Optimisation, contrôle, données

November 17th, 2025

Responsable: Dario Prandi, dario.prandi@centralesupelec.fr
Chargé de TD: Kevin Le Balc'h, kevin.le-balc-h@inria.fr

> **Exercise 1: Primal/Dual problems**
>
> We consider the following optimization problem (OP)
>
> $$\min_{(x,y)\in\mathbb{R}^2} (x-1)^2 + (y-2)^2,$$
>
> subject to the following constraints
>
> $$x \le y, \ \ x + 2y \le 2.$$
>
> 1. Why does (OP) admit a solution?
> 2. Write the Lagrange function associated to (OP).
> 3. Solve (OP) by using KKT's theorem.
> 4. Write the dual Lagrange function.
> 5. Do we have strong duality?

Solution

1. The objective function is continuous and coercive, the constraint set is closed and non-empty, hence (OP) admits a solution.

2. The Lagrange function is

$$L(x,y,\nu_1,\nu_2) = (x-1)^2 + (y-2)^2 + \nu_1(x-y) + \nu_2(x+2y-2), \qquad \nu_1,\nu_2 \ge 0.$$

3. The KKT conditions are

   - Primal feasibility: $x \le y$, $x + 2y \le 2$,
   - Dual feasibility: $\nu_1, \nu_2 \ge 0$,
   - Complementary slackness:

   $$\nu_1(x-y) = 0, \quad \nu_2(x+2y-2) = 0,$$

   - Stationarity:

   $$\nabla_{x,y} L(x,y,\nu_1,\nu_2) = 0 \iff \begin{cases} 2(x-1) + \nu_1 + \nu_2 = 0, \\ 2(y-2) - \nu_1 + 2\nu_2 = 0. \end{cases}$$

   By solving the system given by the last two conditions, we get the solutions $(x^\star, y^\star, \nu_1^\star, \nu_2^\star)$

   $$\left(\frac{3}{2}, \frac{3}{2}, -1, 0\right), \qquad \left(\frac{2}{3}, \frac{2}{3}, -\frac{4}{9}, \frac{10}{9}\right), \qquad (1,2,0,0), \qquad \left(\frac{2}{5}, \frac{4}{5}, 0, \frac{6}{5}\right).$$

   The first two solutions are not dual feasible, the third one is not primal feasible, hence the only solution is

   $$(x^\star, y^\star) = \left(\frac{2}{5}, \frac{4}{5}\right).$$

4. The dual Lagrange function is

$$g(\nu_1,\nu_2) = \inf_{(x,y)\in\mathbb{R}^2} L(x,y,\nu_1,\nu_2).$$

This infimum is attained since L is coercive in (x,y) for any fixed (ν_1,ν_2). By the computations for the stationarity condition, we have then

$$x = 1 - \frac{\nu_1+\nu_2}{2}, \qquad y = 2 + \frac{\nu_1 - 2\nu_2}{2}.$$

Plugging back in L, we get

$$g(\nu_1,\nu_2) = \frac{1}{4}\Big(2(\nu_1+6)\nu_2 - 2\nu_1(\nu_1+2) - 5{\nu_2}^2\Big).$$

5. Yes, strong duality holds since Slater's condition is satisfied by, e.g., $(x,y)=(0,0)$.

Exercise 2: Proximal operator and Moreau regularization

Let $f:\mathbb{R}^n \to \mathbb{R}\cup\{+\infty\}$ be a proper convex function. We define the *proximal operator* of f as

$$\operatorname{prox}_f : x \mapsto \arg\min_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2}\|y-x\|^2 \right\},$$

and the *Moreau regularization* of parameter $\mu>0$ as

$$f_\mu : x \mapsto \inf_{y\in\mathbb{R}^n} \left\{ f(y) + \frac{1}{2\mu}\|y-x\|^2 \right\}.$$

We consider the proximal point algorithm defined by the sequence

$$x^{(k+1)} = \operatorname{prox}_{\mu f}(x^{(k)}).$$

1. Show that prox_f and f_μ are well-defined. For C closed convex and non-empty, and $f=\chi_C$ (the 0-∞ characteristic function of C), identify prox_f and f_μ.
2. Show that $x^\star$ is a minimizer of f if and only if it minimizes f_μ, if and only if $x^\star = \operatorname{prox}_f(x^\star)$.
3. Show that
$$f_\mu(x) = \frac{1}{2\mu}\|x\|^2 - \frac{1}{\mu}\left(\mu f + \frac{1}{2}\|\cdot\|^2\right)^*(x),$$
where $(\cdot)^*$ denotes the Fenchel conjugate.
4. Show that
$$\operatorname{prox}_{\mu f}(x) = \arg\max_y \left\{ x^\top y - \mu f(y) - \frac{1}{2}\|y\|^2 \right\}.$$
5. Show that
$$\nabla f_\mu(x) = \frac{1}{\mu}\left(x - \operatorname{prox}_{\mu f}(x)\right).$$
and interpret the proximal point algorithm as a gradient algorithm.
Hint: Recall that $v\in\partial g(x)$ if and only if $v \in \arg\max_y v^\top y - g(y)$, if and only if $x\in\partial g^*(v)$.
6. Writing
$$f_\mu(x) = \min_{y,z}\left[f(y) + \frac{1}{2\mu}\|z\|^2\right] \qquad \text{s.t.} \qquad x-y=z,$$
and using duality, show that
$$f_\mu(x) = \left(f^* + \frac{\mu}{2}\|\cdot\|^2\right)^*(x).$$
This implies that f_μ has a $\frac{1}{\mu}$-Lipschitz gradient.
7. Show that if f is a proper convex lower-semicontinuous function admitting a minimizer, the proximal point algorithm converges toward a minimizer of f.

Solution

1. The function $y \mapsto f(y) + \frac{1}{2}\|y - x\|^2$ is strongly convex and coercive, hence it admits a unique minimizer, so prox_f is well-defined. A similar argument shows that f_μ is well-defined.
For $f = \chi_C$, we have
$$\text{prox}_f(x) = \arg\min_{y \in C} \|y - x\|^2,$$
which is the projection of x onto C. The Moreau regularization is
$$f_\mu(x) = \inf_{y \in C} \frac{1}{2\mu}\|y - x\|^2 = \frac{\text{dist}(x, C)^2}{2\mu}.$$

2. If $x^\star$ minimizes f, then for all y,
$$f(y) + \frac{1}{2\mu}\|y - x^\star\|_2^2 \geq f(x^\star) = f(x^\star) + \frac{1}{2\mu}\|x^\star - x^\star\|^2.$$
so $x^\star$ is such that $f_\mu(x^\star) = f(x^\star)$ and it minimizes f_μ. Taking $\mu = 1$, this implies $x^\star = \text{prox}_f(x^\star)$.
On the other hand, $x^\star = \text{prox}_f(v)$ if and only if
$$0 \in \partial f(x^\star) + (x^\star - v).$$
Taking $v = x^\star$, this gives $0 \in \partial f(x^\star)$, i.e., $x^\star$ minimizes f.

3. We have
$$f_\mu(x) = \inf_y \left\{ f(y) + \frac{1}{2\mu}\|y\|^2 - \frac{1}{\mu}x^\top y + \frac{1}{2\mu}\|x\|^2 \right\} = \frac{1}{2\mu}\|x\|^2 + \frac{1}{\mu}\inf_y \left\{ \mu f(y) + \frac{1}{2}\|y\|^2 - x^\top y \right\}.$$
The result follows by recognizing the Fenchel conjugate in the infimum.

4. We have
$$\text{prox}_{\mu f}(x) = \arg\min_y \left\{ \mu f(y) + \frac{1}{2}\|y - x\|^2 \right\} = \arg\min_y \left\{ \mu f(y) + \frac{1}{2}\|y\|^2 - x^\top y \right\},$$
and the result follows.

5. From point 4, we have
$$0 \in \mu \partial f(\text{prox}_{\mu f}(x)) + \text{prox}_{\mu f}(x) - x \iff x - \text{prox}_{\mu f}(x) \in \mu \partial f(\text{prox}_{\mu f}(x)).$$
From point 3, taking the gradient, we have
$$\nabla f_\mu(x) = \frac{1}{\mu}x - \frac{1}{\mu}\nabla \left(\mu f + \frac{1}{2}\| \cdot \|^2 \right)^* (x).$$
From the properties of the Fenchel conjugate, we have
$$\nabla \left(\mu f + \frac{1}{2}\| \cdot \|^2 \right)^* (x) = \arg\max_y \left\{ x^\top y - \mu f(y) - \frac{1}{2}\|y\|^2 \right\} = \text{prox}_{\mu f}(x),$$
hence the result. The proximal point algorithm can then be written as
$$x^{(k+1)} = x^{(k)} - \mu \nabla f_\mu(x^{(k)}),$$
which is a gradient descent with step size μ on the Moreau regularization of f.

6. The Lagrangian is
$$L(y, z, \nu) = f(y) + \frac{1}{2\mu}\|z\|^2 + \nu^\top (x - y - z).$$
The dual function is
$$g(\nu) = \inf_{y,z} L(y, z, \nu) = \inf_y \left\{ f(y) - \nu^\top y \right\} + \inf_z \left\{ \frac{1}{2\mu}\|z\|^2 - \nu^\top z \right\} + \nu^\top x = -f^*(\nu) - \frac{\mu}{2}\|\nu\|^2 + \nu^\top x.$$
Hence, the dual problem is
$$\max_\nu g(\nu) = \max_\nu \left\{ x^\top \nu - f^*(\nu) - \frac{\mu}{2}\|\nu\|^2 \right\} = \left(f^\star + \frac{\mu}{2}\| \cdot \|^2 \right)^* (x).$$
Since f^* is convex, the function $h(\nu) = f^*(\nu) + \frac{\mu}{2}\|\nu\|^2$ is strongly convex with parameter μ.

7. As f has $1/\mu$-Lipschitz gradient, the gradient descent with step size μ converges to a minimizer of f for any initial point $x^{(0)}$.

Exercise 3: LQ optimal control problem

Consider the two 2-dimensional linear systems driven by the same scalar input $u(t)$:

$$\dot{x}_1 = A_1 x_1 + B_1 u, \qquad A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad B_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

$$\dot{x}_2 = A_2 x_2 + B_2 u, \qquad A_2 = \begin{bmatrix} -2 & 0 \\ 1 & 0 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Let the augmented state be $X = [x_1^\top, x_2^\top]^\top \in \mathbb{R}^4$, and define

$$\mathcal{A} = \operatorname{diag}(A_1, A_2), \qquad \mathcal{B} = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}.$$

1. For each pair (A_i, B_i), determine whether the system is controllable.
2. Determine whether the two systems are *simultaneously controllable* by the single input u and, if it is not, describe the non-reachable states.
3. Replace A_2 by the parametric matrix

$$A_2(\alpha) = \begin{bmatrix} -2 & \alpha \\ 1 & 0 \end{bmatrix}.$$

   For which $\alpha \in \mathbb{R}$ is the augmented pair $(\mathcal{A}(\alpha), \mathcal{B})$ controllable? Provide the condition(s) discuss their relationship to the controllability of each of the systems separately.

Consider the LQ problem for the augmented system

$$\dot{X} = \mathcal{A}X + \mathcal{B}u, \qquad X(0) = X_0,$$

with final time $T > 0$ and cost

$$J(u) = \tfrac{1}{2} \int_0^T \left(X(t)^\top R X(t) + a\, u(t)^2 \right) dt,$$

where $R = \operatorname{diag}(R_1, R_2)$, each $R_i \in \mathbb{R}^{2\times 2}$ is symmetric positive semidefinite, and $a > 0$.

4. Write the expression for the optimal control of this unconstrained terminal state problem in terms of the matrix $E(t)$, which satisfies the Riccati differential equation

$$\dot{E} = R - \mathcal{A}^\top E - E\mathcal{A} - \frac{1}{r} E\mathcal{B}\mathcal{B}^\top E,$$

   with appropriate boundary conditions at $E(T)$, to be determined. Write the Riccati equation explicitly using the matrices $\mathcal{A} = \operatorname{diag}(A_1, A_2)$, $\mathcal{B}$, and R, and the block decomposition of E:

$$E = \begin{bmatrix} E_{11} & E_{12} \\ E_{21} & E_{22} \end{bmatrix}.$$

5. Discuss why the optimal control u^* exists and is unique for any initial state $X(0)$, even if the augmented system $(\mathcal{A}, \mathcal{B})$ is not controllable.
6. Is it possible for the optimal control to act independently on the the two subsystems? *Hint: How this would reflect on the Riccati equation solution?*

Solution

1. The Kalman matrix for the first system is
$$\mathcal{K}_1 = [B_1, A_1 B_1] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$
which has rank 2, so the first system is controllable. For the second system,
$$\mathcal{K}_2 = [B_2, A_2 B_2] = \begin{bmatrix} 2 & -4 \\ 1 & 2 \end{bmatrix},$$
which also has rank 2, so the second system is controllable.

2. The Kalman matrix for the augmented system is
$$\mathcal{K} = [\mathcal{B}, \mathcal{A}\mathcal{B}, \mathcal{A}^2\mathcal{B}, \mathcal{A}^3\mathcal{B}] = \begin{bmatrix} 0 & 1 & 2 & 1 \\ 1 & 0 & -4 & 2 \\ 0 & 0 & 8 & -4 \\ 0 & 0 & -16 & 8 \end{bmatrix},$$
which has rank 3 (the last two columns are linearly dependent), so the augmented system is not controllable.
The non-reachable states are those orthogonal to the image of $\mathcal{K}$, i.e., those such that $(x_1, x_2, x_3, x_4) \cdot \mathcal{A}^j \mathcal{B} = 0$ for $j = 1, 2, 3$. These are exactly the states satisfying $x_4 = 2x_3$.

3. The augmented system is controllable if and only if the Kalman matrix
$$\mathcal{K}(\alpha) = [\mathcal{B}, \mathcal{A}(\alpha)\mathcal{B}, \mathcal{A}(\alpha)^2\mathcal{B}, \mathcal{A}(\alpha)^3\mathcal{B}]$$
has rank 4. We have $\det(\mathcal{K}(\alpha)) = \alpha^2(\alpha - 8)$, so the system is controllable if and only if $\alpha \neq 0$ and $\alpha \neq 8$.
The case $\alpha = 0$ is the one treated in point 2: the systems are separately but not simultaneously controllable. For $\alpha = 8$, one checks that the second system is not controllable, which entails the non-controllability of the augmented system.

4. The optimal control is given by
$$u^*(t) = -\frac{1}{a}\mathcal{B}^\top E(T-t)X(t),$$
where $E(t)$ satisfies the Riccati equation with terminal condition $E(0) = 0$ (there is no terminal cost). We have that $E_{12}^\top = E_{21}$ and the Riccati equation reads
$$\begin{aligned} \dot{E}_{11} &= R_1 - A_1^\top E_{11} - E_{11}A_1 - \frac{1}{a}E_{11}B_1B_1^\top E_{11} - \frac{1}{a}E_{12}B_2B_2^\top E_{21}, \\ \dot{E}_{12} &= -A_1^\top E_{12} - E_{12}A_2 - \frac{1}{a}E_{11}B_1B_1^\top E_{12} - \frac{1}{a}E_{12}B_2B_2^\top E_{22}, \\ \dot{E}_{22} &= R_2 - A_2^\top E_{22} - E_{22}A_2 - \frac{1}{a}E_{21}B_1B_1^\top E_{12} - \frac{1}{a}E_{22}B_2B_2^\top E_{22}. \end{aligned}$$

5. The optimal control exists and is unique since the cost is strictly convex in u (due to the term au^2 with $a > 0$) and the dynamics are linear in u, for any initial state $X(0)$, independently of the controllability of the system.

6. No, if the optimal control acted independently on the two subsystems, we would have $E_{12} \equiv 0$, which is not possible since the equation for $\dot{E}_{12}$ contains terms depending on E_{11} and E_{22}.

Exercise 4: Bilinear control problem and linear ResNets

Consider the following control problem:
$$\dot{x} = Ux, \qquad x \in \mathbb{R}^d,\, U \in \mathbb{R}^{d\times d}. \tag{2}$$

1. We aim at giving necessary and sufficient conditions for the existence of a time-independent control $U \in L^\infty([0,1], \mathbb{R}^{d\times d})$ steering the system from $x_0 \in \mathbb{R}^d$ to $y_0 \in \mathbb{R}^d$.

   (a) Recall that the solution of (2) is $x(t) = R(t)x(0)$, where $R(t) = R(t,0)$ is the state-transition matrix $R(t)$, which satisfies

   $$\frac{\partial}{\partial t} R(t) = U(t)R(t), \qquad R(0) = \mathrm{Id}\,.$$

   Use this fact to reduce controllability of (2) to a controllability problem for the state-transition matrix to $R(1)$ such that $y_0 = R(1)x_0$. Conclude for the case $x_0 = 0$ and $y_0 \neq 0$.

   (b) For the case $x_0 \neq 0$ and y_0 linearly independent of x_0, show that there exists $v \perp (y_0 - x_0)$ and $\eta \in \mathbb{R}$ such that $U = \eta(y_0 - x_0)v^\top$ steers the system from x_0 to y_0.
   Hint: compute $R(1) = \exp(U)$ using the series expansion of the matrix exponential.

   (c) Consider $d = 2$ and show that in general it is sufficient to consider a control of the form

   $$U = \gamma J + \eta\,\mathrm{Id}, \qquad J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \qquad \gamma, \eta \in \mathbb{R}.$$

   Deduce the form of the control for general $d \geq 2$.
   Hint: Recall that $\exp(\gamma J)$ is the counter-clockwise rotation matrix of angle γ and that $e^{A+B} = e^A e^B$ if the matrices A and B commute.

2. Consider the following simultaneous controllability problem: Given two sets $\{x_1, \ldots, x_d\}$ and $\{y_1, \ldots, y_d\}$ which are bases of $\mathbb{R}^d$ find a *single* time-independent control U steering the system from each x_i to y_i, $i = 1, \ldots, d$.

   (a) Show that the this problem is controllable with time-independent controls if and only if the map $U \in \mathbb{R}^{d\times d} \mapsto \exp U$ is surjective onto the set of invertible matrices.

   (b) Show that this is false in general.

   (c) For $d = 2$, let $X = [x_1, x_2]$ and $Y = [y_1, y_2]$. Show that a sufficient condition for simultaneous controllability is that $X^{-1}Y$ has positive determinant and admits two distinct eigenvalues $\lambda_1, \lambda_2 \in \mathbb{C} \setminus (-\infty, 0)$.

3. Write down the ResNet with L layers corresponding to the neural ODE of equation (2). What can we say about the universal approximating properties of this family of ResNets?
   Hint: Are there general restriction on simultaneous controllability with time-dependent controls?

Solution

1. (a) Since $x(t) = R(t)x(0)$, we have that $y_0 = R(1)x_0$. Hence, controllability of (2) reduces to the existence of U such that $R(1)$ satisfies $y_0 = R(1)x_0$. If $x_0 = 0$ and $y_0 \neq 0$, there is no control U steering the system from x_0 to y_0 since $x(t) = R(t)0 = 0$ for all $t \in [0,1]$.

   (b) Let $v \perp (y_0 - x_0)$ and $\eta \in \mathbb{R}$. We have

   $$U^2 = \eta^2 (y_0 - x_0)v^\top v(y_0 - x_0)^\top = 0,$$

   since $v^\top(y_0 - x_0) = 0$. Hence,

   $$R(1) = \exp(U) = \mathrm{Id} + U = \mathrm{Id} + \eta(y_0 - x_0)v^\top.$$

   Therefore,

   $$R(1)x_0 = x_0 + \eta(y_0 - x_0)v^\top x_0.$$

   Hence, if $v^\top x_0 \neq 0$ we can choose $\eta = 1/(v^\top x_0)$, to get $R(1)x_0 = y_0$.
   Let us show that such v exists. Assume by contradiction that $v^\top x_0 = 0$ for all $v \perp (y_0 - x_0)$. Then,

   $$0 = v^\top x_0 = v^\top y_0 - v^\top(y_0 - x_0) = v^\top y_0, \qquad \forall v \perp (y_0 - x_0).$$

Pick a basis $\{v_1, \dots, v_{d-1}\}$ of the orthogonal complement of $(y_0 - x_0)$. Then, y_0 and x_0 are orthogonal to all v_i, $i = 1, \dots, d-1$, hence they are both parallel to $(y_0 - x_0)$, contradicting their linear independence.

(c) Let $\varrho(\gamma)$ be the rotation matrix, that is,

$$\varrho(\gamma) = \exp(\gamma J) = \begin{bmatrix} \cos(\gamma) & -\sin(\gamma) \\ \sin(\gamma) & \cos(\gamma) \end{bmatrix}.$$

In dimension $d = 2$, any pair of vectors (x_0, y_0) can be written as

$$y_0 = \lambda \varrho(\gamma) x_0, \qquad \lambda > 0,\, \gamma \in \mathbb{R}.$$

Moreover, Id and J are commuting matrices. Thus, choosing

$$U = \log(\lambda)\,\mathrm{Id} + \gamma J,$$

we have

$$R(1) = e^U = e^{\log(\lambda)\,\mathrm{Id}} e^{\gamma J} = \lambda \varrho(\gamma) \implies R(1)x_0 = y_0.$$

2. (a) The problem is controllable with time-independent controls if and only if there exists U such that $R(1) = \exp(U)$ satisfies $y_i = R(1)x_i$ for all $i = 1, \dots, d$. This is equivalent to the existence of U such that

$$R(1)[x_1, \dots, x_d] = [y_1, \dots, y_d].$$

Since both $\{x_i\}$ and $\{y_i\}$ are bases of $\mathbb{R}^d$, the matrices $X = [x_1, \dots, x_d]$ and $Y = [y_1, \dots, y_d]$ are invertible. Hence, the problem is controllable if and only if there exists U such that

$$YX^{-1} = \exp(U).$$

The map $U \mapsto \exp(U)$ is surjective onto the set of invertible matrices if and only if for any invertible matrix M there exists U such that $\exp(U) = M$. Hence, the problem is controllable if and only if the map $U \mapsto \exp(U)$ is surjective onto the set of invertible matrices.

(b) The map $U \mapsto \exp(U)$ is not surjective onto the set of invertible matrices. For instance, there is no U such that $\exp(U) = -\mathrm{Id}$ in dimension $d = 2$ since the eigenvalues of $\exp(U)$ are e^{λ_1} and e^{λ_2}, where λ_1 and λ_2 are the eigenvalues of U. Hence, the eigenvalues of $\exp(U)$ are always positive real numbers.

(c) Let $d = 2$ and let $M = X^{-1}Y$. If $\det(M) > 0$ and M has two distinct eigenvalues $\lambda_1, \lambda_2 \in \mathbb{C} \setminus (-\infty, 0)$, then there exists a matrix $P \in \mathbb{C}^{2\times 2}$ such that

$$M = P \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} P^{-1}.$$

Since $\lambda_1, \lambda_2 \in \mathbb{C} \setminus (-\infty, 0)$, there exist $\mu_1, \mu_2 \in \mathbb{C}$ such that $e^{\mu_1} = \lambda_1$ and $e^{\mu_2} = \lambda_2$. Hence,

$$M = P \exp\left(\begin{bmatrix} \mu_1 & 0 \\ 0 & \mu_2 \end{bmatrix} \right) P^{-1} = \exp(U), \qquad U = P \begin{bmatrix} \mu_0 & 0 \\ -1 & \mu_2 \end{bmatrix} P^{-2},$$

Thus, we have $\exp(U) = M$, which concludes the proof.

3. The ResNet with L layers corresponding to the neural ODE of equation (2) is

$$x_{k+1} = x_k + \frac{1}{L} U_k x_k, \qquad k = 0, \dots, L-1,$$

where $U_k \in \mathbb{R}^{d\times d}$ are the weights of the layers.

Regarding the universal approximating properties of this family of ResNets, we can observe that for any control $U \in L^\infty([0,1], \mathbb{R}^{d\times d})$ the state-transition matrix satisfies $\det R(t) > 0$ for all $t \in [0, 1]$. Indeed,

$$\frac{d}{dt} \det R(t) = \det R(t) \operatorname{tr}(R(t)^{-1} \dot{R}(t)) = \det R(t) \operatorname{tr}(U(t))$$

This implies that

$$\det R(t) = \exp\left(\int_0^t \operatorname{tr}(U(s))\, ds \right) > 0.$$

Therefore, the family of ResNets corresponding to (2) with time-dependent controls cannot approximate maps that send bases to bases with negative determinant. Hence, this family of ResNets is not a universal approximator.