\xi_t = e^{(A^m)t}hv$. Notice $||e^{-(A^m)t}||$(operator norm), is continuous in time and equal to 1 at $t = 0$. Therefore $\exists\overline{\tau} > 0$ such that that for all $t < \overline{\tau}$, $||e^{-(A^m)t}||^2 < 2$. Hence, by the assumption that $\lambda_{min}(B^m) > 0$, this tells us that for all $t \in (0, \overline{\tau})$,

$$\lambda_{min}(B^m)|\triangle\xi_t|_2^2 = \lambda_{min}(B^m)|e^{(A^m)t}hv|_2^2 \geq \lambda_{min}(B^m)\frac{|hv|_2^2}{||e^{-(A^m)t}||^2} > \frac{\lambda_{min}(B^m)}{2}h^2|v|_2^2$$

We note that this choice of $\overline{\tau}$ is independent of h, v, and u^ϵ. Finally, set $s = \frac{\lambda_{min}(B^m)}{2}\hat{\tau}$ where $\hat{\tau} = \min\{\overline{\tau}, \tau\}$, we have

$$S_\tau^m[\phi](x + hv) - 2S_\tau^m[\phi](x) + S_\tau^m[\phi](x - hv)$$

$$\geq \int_0^\tau \lambda_{min}(B^m)|\triangle\xi_t|_2^2 dt + h^2 v^T C v - 2\epsilon \geq \int_0^{\hat{\tau}} \lambda_{min}(B^m)|\triangle\xi_t|_2^2 dt + h^2 v^T C v - 2\epsilon$$

$$> \frac{\lambda_{min}(B^m)}{2}h^2|v|_2^2\hat{\tau} + h^2 v^T C v - 2\epsilon = h^2 v^T(C + sI)v - 2\epsilon$$

Since this holds for all $\epsilon > 0$, we arrive at $S_\tau^m[\phi] \in \mathcal{S}_{-(C+sI)}$ $\square$

7 References

[1] Francis Clarke. Nonsmooth analysis in systems and control theory. *Encyclopedia of Complexity and System Science*, 2008.

[2] William M. McEneaney. A curse-of-dimensionality-free numerical method for solution of certain hjb pdes. *SIAM Journal on Control and Optimization*, 45(4):1239–1276, 2007.

[3] William M. McEneaney. Convergence rate for a curse-of-dimensionality-free method for hamilton-jacobi-bellman pdes represented as maxima of quadratic forms. *SIAM Journal on Control and Optimization*, 48(4):2651–2685, 2009.

[4] William M. McEneaney and Jonathan Kluberg. Convergence rate for a curse-of-dimensionality-free method for a class of hjb pdes. *SIAM Journal on Control and Optimization*, 48(5):3052–3079, 2009.

[5] Roger J-B Wets R. Tyrrell Rockafellar. *Variational Analysis*. Springer Science Business Media, 1997.

[6] Eduardo D. Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems*. Springer New York, 1998.

[7] Victor Magron Benjamin Werner Xavier Allamigeon, Stephane Gaubert. Certification of real inequalities - templates and sums of squares. arXiv:1403.5899v2 [math.OC], 2014.