Second order necessary conditions for optimal control problems with endpoints-constraints and convex control-constraints *

Li Deng†

Abstract In this manuscript, we consider a control system governed by a general ordinary differential equation on a Riemannian manifold, with its endpoints satisfying some inequalities and equalities, and its control constrained to a closed convex set. We concern on an optimal control problem of this system, and obtain the second order necessary condition in the sense of convex variation (Theorem 2.2). To this end, we first obtain a second order necessary condition of an optimization problem (Theorem 4.2) via separation theorem of convex sets. Then, we derive our necessary condition by transforming the optimal control problem into an optimization problem. It is worth to point out that, our necessary condtition evolves the curvature tensor, which is trivial in Euclidean case. Moreover, even M is a Euclidean space, our result is still of interest. Actually, we give an example (Example 2.1) which shows that, when an optimal control stays at the boundary of the control set, the existing results are invalid while Theorem 2.2 works.

Keywords Optimal control, Second order necessary condition, Endpoints-constraints, Convex constraints, Riemannian manifold

MSC (2010) **49K15, 49K30, 93C15, 58E25, 70Q05**

1 Introduction

*This work is supported by the National Science Foundation of China under grants 11401491 and 11231007, and the Fundamental research funds for the Central Universities under grant 2682014CX052.

†School of Mathematics, Southwest Jiaotong University, Chengdu 611756, Sichuan Province, China. *E-mail:* dengli@swjtu.edu.cn.

In this paper, we consider a control system described by a general ordinary differential equation with the state restricted to a manifold, with the initial and terminal states restricted to inequality-type and equality-type constraints, and with the control constrained pointwisely to a convex set. For this control system, we study the second order necessary optimality condition for an optimal control problem.

Before elaborating on our problem, we introduce some notions on manifolds. Let $n \in \mathbb{N}$ and M be a complete simply connected, n-dimensional manifold with Riemannian metric g. Let ∇ be the Levi-Civita connection on M related to g, $\rho(\cdot,\cdot)$ be the distance function on M, and T_xM and T_x^*M be respectively the tangent and cotangent spaces of M at $x \in M$. Denote by $\langle\cdot,\cdot\rangle$ and $|\cdot|$ the inner product and the norm over T_xM related to g, respectively. Also, denote by $TM \equiv \bigcup\limits_{x\in M} T_xM$, $T^*M \equiv \bigcup\limits_{x\in M} T_x^*M$ and $C^\infty(M)$ the tangent bundle, the cotangent bundle and the set of smooth functions on M, respectively.

Let $j,k \in \mathbb{N}$, $T>0$, U be a subset of $\mathbb{R}^m$ ($m \in \mathbb{N}$), and $f : [0,T]\times M\times U \to TM$, $\phi_i : M\times M \to \mathbb{R}$ ($i = 0,1,\cdots,j$) and $\psi = (\psi_1,\cdots,\psi_k)^\top : M\times M \to \mathbb{R}^k$ be maps (satisfying suitable assumptions to be given later). Set by

$$\mathcal{U} = \{v : [0,T] \to U \,|\, v(\cdot)\,\text{is measurable}\} \tag{1.1}$$

the set of all possible controls. Consider the following optimal control problem

(OCP) Find a control $\bar{u}(\cdot)$ and a trajectory $\bar{y}(\cdot)$ minimizing

$$J(y(\cdot),u(\cdot)) \equiv \phi_0(y(0),y(T)),$$

subject to

$$\dot{y}(t) = f(t,y(t),u(t)),\ a.e.\,t\in[0,T];\ u(\cdot)\in\mathcal{U}; \tag{1.2}$$

and

$$\begin{cases} \phi_i(y(0),y(T)) \le 0,\ i=1,\cdots,j, \\ \psi(y(0),y(T)) = 0. \end{cases} \tag{1.3}$$

$\bar{u}(\cdot)$, $\bar{y}(\cdot)$ and $(\bar{u}(\cdot),\bar{y}(\cdot))$ are respectively called optimal control, optimal trajectory and optimal pair.

For problem (OCP), Deng and Zhang ([5]) obtained the second order necessary condition via spike variation, by applying the separation theorem of convex sets to a suitably chosen set related to the second order spike variation. In this paper, we are concerned with the second order necessary condition obtained by convex variation.

In optimal control theory, optimality conditions are usually obtained by two kinds of variations of trjectories: spike and convex variations. In this paper, we call a necessary condition obained by convex variation (resp. spike variation) a necessary condition in the sense of convex variation (resp. spike variation). For the first order necessary optimality condition, the condition in the sense of spike variation is always more precise than that in the sense of convex variation, i.e. the latter can be deduced directly from the former. However, for the second order necessary condition, it is hard to assess which condition is better. As already explained in [5, Section 1], the second order necessary condition only makes sense along a critical direction, in which the first order necessary condition is trivial. Since different variations lead to different first order necessary conditions, critical directions in the senses of different variations are distinct. Consequently the corresponding second order necessary conditions are different. They are always complementary to each other. Two examples in [10] (or [4, Section 3.5]) show that, these two kinds of second order necessary conditions can not cover each other. Our main result (Theorem 2.2) is the second order necessary condition in the sense of convex variation, which can be viewed as a complement to [5, Theorem 2.2].

For problem (OCP), when M is a Euclidean space, the second order necessary condition in the sense of convex variation is given by [7, Theorem 3.1]. Comparing to it, our main result (i.e. Theorem 2.2) has two differences: i) The curvature tensor of the Riemannian manifold arises. This is the same case as the corresponding second order necessary condition in the sense of spike variation (i.e. [5, Theorem 2.2]); ii) There is another extra term (i.e. the first integral of the left hand side of inequality (2.23)), which always makes sense when an optimal control stays at the boundary of the control set, and provides additional informations. We explain this by Example 2.1, in which Theorem 2.2 works, while [7, Theorem 3.1] fails.

The appearance of the curvature tensor results from the second order variation of trajectories evolved on Riemannian manifolds. The extra term (the first integral of the left hand side of inequality (2.23)) comes from the use of "the second-order adjacent subset" introduced in [1, Definition 4.7.2, p. 171]. More precisely, we prove Theorem 2.2 by using the second order necessary condition of a solution to an optimization problem with inequality-type and equality-type constraints (i.e. Theorem 4.2), which evolves "the second-order adjacent subset". With this nontrivial term, Theorem 4.2 generalizes [7, Theorem 4.1], see Remark 4.1 for details.

This paper is organised as follows. The main results are stated in Section 2, the variations of trajectories to the second order are given in Section 3, and Section 4 is devoted to the proof of the main results.

2 Statement of the main results

2.1 Notations and assumptions

We first introduce some notions. Denote by $i(x)$, $|\mathcal{T}(x)|$, $\nabla\mathcal{T}$, R, the injectivity radius (at the point $x \in M$), the norm of the tensor field $\mathcal{T}$ at the point $x \in M$ (see [3, (2.5)]), the covariant derivative of the tensor field $\mathcal{T}$ and the curvature tensor (of (M, g)), respectively. For any $x, y \in M$ with $\rho(x, y) < \min\{i(x), i(y)\}$, by [3, Lemma 2.1], there exists a unique shortest geodesic connecting x and y. We denote the parallel translation of a tensor from x to y along this geodesic by L_{xy}. For a smooth function $h : M \times M \to \mathbb{R}$ of two arguments, we denote by $\nabla_i h(y_1, y_2)$ the covariant derivative of h with respect to the i^{th} argument y_i with $i = 1, 2$. Namely, we have

$$\nabla_i h(y_1, y_2)(X) = X(y_i)(h(y_1, y_2)), \ \forall\, X \in TM.$$

Thus, $\nabla_i h(y_1, y_2) \in T^*_{y_i}M$. For a smooth vector valued map $\psi = (\psi_1, \cdots, \psi_k)^\top : M \times M \to \mathbb{R}^k$ $(k > 0)$, we denote the first and second order covariant derivatives with respect to the i^{th} argument $(i = 1, 2)$ respectively by

$$\begin{aligned} &\nabla_i \psi(X) = (\nabla_i \psi_1(X), \cdots, \nabla_i \psi_k(X))^\top, \\ &\nabla_i^2 \psi(X, Y) = (\nabla_i^2 \psi_1(X, Y), \cdots, \nabla_i^2 \psi_k(X, Y))^\top, \end{aligned} \tag{2.1}$$

for all $X, Y \in TM$. The corresponding norms are given respectively by $|\nabla_i \psi| = \sum_{l=1}^k |\nabla_i \psi_l|$ and $|\nabla_i^2 \psi| = \sum_{l=1}^k |\nabla_i^2 \psi_l|$. For the definitions of the above notions, please see [3, Section. 2].

To present the optimality conditions of problem (OCP), we need to introduce two functions: Lagrange and Hamiltonion functions. When $k > 0$, the Lagrange function $\mathcal{L} : M \times M \times \mathbb{R}^{1+j+k} \to \mathbb{R}$ is defined by $\mathcal{L}(y_1, y_2; \ell) \equiv \sum_{i=0}^j \ell_i \phi_i(y_1, y_2) + \ell_\psi^\top \psi(y_1, y_2)$, where $\ell = (\ell_0, \cdots, \ell_j, \ell_\psi^\top)^\top$. When $k = 0$, the corresponding Lagrange function $\mathcal{L} : M \times M \times \mathbb{R}^{1+j} \to \mathbb{R}$ is $\mathcal{L}(y_1, y_2; \ell) \equiv \sum_{i=0}^j \ell_i \phi_i(y_1, y_2)$, where $\ell = (\ell_0, \cdots, \ell_j)^\top$. For each $\ell \in \mathbb{R}^{1+j+k}$, we also denote by $d_i\mathcal{L}(y_1, y_2; \ell)$ the differential of $\mathcal{L}$ with respect to the variable y_i $(i = 1, 2)$, i.e. $d_i\mathcal{L}(y_1, y_2; \ell) \in T^*_{y_i}M$ satisfies $d_i\mathcal{L}(y_1, y_2; \ell)(X) = X(y_i)(\mathcal{L}(y_1, y_2; \ell))$ for all $X \in TM$. The Hamiltonian function $H : [0, T] \times T^*M \times U \to \mathbb{R}$ is given by

$$H(t, y, p, u) \equiv p(f(t, y, u)), \quad \forall\, (t, y, p, u) \in [0, T] \times T^*M \times U. \tag{2.2}$$

Throughout this paper, we denote by $\tilde{X} \in T^*M$ (resp. $\tilde{X} \in TM$) the dual covector (resp. vector) of $X \in TM$ (resp. $X \in T^*M$), see [5, Section 2.1] for the detailed definition.

Then, we recall some definitions concerning tangent sets. For more details, we refer to [1]. Let $\mathcal{X}$ be a metric space with a metric d, and $K \subset \mathcal{X}$ be a subset. The distance between a point $x \in \mathcal{X}$ and K is defined by $dist_K(x) := \inf\{d(x, y); y \in K\}$. Let $\{K_h\}_{h>0}$

be a family of subsets of $\mathcal{X}$. The lower limit of $\{K_h\}_{h>0}$ is given by

$$Liminf_{h\to 0^+}K_h := \{v\in\mathcal{X};\ \lim_{h\to 0^+} dist_{K_h}(v)=0\}.$$

When $\mathcal{X}$ is a normed vector space, the adjacent cone to a subset $K\subset\mathcal{X}$ at a point $x\in\overline{K}$ (i.e. x belongs to the closure of K) is defined by (see [1, p.127])

$$T_K^{\flat}(x) := Liminf_{h\to 0^+}\frac{K-x}{h}. \tag{2.3}$$

Moreover, for $v\in T_K^{\flat}(x)$, the second-order adjacent subset to K at (x,v) is given by (see [1, Definition 4.7.2, p. 171])

$$T_K^{\flat(2)}(x,v) := Liminf_{h\to 0^+}\frac{K-x-hv}{h^2}. \tag{2.4}$$

By the definition of the lower limit of a family of subsets, one can respectively characterise $T_K^{\flat}(x)$ and $T_K^{\flat(2)}(x,v)$ in terms of sequences:

(i) $v\in T_K^{\flat}(x)$ if and only if, for any $h_n\to 0^+$ as $n\to+\infty$, there exists $v_n\in\mathcal{X}$ approaching to v, such that $x+h_nv_n\in K$ for each $n\geq 1$;

(ii) Given $v\in T_K^{\flat}(x)$, $w\in T_K^{\flat(2)}(x,v)$ if and only if, for any $h_n\to 0^+$ as $n\to+\infty$, there exists $w_n\in\mathcal{X}$ approaching to w, such that $x+h_nv+h_n^2w_n\in K$ for each $n\geq 1$.

The main assumptions are exhibited as follows:

$(C1)$ $U\subset\mathbb{R}^m$ is convex and closed.

$(C2)$ The map $f(=f(t,x,u)):[0,T]\times M\times U\to TM$ is measurable in t, and C^1 in (x,u). Moreover, there exists a constant $K>1$ such that,

$$\begin{aligned}
&|L_{x_1\hat{x}_1}f(s,x_1,u_1)-f(s,\hat{x}_1,u_2)|\leq K(\rho(x_1,\hat{x}_1)+|u_1-u_2|),\\
&|f(s,x_0,u_1)|\leq K,\\
&|\phi_i(x_1,x_2)-\phi_i(\hat{x}_1,\hat{x}_2)|\leq K(\rho(x_1,\hat{x}_1)+\rho(x_2,\hat{x}_2)),\ i=0,\cdots,j,\\
&|\psi(x_1,x_2)-\psi(\hat{x}_1,\hat{x}_2)|\leq K(\rho(x_1,\hat{x}_1)+\rho(x_2,\hat{x}_2)),
\end{aligned} \tag{2.5}$$

for all $s\in[0,T]$, $u_1,u_2\in U$, and $x_l,\hat{x}_l\in M$ satisfying $\rho(x_l,\hat{x}_l)<\min\{i(x_l),i(\hat{x}_l)\}$ for $l=1,2$, where $x_0\in M$ is fixed.

$(C3)$ The map $f=(f(s,x,u))$ is C^2 in $(x,u)\in M\times U$, and $\phi_0,\cdots,\phi_j,\psi$ are C^2 over $M\times M$. Furthermore, there exists a positive constant $K>1$ such that

$$\begin{aligned}
&|\nabla_xf(s,x_1,u_1)-L_{\hat{x}_1x_1}\nabla_xf(s,\hat{x}_1,u_2)|\leq K(\rho(x_1,\hat{x}_1)+|u_1-u_2|),\\
&|\nabla_uf(s,x_1,u_1)-L_{\hat{x}_1x_1}\nabla_uf(s,\hat{x}_1,u_2)|\leq K(\rho(x_1,\hat{x}_1)+|u_1-u_2|),\\
&|\nabla_1\phi_i(x_1,x_2)-L_{\hat{x}_1x_1}\nabla_1\phi_i(\hat{x}_1,x_2)|\leq K\rho(x_1,\hat{x}_1),\ i=0,1,\cdots,j,\\
&|\nabla_2\phi_i(x_1,x_2)-L_{\hat{x}_2x_2}\nabla_2\phi_i(x_1,\hat{x}_2)|\leq K\rho(x_2,\hat{x}_2),\ i=0,1,\cdots,j,\\
&|\nabla_1\psi(x_1,x_2)-L_{\hat{x}_1x_1}\nabla_1\psi(\hat{x}_1,x_2)|\leq K\rho(x_1,\hat{x}_1),\\
&|\nabla_2\psi(x_1,x_2)-L_{\hat{x}_2x_2}\nabla_2\psi(x_1,\hat{x}_2)|\leq K\rho(x_2,\hat{x}_2),
\end{aligned} \tag{2.6}$$

for all $x_1, \hat{x}_1, x_2, \hat{x}_2 \in M$ with $\rho(x_1, \hat{x}_1) < \min\{i(x_1), i(\hat{x}_1)\}$ and $\rho(x_2, \hat{x}_2) < \min\{i(x_2), i(\hat{x}_2)\}$, and $(s, u_1, u_2) \in [0, T] \times U \times U$, where $\nabla_u f(s, x, u)$ is defined by

$$\begin{aligned} & \nabla_u f(s,x,u)(\eta, V) \\ = \;& \lim_{\epsilon\to 0^+} \tfrac{1}{\epsilon}\Big(f(s,x,u+\epsilon V)(\eta) - f(s,x,u)(\eta)\Big),\ \forall\, (\eta, V) \in T_x^*M \times \mathbb{R}^m, \end{aligned} \tag{2.7}$$

with its norm

$$|\nabla_u f(s,x,u)| \equiv \sup\{\nabla_u f(s,x,u)(\eta,V);\, (\eta,V) \in T_x^*M \times \mathbb{R}^m, |\eta| + |V| \le 1\},$$

and $\nabla_x f(s, x, u)$ is the covariant derivative of $f(s, x, u)$ with respect to the state variable $x \in M$, and is defined by

$$\nabla_x f(s,x,u)(\eta, X) = \nabla_X f(s,\cdot,u)(\eta), \quad \forall\, (\eta, X) \in T_x^*M \times T_xM, \tag{2.8}$$

with its norm given by [3, (2.5)].

We should mention that, for $l = 1, 2$ and $i = 0, \cdots, j$, by [3, Lemma 4.1], we obtain from (2.5) and (2.6) that, f and $\nabla_x f$ are both Lipschitz continuous with respect to the variable $(x, u) \in M \times U$, and ϕ_i, ψ, $\nabla_l \phi_i$ and $\nabla_l \psi$ are Lipschitz continuous. These conditions can be checked by computing the norms of $\nabla_x f$, $\nabla_x^2 f$, $\nabla_l \phi_i$, $\nabla_l^2 \phi_i$, $\nabla_l \psi$ and $\nabla_l^2 \psi$.

2.2 Main results

In this subsection, we fix an optimal pair $(\bar{u}(\cdot), \bar{y}(\cdot))$. For a function defined on $[0, T] \times M \times U$, we denote by

$$\varphi[t] \equiv \varphi(t, \bar{y}(t), \bar{u}(t)), \quad \forall\, t \in [0, T] \tag{2.9}$$

for abbreviation. Set

$$I_A \equiv \{i \in \{1, \cdots, j\};\, \phi_i(\bar{y}(0), \bar{y}(T)) = 0\} \cup \{0\}, \tag{2.10}$$

$$I_N \equiv \{0, 1, \cdots, j\} \setminus I_A. \tag{2.11}$$

Given a vector $\ell = (\ell_0, \cdots, \ell_j, \ell_\psi^\top)^\top \in \mathbb{R}^{1+j+k}$, we denote by $p^\ell(\cdot)$ the solution to

$$\begin{cases} \nabla_{\dot{\bar{y}}(t)} p^\ell = -\nabla_x f[t](p^\ell(t), \cdot),\ a.e.\, t \in (0, T), \\ p^\ell(T) = d_2 \mathcal{L}(\bar{y}(0), \bar{y}(T); \ell), \end{cases} \tag{2.12}$$

with $\nabla_x f$ given by (2.8). It is a covector field along $\bar{y}(\cdot)$, i.e. $p^\ell(t) \in T^*_{\bar{y}(t)}M$ for each $t \in [0, T]$. Furthermore, for a function φ defined on $[0, T] \times T^*M \times U$, we set

$$\varphi[t, \ell] = \varphi(t, \bar{y}(t), p^\ell(t), \bar{u}(t)),\ \forall\, t \in [0, T] \tag{2.13}$$

for abbreviation.

The first order necessary condition of an optimal pair in the secnse of convex variation is stated as follows.

Theorem 2.1 *Assume $U \subset \mathbb{R}^m$ $(m \in \mathbb{N})$ is convex, and condition $(C2)$ holds. If $(\bar{u}(\cdot), \bar{y}(\cdot))$ is optimal pair for problem (OCP) with $\bar{u}(\cdot) \in L^2(0,T;\mathbb{R}^m) \cap \mathcal{U}$, then there exists $\ell = (\ell_0, \ell_1, \cdots, \ell_j, \ell_\psi^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ such that*

$$\ell_i \leq 0, \text{ if } i \in I_A; \quad \ell_i = 0, \text{ if } i \in I_N, \tag{2.14}$$

and

$$\nabla_u H[t,\ell](v(t)) \leq 0, \text{ a.e. } t \in [0,T], \tag{2.15}$$

holds for all $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$ with $v(t) \in T_U^\flat(\bar{u}(t))$ a.e $t \in [0,T]$, where $p^\ell(\cdot)$ is a covector field along $\bar{y}(\cdot)$ satisfying (2.12) and initial condition

$$p^\ell(0) = -d_1\mathcal{L}(\bar{y}(0), \bar{y}(T); \ell), \tag{2.16}$$

and $\nabla_u H(t,y,\eta,u)(V)$ (with $(t,y,\eta,u,V) \in [0,T] \times T^*M \times U \times \mathbb{R}^m$) is defined by

$$\nabla_u H(t,y,\eta,u)(V) = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon}\Big(H(t,y,\eta,u+\epsilon V) - H(t,y,\eta,u)\Big). \tag{2.17}$$

From the viewpoint of calculus, when the first order necessary condition is trivial in some direction, it is necessary to find the second order necessary condition along this direction. Thus, in what follows, we give the definition of critical direction in the sense of convex variation.

Definition 2.1 *Assume that $(\bar{u}(\cdot), \bar{y}(\cdot))$ is an optimal pair of problem (OCP) with $\bar{u}(\cdot) \in L^2(0,T;\mathbb{R}^m) \cap \mathcal{U}$, and that all the assumptions in Theorem 2.1 hold. A function $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$ is called a singular direction in the sense of convex variation, if it satisfies*

$$\begin{aligned}
&v(t) \in T_U^\flat(\bar{u}(t)), \quad a.e.\, t \in [0,T], \\
&\nabla_1\phi_i(\bar{y}(0), \bar{y}(T))(X_v(0)) + \nabla_2\phi_i(\bar{y}(0), \bar{y}(T))(X_v(T)) \leq 0, \ \forall\, i \in I_A, \\
&\nabla_1\psi(\bar{y}(0), \bar{y}(T))(X_v(0)) + \nabla_2\psi(\bar{y}(0), \bar{y}(T))(X_v(T)) = 0 \quad (\text{omit if } k = 0),
\end{aligned} \tag{2.18}$$

where $\nabla_i\psi$ is defined by (2.1), and $X_v(\cdot)$ is a vector field along $\bar{y}(\cdot)$ (i.e. $X_v(t) \in T_{\bar{y}(t)}M$ for all $t \in [0,T]$) and verifies

$$\nabla_{\dot{\bar{y}}(t)} X_v = \nabla_x f[t](\cdot, X_v(t)) + \nabla_u f[t](\cdot, v(t)), \text{ a.e. } t \in (0,T), \tag{2.19}$$

with $\nabla_x f$ and $\nabla_u f$ given respectively by (2.8) and (2.7).

Along a critical direction $v(\cdot)$ defined above, the first order necessary condition in the sense of convex variation is trival. To show this, we need a definition as follows.

Definition 2.2 *Assume that all the assumptions in Theorem 2.1 hold, and that $(\bar{u}(\cdot), \bar{y}(\cdot))$ is an optimal pair of problem (OCP). A vector $\ell = (\ell_0, \ell_1, \cdots, \ell_j, \ell_\psi^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ is called a Lagrange multiplier in the sense of convex variation, if it satisfies (2.14), and (2.15) holds for all $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$ with $v(t) \in T_U^\flat(\bar{u}(t))$ a.e $t \in [0,T]$, where $p^\ell(\cdot)$ satisfies (2.12) and (2.16).*

Thus, we can understand the first order necessary condition in the sense of convex variation (i.e. Theorem 2.1) as follows: if $(\bar{u}(\cdot), \bar{y}(\cdot))$ is an optimal pair, there exists a Lagrange multiplier in the sense of convex variation. Moreover, if $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$ is a crictical direction in the sense of convex variation, with $X_v(\cdot)$ satisfying (2.18) and (2.19), then for any Lagrange multiplier $\ell \in \mathbb{R}^{1+j+k} \setminus \{0\}$, by (2.18), (2.12), (2.16), (2.19), (2.15) and integration by parts, we have

$$\begin{aligned} 0 \leq & \nabla_1 \mathcal{L}(\bar{y}(0), \bar{y}(T); \ell)(X_v(0)) + \nabla_2 \mathcal{L}(\bar{y}(0), \bar{y}(T); \ell)(X_v(T)) \\ = & \int_0^T \nabla_u H[t, \ell](v(t)) dt \leq 0, \end{aligned}$$

which implies $\nabla_u H[t,\ell](v(t)) = 0$ a.e $t \in [0,T]$. Thus, along direction $v(\cdot)$, the first order necessary condition in the sense of convex variation is trivial.

Then, along this direction $v(\cdot)$, we shall study the second order necessary condition. To this end, associated to $v(\cdot)$ and $X_v(\cdot)$, we set

$$\begin{aligned} & I_0' \equiv I_N \cup \{i \in I_A;\ \nabla_1 \phi_i(\bar{y}(0), \bar{y}(T))(X_v(0)) + \nabla_2 \phi_i(\bar{y}(0), \bar{y}(T))(X_v(T)) < 0\}; \\ & I_0'' \equiv \{0, 1, \cdots, j\} \setminus I_0'. \end{aligned} \tag{2.20}$$

The second order necessary condition of optimal pairs is as follows.

Theorem 2.2 *Assume that conditions $(C1)-(C3)$ hold, and that $(\bar{u}(\cdot), \bar{y}(\cdot))$ with $\bar{u}(\cdot) \in L^2(0,T;\mathbb{R}^m) \cap \mathcal{U}$ is an optimal pair of problem (OCP). Let $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$ be a critical direction in the sense of convex variation, with $X_v(\cdot)$ satisfying (2.18) and (2.19). Assume that there exist $\ell(\cdot) \in L^2(0,T;\mathbb{R}^m)$ and $\epsilon_0 > 0$ such that*

$$dist_U(\bar{u}(t) + \epsilon v(t)) \leq \epsilon^2 \ell(t), \forall \epsilon \in [0, \epsilon_0], a.e.\, t \in [0,T], \tag{2.21}$$

and that the set $\mathcal{B} \equiv \{\sigma(\cdot) \in L^2(0,T;\mathbb{R}^m);\ \sigma(t) \in T_U^{\flat(2)}(\bar{u}(t), v(t))\, a.e.\, t \in [0,T]\} \neq \emptyset$. Then, there exists a Lagrange multiplier $\ell = (\ell_{\phi_0}, \ell_{\phi_1}, \cdots, \ell_{\phi_j}, \ell_\psi^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ satisfying

$$\ell_{\phi_i} = 0,\ if\, i \notin I_0'', \tag{2.22}$$

such that

$$\begin{aligned} & \textstyle\int_0^T \nabla_u H[t,\ell](\sigma(t)) dt + \frac{1}{2} \int_0^T \big\{ \nabla_x^2 H[t,\ell](X_v(t), X_v(t)) \\ & + 2 \nabla_u \nabla_x H[t,\ell](X_v(t), v(t)) + \nabla_u^2 H[t,\ell](v(t), v(t)) \\ & - R(\tilde{p}^\ell(t), X_v(t), f[t], X_v(t)) \big\} dt + \tfrac{1}{2} \nabla_1^2 \mathcal{L}(\bar{y}(0), \bar{y}(T); \ell)(X_v(0), X_v(0)) \\ & + \nabla_1 \nabla_2 \mathcal{L}(\bar{y}(0), \bar{y}(T); \ell)(X_v(T), X_v(0)) \\ & + \tfrac{1}{2} \nabla_2^2 \mathcal{L}(\bar{y}(0), \bar{y}(T); \ell)(X_v(T), X_v(T)) \leq 0 \end{aligned} \tag{2.23}$$

holds for all $\sigma(\cdot) \in \mathcal{B}$, where $p^\ell(\cdot)$ satisfies (2.12) and (2.16), $\tilde{p}^\ell(t)$ ($t \in [0,T]$) is the dual vector of $p^\ell(t)$, $H[t,\ell]$ is defined in (2.2) and (2.13), for each $(t,y,\eta,u) \in [0,T]\times T^*M\times U$,

$$\begin{aligned}
&\nabla_x H(t,y,\eta,u)(X) = \nabla_x f(t,y,u)(\eta, X), \quad \forall\, X \in T_yM,\\
&\nabla_u\nabla_x H(t,y,\eta,u)(X,V) = \frac{d}{ds}\Big|_{s=0}\nabla_x H(t,y,\eta,u+sV)(X),\ \forall\,(X,V)\in T_yM\times\mathbb{R}^m,\\
&\nabla_u^2 H(t,y,\eta,u)(V,V) = \frac{d^2}{ds^2}\Big|_{s=0} H(t,y,\eta,u+sV), \quad \forall\, V\in\mathbb{R}^m,\\
&\nabla_x^2 H(t,y,\eta,u)(X,X) = \nabla_x^2 f(t,y,u)(\eta,X,X), \quad \forall\, X\in T_yM,
\end{aligned}$$

and $\nabla_i\nabla_j\mathcal{L}(\bar{y}(0),\bar{y}(T);\ell)$ ($i,j=1,2$) is defined by [5, (5.11)].

Remark 2.1 *[7, Theorem 3.1] considers a special case of problem (OCP): M is a Euclidean space. Theorems 2.1 & 2.2 differ from [7, Theorem 3.1] in three aspects. First, Theorem 2.1 extends [7, Theorem 3.1] from a Euclidean space to a Riemannian manifold. What comes new is that, the curvature tensor of the Riemannian manifold appears in the second order necessary condition. Second, [7, Theorem 3.1] says that, if $\bar{u}(\cdot)$ is an optimal control, the first order necessary condtion is that, there exists a nontrivial vector $\ell = (\ell_{\phi_0},\cdots,\ell_{\phi_j},\ell_\psi^\top)^\top \in \mathbb{R}^{1+j+k}\setminus\{0\}$ satisfying (2.14), such that the following inequality holds:*

$$\int_0^T \nabla_u H[t,\ell](v(t))dt \le 0,\ \forall\, v(\cdot)\in L^\infty(0,T;\mathbb{R}^m)\cap(\mathcal{U}-\{\bar{u}(\cdot)\}). \tag{2.24}$$

While (2.15) is of pointwise form, which is easier to be checked. Third, when there exists a set $A\subset[0,T]$ with its Lebesgue measure bigger than zero, such that $\bar{u}(t)$ belongs to the boundary of U for all $t\in A$, the set $\{v(\cdot)\in L^2(0,T;\mathbb{R}^m)|\, v(t)\in T_U^\flat(\bar{u}(t))\, a.e.\, t\in[0,T]\}$ is some times larger than $L^\infty(0,T;\mathbb{R}^m)\cap(\mathcal{U}-\bar{u}(\cdot))$ (see Example 2.1). Consequently, when a singular direction $v(\cdot)$ in the sense of convex variation satisfies $v(t)\in T_U^\flat(\bar{u}(t))\setminus\{U-\{\bar{u}(t)\}\}$ for $t\in A$, compared to [7, Theorem 3.1], we still have further information about an optimal pair (see (2.23)). We shall use Example 2.1 below to illustrate it more explicitly.

Example 2.1 *Given $T\in(0,3-\sqrt{5})$ and $\theta>2$, consider the control system*

$$\begin{cases} \dot{y}_1(t) = u_2(t), & a.e.\, t\in(0,T),\\ \dot{y}_2(t) = -y_1^2(t)+4y_1(t)u_2(t)-\theta u_1(t)^2, & a.e.\, t\in(0,T), \end{cases} \tag{2.25}$$

where $(u_1(t),u_2(t))\in B(1)\stackrel{\triangle}{=}\{(x,y)^\top\in\mathbb{R}^2|x^2+y^2\le 1\}$ a.e. $t\in(0,T)$. Set by $\phi_0(y_1(0),y_2(0),y_1(T),y_2(T)) = y_2(T)$ and $\psi(y_1(0),y_2(0),y_1(T),y_2(T)) = (y_1(0)-1,y_2(0))^\top$. Then, the optimal control problem is to minimize $\phi_0(y_1(0),y_2(0),y_1(T),y_2(T))$, where

$(y_1(\cdot), y_2(\cdot), u_1(\cdot), u_2(\cdot))$ *is subject to (2.25) and* $\psi(y_1(0), y_2(0), y_1(T), y_2(T)) = 0$*. Consider the control* $\bar{u}(t) = (\bar{u}_1(t), \bar{u}_2(t))^\top \equiv (0, -1)^\top$*. The corresponding trajectroy is*

$$(\bar{y}_1(t), \bar{y}_2(t)) = (1 - t, -\frac{1}{3}t^3 + 3t^2 - 5t),\ \forall\, t \in [0, T].$$

Then, we shall use [7, Theorem 3.1] and Theorem 2.2 respectively to check whether $\bar{u}(\cdot)$ *is optimal.*

Solution. It can be checked that $T^\flat_{B(1)}((0, -1)^\top) = \{(x, y)^\top | x \in \mathbb{R}, y \geq 0\}$, which is strictly larger than $B(1) - (0, -1)^\top$. We also have $T^{\flat(2)}_{B(1)}((0, -1)^\top, (1, 0)^\top) = \{(x, y) | x \in \mathbb{R}, y \geq \frac{1}{2}\}$. By the [7, Theorem 3.1(i)](or Theorem 2.1), there exists a unique $\ell = (\ell_0, \ell_1, \ell_2)^\top \in \mathbb{R}^3 \setminus \{0\}$ (up to a positive factor) satisfying $\ell_0 \leq 0$ and (2.24) (or (2.15)), and we also have $-\ell_2 = \ell_0$ and $\ell_1 = (6T - T^2)\ell_0$. Without loss of generality, we assume $\ell_0 = -1$. We observe that, for all $v(\cdot) \in L^\infty(0, T; \mathbb{R}^2) \cap (\mathcal{U} - \{\bar{u}(\cdot)\})$, the relation $\int_0^T \nabla_u H[t, \ell](v(t))dt < 0$ holds, which means that, [7, Theorem 3.1(ii)](the second order necessary condition) can not be applied to $\bar{u}(\cdot)$. However, by Theorem 2.1 we know that $v(t) \equiv (1, 0)^\top \in T^\flat_{B(1)}(\bar{u}(t)) \setminus (B(1) - \{\bar{u}(t)\})$ $(t \in [0, T])$ is a singular direction. Set by $\sigma(t) \equiv (0, \frac{1}{2})^\top \in T^{\flat(2)}_{B(1)}(\bar{u}(t), (1, 0)^\top)$ $(t \in [0, T])$. The left hand side of (2.23) is reduced to $T(-\frac{1}{3}T^2 + \frac{5}{2}T + \theta - 2) > 0$, which contradicts (2.23). Consequently, $\bar{u}(\cdot)$ is not an optimal control. $\square$

Then, we would apply Theorem 2.2 to the following problem

(OCPE) Minimize $J(u(\cdot)) = \int_0^T f^0(t, y(t), u(t))dt$ over $u(\cdot) \in \mathcal{U}$ subject to (1.2), $y(0) = y_0$ and $y(T) = y_1$, where $y_0, y_1 \in M$ are fixed, and $f : \mathbb{R}^+ \times M \times U \to TM$ and $f^0 : \mathbb{R}^+ \times M \times U \to \mathbb{R}$ are given maps.

The Hamiltonian function $H^e : [0, T] \times T^*M \times U \times \mathbb{R} \to \mathbb{R}$ is given by $H^e(t, y, p, u, \ell) = p(f(t, y, u)) + \ell f^0(t, y, u)$, where $(t, y, p, u, \ell) \in [0, T] \times T^*M \times U \times \mathbb{R}$. We need the following assumption:

(C_e) The maps f and f^0 are measurable in t, C^2 in (x, u), and there exist a constant $K > 1$ and $x_0 \in M$ such that

$$\begin{aligned}
&|f^0(s, x_1, u_1) - f^0(s, x_2, u_2)| \leq K(\rho(x_1, x_2) + |u_1 - u_2|),\\
&|f^0(s, x_0, u_1)| \leq K,\\
&|\nabla_x f^0(s, x_1, u_1) - L_{\hat{x}_1 x_1} \nabla_x f^0(s, \hat{x}_1, u_2)| \leq K\big(\rho(x_1, x_2) + |u_1 - u_2|\big),\\
&|\nabla_u f^0(s, x_1, u_1) - \nabla_u f^0(s, \hat{x}_1, u_2)| \leq K\big(\rho(x_1, x_2) + |u_1 - u_2|\big),
\end{aligned}$$

hold for all $x_1, \hat{x}_1, x_2 \in M$, $u_1, u_2 \in U$ and $s \in \mathbb{R}^+$, with $\rho(x_1, \hat{x}_1) < \min\{i(x_1), i(\hat{x}_1)\}$.

Then, the corresponding second order necessary condition is stated as follows.

Corollary 2.1 *Assume that conditions $(C1)$ and (C_e) hold, that there exist a constant $K > 1$ and $x_0 \in M$ such that the first two lines of (2.5) and (2.6) hold for all $x_1, \hat{x}_1 \in M$ with $\rho(x_1, \hat{x}_1) < \min\{i(x_1), i(\hat{x}_1)\}$, $u_1, u_2 \in U$ and $s \in \mathbb{R}^+$, and that $(\bar{u}(\cdot), \bar{y}(\cdot))$ is an optimal pair of problem $(OCPE)$. Then, there exist $\ell_0 \leq 0$ and $\varphi \in T^*_{y_1}M$ such that $(\ell_0, \varphi) \neq 0$ and*

$$\nabla_u H^e[t, \ell_0, \varphi](w(t)) \leq 0, \quad a.e.\ t \in [0, T], \tag{2.26}$$

holds for any $w(\cdot) \in L^2(0, T; \mathbb{R}^m)$ with $w(t) \in T^\flat_U(\bar{u}(t))$ a.e. $t \in [0, T]$, where we have used notion (2.9), $p_{\ell_0\varphi}(\cdot)$ solves

$$\begin{cases} \nabla_{\dot{\bar{y}}(t)} p_{\ell_0\varphi} = -\nabla_x f[t](p_{\ell_0\varphi}(t), \cdot) - \ell_0 \nabla_x f^0[t],\ a.e.\ t \in (0, T), \\ p_{\ell_0\varphi}(T) = \varphi, \end{cases} \tag{2.27}$$

and we have adopted $[t, \ell_0, \varphi] = (t, \bar{y}(t), p_{\ell_0\varphi}(t), \bar{u}(t), \ell_0)$ for abbreviation. Moreover, for any $v(\cdot) \in L^2(0, T; \mathbb{R}^m)$ with $v(t) \in T^\flat_U(\bar{u}(t))$ a.e. $t \in [0, T]$ and vector field $X_v(\cdot)$ along $\bar{y}(\cdot)$ satisfying (2.19), $X_v(0) = 0$, $X_v(T) = 0$, and $\int_0^T \big(\nabla_x f^0[t](X_v(t)) + \nabla_u f^0[t](v(t))\big)\, dt \leq 0$, there exist $(\hat{\ell}_0, \hat{\varphi}) \in \big((-\infty, 0] \times T^*_{y_1}M\big) \setminus \{0\}$ satisfying (2.26) and (2.27) (with (ℓ_0, φ) replaced by $(\hat{\ell}_0, \hat{\varphi})$), such that

$$\begin{aligned} &\int_0^T \nabla_u H^e[t, \hat{\ell}_0, \hat{\varphi}]\sigma(t)dt + \frac{1}{2}\int_0^T \Big(\nabla_x^2 H^e[t, \hat{\ell}_0, \hat{\varphi}](X_v(t), X_v(t)) \\ &+ 2\nabla_u \nabla_x H^e[t, \hat{\ell}_0, \hat{\varphi}](X_v(t), v(t)) + \nabla_u^2 H^e[t, \hat{\ell}_0, \hat{\varphi}](v(t), v(t)) \\ &- R(\tilde{p}_{\hat{\ell}_0\hat{\varphi}}(t), X_v(t), f[t], X_v(t))\Big)\, dt \leq 0, \end{aligned} \tag{2.28}$$

holds for all $\sigma(\cdot) \in \mathcal{B}$, where $p_{\hat{\ell}_0\hat{\varphi}}(\cdot)$ is the solution to (2.27) with (ℓ_0, φ) replaced by $(\hat{\ell}_0, \hat{\varphi})$, $\tilde{p}_{\hat{\ell}_0\hat{\varphi}}(t)$ ($t \in [0, T]$) is the dual vector of $p_{\hat{\ell}_0\hat{\varphi}}(t)$, and for $(t, y, \eta, u, \ell) \in [0, T] \times T^*M \times U \times \mathbb{R}$, the corresponding values of $\nabla_u H^e$, $\nabla_x H^e$, $\nabla_x^2 H^e$, $\nabla_u \nabla_x H^e$ and $\nabla_u^2 H^e$ at (t, y, η, u, ℓ) are respectively defined by

$$\begin{aligned} &\nabla_u H^e(t, y, \eta, u, \ell)(V) = \frac{d}{ds}\Big|_{s=0} H^e(t, y, \eta, u + sV, \ell), \\ &\nabla_x H^e(t, y, \eta, u, \ell)(X) = \nabla_x f(t, y, u)(\eta, X) + \ell \nabla_x f^0(t, y, u)(X), \\ &\nabla_x^2 H^e(t, y, \eta, u, \ell)(X, X) = \nabla_x^2 f(t, y, u)(\eta, X, X) + \ell \nabla_x^2 f^0(t, y, u)(X, X), \\ &\nabla_u \nabla_x H^e(t, y, \eta, u, \ell)(X, V) = \frac{d}{ds}\Big|_{s=0} \nabla_x H^e(t, y, \eta, u + sV, \ell)(X), \\ &\nabla_u^2 H^e(t, y, \eta, u, \ell)(V, V) = \frac{d^2}{ds^2}\Big|_{s=0} H(t, y, \eta, u + sV, \ell), \end{aligned}$$

where $(X, V) \in T_yM \times \mathbb{R}^m$.

By Theorem 2.2, we can use the idea in [5, Section 3.1] to prove Corollary 2.1, and we omit its proof.

For the case that U is open (not necessarily convex), [3, Theorem 3.2] gives the second order necessary condition of optimal pairs of problem $(OCPE)$, which is in fact obtained through convex variation. Thus, Corollary 2.1 is an complement to it.

3 Variations of Trajectories

In this section, we will compute variations of (1.2), in the sense of convex variation.

Proposition 3.1 *Assume that conditions* $(C2)$ *and* $(C3)$ *hold, and that* $U \subset \mathbb{R}^m$ *(*$m \in \mathbb{N}$*) is convex. Let* $\bar{u}(\cdot) \in L^2(0,T;\mathbb{R}^m) \cap \mathcal{U}$ *and* $(\bar{u}(\cdot), \bar{y}(\cdot))$ *satisfy (1.2). Fix* $W \in T_{\bar{y}(0)}M$ *and* $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$*. Let* $X_v(\cdot)$ *be a vector field along* $\bar{y}(\cdot)$ *and satisfy (2.19). Assume a set* $\{\sigma_\epsilon(\cdot)\}_{\epsilon>0} \subset L^2(0,T;\mathbb{R}^m)$ *is bounded in* $L^2(0,T;\mathbb{R}^m)$*, with* $\sup_{\epsilon>0} \|\sigma_\epsilon(\cdot)\|_{L^2(0,T;\mathbb{R}^m)} = C_\sigma$*. Denote by* $y_\epsilon(\cdot)$ *the solution to (1.2) corresponding to the control* $u_\epsilon(\cdot) := \bar{u}(\cdot) + \epsilon v(\cdot) + \epsilon^2 \sigma_\epsilon(\cdot)$ *and the initial state* $y_\epsilon(0) = \exp_{\bar{y}(0)}(\epsilon X_v(0) + \epsilon^2 W)$*. Also denote by* $Y^{X_v}_{\sigma_\epsilon W}(\cdot)$ *the solution to*

$$\begin{cases} \nabla_{\dot{\bar{y}}(t)} Y^{X_v}_{\sigma_\epsilon W}(Z) = \nabla_x f[t](Z, Y^{X_v}_{\sigma_\epsilon W}(t)) + \nabla_u f[t](Z, \sigma_\epsilon(t)) + \nabla_u \nabla_x f[t](Z, X_v(t), v(t)) \\ \qquad\qquad\qquad -\frac{1}{2} R(\tilde{Z}, X_v(t), \dot{\bar{y}}(t), X_v(t)) + \frac{1}{2}\nabla_x^2 f[t](Z, X_v(t), X_v(t)) \\ \qquad\qquad\qquad +\frac{1}{2}\nabla_u^2 f[t](Z, v(t), v(t)), \quad a.e.\, t \in (0,T], \forall\, Z \in T^*M, \\ Y^{X_v}_{\sigma_\epsilon W}(0) = W, \end{cases} \tag{3.1}$$

where we adopt notion (2.9), and

$$\nabla_u^2 f[t](Z, v(t), v(t)) = \frac{\partial^2}{\partial s^2}\Big|_{s=0} f\big(t, \bar{y}(t), \bar{u}(t) + s v(t)\big)(Z),$$
$$\nabla_u \nabla_x f[t]\big(Z, X_v(t), v(t)\big) = \frac{\partial}{\partial s}\Big|_{s=0} \nabla_x f\big(t, \bar{y}(t), \bar{u}(t) + s v(t)\big)\big(Z, X_v(t)\big).$$

Then, for any $\alpha > 0$*, there exists* $\epsilon^0 > 0$ *such that*

$$|V_\epsilon(t) - \epsilon X_v(t) - \epsilon^2 Y^{X_v}_{\sigma_\epsilon W}(t)| \le \alpha \epsilon^2, \quad \forall\, t \in [0,T], \forall\, \epsilon \in [0, \epsilon^0], \tag{3.2}$$

where

$$V_\epsilon(t) := \exp^{-1}_{\bar{y}(t)} y_\epsilon(t), \quad t \in [0,T]. \tag{3.3}$$

Proof. The proof is split into three steps.

Step 1. We claim that, there exists $\hat{\epsilon} > 0$ such that (3.3) can be defined for $\epsilon \in [0, \hat{\epsilon}]$, and

$$\begin{aligned} |V_\epsilon(t)| &= \rho(y_\epsilon(t), \bar{y}(t)) \\ &\le \epsilon e^{\frac{1}{2}(K+\frac{1}{2})T} \Big(|X_v(0) + \epsilon W|^2 + 4K^2 \|v\|^2_{L^2(0,T;\mathbb{R}^m)} + 4K^2 C_\sigma^2\Big)^{\frac{1}{2}}, \; t \in [0,T], \end{aligned} \tag{3.4}$$

for all $\epsilon \in [0, \hat{\epsilon}]$.

In fact, let $\epsilon_0 \in (0, 1]$ be such that $\epsilon_0|X_v(0)| + \epsilon_0^2|W| < i(\bar{y}(0))$, where the injectivity radius $i(y_0)$ of the point y_0 is defined in [3, Section 2.1]. Then, by [3, Lemma 5.2], the triangle inequality of $\rho(\cdot, \cdot)$ and [3, Lemma 2.2], we have

$$\begin{array}{rl} & \rho(y_\epsilon(t), \bar{y}(0)) \leq \rho(y_\epsilon(t), y_\epsilon(0)) + \rho(y_\epsilon(0), \bar{y}(0)) \\ \leq & (1 + \rho(x_0, \bar{y}(0)) + \epsilon|X_v(0)| + \epsilon^2|W|)e^{Kt}, \quad \forall\, t \in [0, +\infty), \quad \forall\, \epsilon \in [0, \epsilon_0]. \end{array} \tag{3.5}$$

By HopfRinow theorem (see [8, Theorem. 16, p. 137]), there exists $\delta > 0$ such that $i(y) \geq \delta$ for all $y \in M$ with $\rho(\bar{y}(0), y) \leq (1 + \rho(x_0, \bar{y}(0)) + \epsilon_0|X_v(0)| + \epsilon_0^2|W|)e^{KT}$. Similar to (3.5), for any $t, \hat{t} \in [0, +\infty)$, we have

$$\rho(y_\epsilon(t), \bar{y}(t)) \leq \Big(2 + 2\rho(x_0, \bar{y}(0)) + \epsilon|X_v(0)| + \epsilon^2|W|\Big)(e^{Kt} - e^{K\hat{t}}) + \rho(y_\epsilon(\hat{t}), \bar{y}(\hat{t})), \tag{3.6}$$

for all $\epsilon \in [0, \epsilon_0]$.

Let $\hat{t} = 0$ in (3.6). We take $\epsilon_1 > 0$ such that $\epsilon_1|X_v(0)| + \epsilon_1^2|W| = \frac{\delta}{2}$, and take $T_1 > 0$ such that $(2 + 2\rho(x_0, \bar{y}(0)) + \epsilon_1|X_v(0)| + \epsilon_1^2|W|\Big)(e^{KT_1} - 1) = \frac{\delta}{2}$. Thus, we can define (3.3) over $[0, T_1] \cap [0, T]$, and obtain by [3, (2.17) & (2.19)], [5, (2.2)] and $(C2)$ that

$$\begin{array}{rcl} \frac{d}{dt}\rho^2(y_\epsilon(t), \bar{y}(t)) & = & \nabla_1\rho^2(y_\epsilon(t), \bar{y}(t))\Big(f(t, y_\epsilon(t), u_\epsilon(t)) - L_{\bar{y}(t)y_\epsilon(t)}f(t, \bar{y}(t), u_\epsilon(t))\Big) \\ & & +\nabla_2\rho^2(y_\epsilon(t), \bar{y}(t))\Big(f(t, \bar{y}(t), u_\epsilon(t)) - f[t]\Big) \\ & \leq & (K + \frac{1}{2})\rho^2(y_\epsilon(t), \bar{y}(t)) + 4\epsilon^2K^2(|v(t)|^2 + |\sigma_\epsilon(t)|^2), \end{array} \tag{3.7}$$

for all $t \in [0, T_1] \cap [0, T]$ and $\epsilon \in [0, \epsilon_1]$. By the Gronwall's inequality and [3, Lemma 2.2], we obtain that (3.4) holds for all $t \in [0, T_1] \cap [0, T]$ and $\epsilon \in [0, \epsilon_1]$.

If $T_1 < T$, we set $\hat{t} = T_1$ in (3.6), $\hat{\epsilon} = \min\{\epsilon_1, \frac{\delta}{2C_{v,\sigma}}\}$ with $(C_{v,\sigma})^2 = e^{(K+\frac{1}{2})T}\Big(2|X_v(0)|^2 + 2\epsilon_1^2|W|^2 + 4K^2\|v\|^2_{L^2(0,T;\mathbb{R}^m)} + 4K^2C_\sigma^2\Big)$, and $T_2 > T_1$ satisfying $\Big(2 + 2\rho(x_0, \bar{y}(0)) + \hat{\epsilon}|X_v(0)| + \hat{\epsilon}^2|W|\Big)(e^{KT_2} - e^{KT_1}) = \frac{\delta}{2}$. Then we can define (3.3) over $[0, T_2] \cap [0, T]$ for $\epsilon \in [0, \hat{\epsilon}]$, and consequently (3.7) holds for $t \in [0, T_2] \cap [0, T]$ and $\epsilon \in [0, \hat{\epsilon}]$. Analogously we get (3.4) for $t \in [0, T_2] \cap [0, T]$ and $\epsilon \in [0, \hat{\epsilon}]$. Recursively, if $T_i < T$, we take $\hat{t} = T_i$ $(i \geq 2)$, and set $T_{i+1} > T_i$ such that

$$\Big(2 + 2\rho(x_0, \bar{y}(0)) + \hat{\epsilon}|X_v(0)| + \hat{\epsilon}^2|W|\Big)(e^{KT_{i+1}} - e^{KT_i}) = \frac{\delta}{2}. \tag{3.8}$$

Then, one can define (3.3) over $[0, T_{i+1}] \cap [0, T]$. Consequently, we obtain (3.4) for $t \in [0, T_{i+1}] \cap [0, T]$ and $\epsilon \in [0, \hat{\epsilon}]$. It follows from (3.8) that there eixsts $I \in \mathbb{N}$, such that $T_I > T$, and consequently, (3.4) holds for $t \in [0, T]$ and $\epsilon \in [0, \hat{\epsilon}]$.

Step 2. Let $\{e_1, \cdots, e_n\} \subset T_{\bar{y}(0)}M$ be an orthonormal basis at $\bar{y}(0)$, i.e. $\langle e_i, e_j\rangle = \delta_i^j$ for $i, j = 1, \cdots, n$, where δ_i^j is the usual Kronecker symbol. Denote by $\{d_i\}_{i=1}^n \subset T^*_{\bar{y}(0)}M$ the dual basis to it. For $t \in [0,T]$, set respectively by $e_i(t) = L^{\bar{y}(\cdot)}_{\bar{y}(0)\bar{y}(t)} e_i$ and $d_i(t) = L^{\bar{y}(\cdot)}_{\bar{y}(0)\bar{y}(t)} d_i$ for $i = 1, \cdots, n$, where $L^{\bar{y}(\cdot)}_{\bar{y}(0)\bar{y}(t)}$ is the parallel translation along the curve $\bar{y}(\cdot)$ and from $\bar{y}(0)$ to $\bar{y}(t)$, see [3, Section 2.2] for its detailed definition, and then $\nabla_{\dot{\bar{y}}(t)} e_i(\cdot) = 0$ and $\nabla_{\dot{\bar{y}}(t)} d_i(\cdot) = 0$ for $i = 1, \cdots, n$. We deduce from [3, (2.7)&(2.6)] that, $\{e_i(t)\}_{i=1}^n \subset T_{\bar{y}(t)}M$ is an orthonormal basis, and $\{d_i(t)\}_{i=1}^n \subset T^*_{\bar{y}(t)}M$ is the dual basis to it.

For $\epsilon \in [0, \hat{\epsilon}]$ and $t \in [0,T]$, it follows from "Step 1", [3, Lemma 2.1] and the definition of exponential map (see [3, Section 2.1]) that, there exists a unique geodesic

$$\beta(\theta;t) = \exp_{\bar{y}(t)}(\theta V_\epsilon(t)), \quad \forall\, \theta \in [0,1], \tag{3.9}$$

connecting $\beta(0;t) = \bar{y}(t)$ and $\beta(1;t) = y_\epsilon(t)$. For $\theta \in [0,1]$, denote by $L_{\bar{y}(t)\beta(\theta;t)} : T_{\bar{y}(t)}M \to T_{\beta(\theta;t)}M$ the parallel translation along the geodesic $\beta(\cdot;t)$ ($t \in [0,T]$ is fixed), from $\bar{y}(t)$ to $\beta(\theta;t)$. By [3, (2.6)] we know that $\{e_i(t)\}_{i=1}^n$ and $\{L_{\bar{y}(t)\beta(\theta;t)} e_i(t)\}_{i=1}^n$ are respectively orthonormal bases at $T_{\bar{y}(t)}M$ and $T_{\beta(\theta;t)}M$ for each $t \in [0,T]$ and $\theta \in [0,1]$. Thus, we can write

$$V_\epsilon(t) = \sum_{i=1}^n a_i^\epsilon(t) e_i(t),\ t \in [0,T], \tag{3.10}$$

where $\{a_1^\epsilon(t), \cdots, a_n^\epsilon(t)\}$ satisfies $\sum_{i=1}^n a_i^\epsilon(t)^2 = |V_\epsilon(t)|^2,\ \forall\, t \in [0,T]$. By the linearity of $L_{\bar{y}(t)\beta(\theta;t)} : T_{\bar{y}(t)}M \to T_{\beta(\theta;t)}M$ and [3, (2.8)] we have

$$L_{\bar{y}(t)\beta(\theta;t)} V_\epsilon(t) = \sum_{i=1}^n a_i^\epsilon(t) L_{\bar{y}(t)\beta(\theta;t)} e_i(t),\ |L_{\bar{y}(t)\beta(\theta;t)} V_\epsilon(t)| = |V_\epsilon(t)|,\ \forall\, t \in [0,T]. \tag{3.11}$$

Fix any $t \in [0,T]$ and $i \in \{1, \cdots, n\}$. By [3, Lemma 2.2], Newton-Leibniz formula and the exchange of integral order, we derive

$$\begin{array}{rl} & \langle \nabla_{\dot{\bar{y}}(t)} V_\epsilon, e_i(t) \rangle \\ = & \frac{d}{dt} \langle V_\epsilon, e_i(t) \rangle \\ = & -\frac{d}{dt}\Big(\frac{1}{2} \nabla_1 \rho^2(\bar{y}(t), y_\epsilon(t))(d_i(t)) \Big) \\ := & P_1^i(t) + \nabla_u f[t]\Big(d_i(t), \epsilon v(t) + \epsilon^2 \sigma_\epsilon(t)\Big) \\ & + \int_0^1 \nabla_u^2 f\{t,\tau\}_\epsilon \Big(d_i(t), \epsilon v(t) + \epsilon^2 \sigma_\epsilon(t), \epsilon v(t) + \epsilon^2 \sigma_\epsilon(t)\Big)(1-\tau)d\tau, \end{array} \tag{3.12}$$

where

$$\{t,\tau\}_\epsilon := (t, \bar{y}(t), \bar{u}(t) + \tau\epsilon v(t) + \tau\epsilon^2 \sigma_\epsilon(t)), \quad \forall\, t \in [0,T],\ \tau \in [0,1], \tag{3.13}$$

and

$$\begin{array}{rl} & P_1^i(t) \\ := & -\frac{1}{2}\Big\{ \nabla_2 \nabla_1 \rho^2(\bar{y}(t), y_\epsilon(t))(e_i(t), f(t, y_\epsilon(t), u_\epsilon(t))) \\ & -\nabla_2 \nabla_1 \rho^2(\bar{y}(t), \bar{y}(t))(e_i(t), f(t, \bar{y}(t), u_\epsilon(t))) \\ & +\nabla_1^2 \rho^2(\bar{y}(t), y_\epsilon(t))(e_i(t), f[t]) - \nabla_1^2 \rho^2(\bar{y}(t), \bar{y}(t))(d_i(t), f[t]) \Big\}. \end{array}$$

From the definitions of geodesic (see [3, (2.1)]) and parallel translation (see [3, Section 2.2]) and [3, Lemma 2.1], we obtain

$$\frac{\partial}{\partial\theta}\beta(\theta;t) = L_{\bar{y}(t)\beta(\theta;t)}\frac{\partial}{\partial\theta}\Big|_0\beta(\theta;t) = L_{\bar{y}(t)\beta(\theta;t)}V_\epsilon(t), \quad \theta\in[0,1]. \tag{3.14}$$

Applying [3, Lemma 2.2] and Newton-Leibniz formula to $P_1^i(t)$, we have

$$P_1^i(t) = I_1^i(t) + I_2^i(t) + I_3^i(t),$$

where

$$\begin{aligned} I_1^i(t) = &-\frac{1}{2}\int_0^1 \Big[\nabla_2\nabla_1^2\rho^2(\bar{y}(t),\beta(\theta;t))\Big(e_i(t), f[t], \frac{\partial}{\partial\theta}\beta(\theta;t)\Big) \\ &- \nabla_2\nabla_1^2\rho^2(\bar{y}(t),\beta(0;t))\Big(e_i(t), f[t], \frac{\partial}{\partial\theta}\Big|_0\beta(\theta;t)\Big)\Big]d\theta, \end{aligned}$$

$$\begin{aligned} I_2^i(t) = &-\frac{1}{2}\int_0^1 \Big[\nabla_2^2\nabla_1\rho^2(\bar{y}(t),\beta(\theta;t))\Big(e_i(t), f(t,\beta(\theta;t),u_\epsilon(t)), \frac{\partial}{\partial\theta}\beta(\theta;t)\Big) \\ &- \nabla_2^2\nabla_1\rho^2(\bar{y}(t),\beta(0;t))\Big(e_i(t), f(t,\beta(0;t),u_\epsilon(t)), \frac{\partial}{\partial\theta}\Big|_0\beta(\theta;t)\Big)\Big]d\theta, \end{aligned}$$

and

$$\begin{aligned} I_3^i(t) =& \nabla_x f[t](d_i(t),V_\epsilon(t)) + \int_0^1 \nabla_u\nabla_x f(t,\bar{y}(t),u_\epsilon^\theta(t))\Big(d_i(t),V_\epsilon(t),\epsilon v(t) \\ &+ \epsilon^2\sigma_\epsilon(t)\Big)d\theta - \frac{1}{2}\int_0^1 \Big[\nabla_2\nabla_1\rho^2(\bar{y}(t),\beta(\theta;t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\theta}\beta(\theta;t)} f(t,\cdot,u_\epsilon(t))\Big) \\ &- \nabla_2\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\theta}|_0\beta(\theta;t)} f(t,\cdot,u_\epsilon(t))\Big)\Big]d\theta, \end{aligned}$$

with $u_\epsilon^\theta(t) = \bar{u}(t) + \theta(\epsilon v(t) + \epsilon^2\sigma_\epsilon(t)), \theta\in[0,1]$. We use Newton-Leibniz formula again to the above three items, exchange the integration order, and get

$$I_1^i(t) = -\frac{1}{4}\nabla_2^2\nabla_1^2\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), f[t], V_\epsilon(t), V_\epsilon(t)\Big) + \hat{I}_1^i(t), \tag{3.15}$$

$$I_2^i(t) = -\frac{1}{4}\nabla_2^3\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), f[t], V_\epsilon(t), V_\epsilon(t)\Big) + \hat{I}_2^i(t), \tag{3.16}$$

and

$$\begin{aligned} & I_3^i(t) \\ = \; & \nabla_x f[t](e_i(t),V_\epsilon(t)) + \epsilon\nabla_u\nabla_x f[t](d_i(t),V_\epsilon(t),v(t)) \\ & -\tfrac{1}{4}\nabla_2\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\tau}|_0\beta(\tau;t)}\nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)} f(t,\cdot,\bar{u}(t))\Big) + \hat{I}_3^i(t), \end{aligned} \tag{3.17}$$

where we have used (3.10), (3.11), (3.14), and the fact that $\beta(\cdot;t)$ is a geodesic,

$$\begin{aligned} \hat{I}_1^i(t) = &-\frac{1}{2}\sum_{k,l=1}^n\int_0^1 \Big[\nabla_2^2\nabla_1^2\rho^2(\bar{y}(t),\beta(\theta;t))\Big(e_i(t), f[t], L_{\bar{y}(t)\beta(\theta;t)}e_k(t), L_{\bar{y}(t)\beta(\theta;t)}e_l(t)\Big) \\ &- \nabla_2^2\nabla_1^2\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), f[t], e_k(t), e_l(t)\Big)\Big](1-\theta)d\theta\, a_k^\epsilon(t)a_l^\epsilon(t), \end{aligned}$$

$$
\begin{aligned}
\hat{I}_2^i(t) =& -\frac{1}{4}\nabla_2^3\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), f(t,\bar{y}(t),u_\epsilon(t)) - f[t], V_\epsilon(t), V_\epsilon(t)\Big)\\
&-\sum_{k,l=1}^{n}\frac{1}{2}\int_0^1\Big[\nabla_2^3\nabla_1\rho^2(\bar{y}(t),\beta(\tau;t))\Big(e_i(t), f(t,\beta(\tau;t),u_\epsilon(t)),\\
&L_{\bar{y}(t)\beta(\theta;t)}e_k(t), L_{\bar{y}(t)\beta(\theta;t)}e_l(t)\Big) - \nabla_2^3\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), f(t,\bar{y}(t),u_\epsilon(t)),\\
&e_k(t), e_l(t)\Big)\Big]a_k^\epsilon(t)a_l^\epsilon(t)(1-\tau)d\tau - \frac{1}{2}\int_0^1\nabla_2^2\nabla_1\rho^2(\bar{y}(t),\beta(\tau;t))\Big(e_i(t),\\
&\nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}f(t,\cdot,u_\epsilon(t)), \frac{\partial}{\partial\tau}\beta(\tau;t)\Big)(1-\tau)d\tau,
\end{aligned}
$$

and

$$
\begin{aligned}
&\hat{I}_3^i(t)\\
=& -\frac{1}{4}\nabla_2\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\tau}|_0\beta(\tau;t)}\nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}[f(t,\cdot,\bar{u}(t)) - f(t,\cdot,u_\epsilon(t))]\Big)\\
&-\frac{1}{2}\int_0^1\Big[\nabla_2\nabla_1\rho^2(\bar{y}(t),\beta(\tau;t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}\nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}f(t,\cdot,u_\epsilon(t))\Big)\\
&-\nabla_2\nabla_1\rho^2(\bar{y}(t),\bar{y}(t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\tau}|_0\beta(\tau;t)}\nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}f(t,\cdot,u_\epsilon(t))\Big)\Big](1-\tau)d\tau\\
&+\epsilon\int_0^1\Big[\Big(\nabla_u\nabla_x f(t,\bar{y}(t),u_\epsilon^\theta(t)) - \nabla_u\nabla_x f[t]\Big)\Big(d_i(t), V_\epsilon(t), v(t)\Big)\\
&+\epsilon\nabla_u\nabla_x f(t,\bar{y}(t),u_\epsilon^\theta(t))\Big(d_i(t), V_\epsilon(t), \sigma_\epsilon(t)\Big)\Big]d\theta\\
&-\frac{1}{2}\int_0^1\nabla_2^2\nabla_1\rho^2(\bar{y}(t),\beta(\tau;t))\Big(e_i(t), \nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}f(t,\cdot,u_\epsilon(t)),\\
&\frac{\partial}{\partial\tau}\beta(\tau;t)\Big)(1-\tau)d\tau.
\end{aligned}
$$

Following the same argument as that used in [3, (5.18)], we have

$$
\begin{aligned}
&\langle Z(\beta(s;t)), \nabla_{\frac{\partial}{\partial\tau}|_s\beta(\tau;t)}\nabla_{\frac{\partial}{\partial\tau}\beta(\tau;t)}f(t,\cdot,u)\rangle\\
=\;&\nabla^2 f(t,\beta(s;t),u)\Big(\tilde{Z}, \frac{\partial}{\partial\tau}\Big|_s\beta(\tau;t), \frac{\partial}{\partial\tau}\Big|_s\beta(\tau;t)\Big),
\end{aligned}
\tag{3.18}
$$

where $(s,u,Z)\in[0,1]\times U\times TM$, and $\tilde{Z}$ is the dual covector of Z.

Step 3. Recall (2.19), (3.1), (3.12), (3.15) - (3.17). For $\epsilon\in[0,\hat{\epsilon}]$, by (3.18), [3, Lemma

2.2 & Lemma 2.3] and Newton-Leibniz formula, we get

$$
\begin{aligned}
& \langle V_\epsilon(t) - \epsilon X_v(t) - \epsilon^2 Y^{X_v}_{\sigma_\epsilon W}(t), e_i(t)\rangle \\
= & \int_0^t \langle \nabla_{\dot{y}(s)} V_\epsilon - \epsilon \nabla_{\dot{y}(s)} X_v - \epsilon^2 \nabla_{\dot{y}(s)} Y^{X_v}_{\sigma_\epsilon W}, e_i(s)\rangle ds \\
= & \int_0^t \Big\{ \nabla_x f[s]\Big(d_i(s), V_\epsilon(s) - \epsilon X_v(s) - \epsilon^2 Y^{X_v}_{\sigma_\epsilon W}(s)\Big) \\
& + \epsilon^2 \int_0^1 \Big[\nabla_u^2 f\{s,\tau\}_\epsilon\Big(d_i(s), v(s), v(s)\Big) - \nabla_u^2 f[s]\Big(d_i(s), v(s), v(s)\Big)\Big](1-\tau)d\tau \\
& - \tfrac{1}{2} R(e_i(s), V_\epsilon(s), f[s], V_\epsilon(s)) + \tfrac{\epsilon^2}{2} R(e_i(s), X_v(s), f[s], X_v(s)) \\
& + \tfrac{1}{2}\nabla_x^2 f[s]\Big(d_i(s), V_\epsilon(s), V_\epsilon(s)\Big) - \tfrac{\epsilon^2}{2}\nabla_x^2 f[s]\Big(d_i(s), X_v(s), X_v(s)\Big) \\
& + \epsilon \nabla_u \nabla_x f[s]\Big(d_i(s), V_\epsilon(s) - \epsilon X_v(s), v(s)\Big)\Big\} ds + \epsilon^2 P_2^i(t),
\end{aligned}
\tag{3.19}
$$

where

$$
\begin{aligned}
& P_2^i(t) \\
= & \int_0^t \Big\{ \int_0^1 \Big[2\epsilon \nabla_u^2 f\{s,\tau\}_\epsilon\Big(d_i(s), v(s), \sigma_\epsilon(s)\Big) \\
& + \epsilon^2 \nabla_u^2 f\{s,\tau\}_\epsilon\Big(d_i(s), \sigma_\epsilon(s), \sigma_\epsilon(s)\Big)\Big](1-\tau)d\tau + \frac{1}{\epsilon^2}\Big(\hat{I}_1^i(s) + \hat{I}_2^i(s) + \hat{I}_3^i(s)\Big)\Big\} ds.
\end{aligned}
$$

It follows from $(C2)-(C4)$ and (3.4) that $|P_1^i(t)|$ is bounded for $t \in [0,T]$. By applying Gronwall's inequality to (3.19), we obtain that, there exists a positive constant C such that

$$
|V_\epsilon(t) - \epsilon X_v(t)| \le C\epsilon^2, \ \forall\, t \in [0,T], \ \forall\, \epsilon \in [0,\hat{\epsilon}].
$$

Applying Gronwall's inequality again to (3.19), we obtain from the above inequality, Lebesgue's dominated convergence theorem, (3.4) and $(C2)-(C4)$ that, given any $\alpha > 0$, there exists $\epsilon^0 \in (0,\hat{\epsilon}]$ such that (3.2) holds. The proof is concluded. $\Box$

4 Proof of Theorem 2.2

In Section 4.1, we obtain the second order necessary condition of an optimization problem (problem (OP)), see Theorem 4.2. In Section 4.2, we transform problem (OCP) into an optimization problem, which is a special case of problem (OP), and prove Theorem 2.2 by Theorem 4.2.

4.1 An optimization problem

Let $\mathcal{X}$ be a Banach space, and $E \subset \mathcal{X}$ be a convex subset of it. Given maps $\hat{\phi}_i : \mathcal{X} \to \mathbb{R}$ with $i = 0, \cdots, j$, and $\hat{\psi} = (\hat{\psi}_1, \cdots, \hat{\psi}_k)^\top : \mathcal{X} \to \mathbb{R}^k$ $(k \in \mathbb{N} \cup \{0\})$, consider the following optimization problem.

(OP) Find $\bar{e} \in E$ such that it minimizes $\hat{\phi}_0(e)$ with $e \in E$ subject to

$$\begin{aligned} &\hat{\phi}_i(e) \leq 0,\ i = 1, \cdots, j, \\ &\hat{\psi}(e) = 0. \end{aligned} \tag{4.1}$$

$\bar{e}$ is called a solution or a minimizer of problem (OP).

Set $\hat{\Phi} : \mathcal{X} \to \mathbb{R}^{1+j+k}$ by $\hat{\Phi} = (\hat{\phi}_0, \hat{\phi}_1, \cdots, \hat{\phi}_j, \hat{\psi}_1, \cdots, \hat{\psi}_k)^\top$. Given any index set $I \subseteq \{0, 1, \cdots, j\}$, we denote by $\hat{\Phi}_I = (\bar{\phi}_0, \bar{\phi}_1, \cdots, \bar{\phi}_j, \hat{\psi}_1, \cdots, \hat{\psi}_k)^\top$, where $\bar{\phi}_i = \hat{\phi}_i$ if $i \in I$, and $\bar{\phi}_i = 0$ if $i \notin I$.

Assume $\bar{e} \in E$ is a minimizer of problem (OP). Set by

$$\begin{aligned} &\hat{I}_A \equiv \{i \in \{1, \cdots, j\}; \hat{\phi}_i(\bar{e}) = 0\} \cup \{0\}, \\ &\hat{I}_N \equiv \{i \in \{1, \cdots, j\}; \hat{\phi}_i(\bar{e}) < 0\}. \end{aligned}$$

To state the necessary condition of problem (OP), we introduce the following condition.

(C5) $\hat{\Phi}$ is Fréchet differentiable at $\bar{e} \in E$, and we denote its Fréchet derivative at $\bar{e} \in E$ by $D\hat{\Phi}(\bar{e}) = \big(D\hat{\phi}_0(\bar{e}), \cdots, D\hat{\phi}_j(\bar{e}), D\hat{\psi}_1(\bar{e}), \cdots, D\hat{\psi}_k(\bar{e})\big)^\top$, where $D\hat{\phi}_i(\bar{e})$ $(i = 0, 1, \cdots, j)$ and $D\hat{\psi}_l(\bar{e})$ $(l = 1, \cdots, k)$ are the Fréchet derivatives of $\hat{\phi}_i$ and $\hat{\psi}_l$ at $\bar{e}$ respectively. For any $x \in \mathcal{X}$,

$$D\hat{\Phi}(\bar{e})(x) \equiv \big(D\hat{\phi}_0(\bar{e})(x), \cdots, D\hat{\phi}_j(\bar{e})(x), D\hat{\psi}_1(\bar{e})(x), \cdots, D\hat{\psi}_k(\bar{e})(x)\big)^\top.$$

For each $y \in \mathcal{X}$, there exists

$$D^2\hat{\Phi}(\bar{e})(y) = \big(D^2\hat{\phi}_0(\bar{e})(y), \cdots, D^2\hat{\phi}_j(\bar{e})(y), D^2\hat{\psi}_1(\bar{e})(y), \cdots, D^2\hat{\psi}_k(\bar{e})(y)\big)^\top \in \mathbb{R}^{1+j+k},$$

such that the following relation holds: for any $\alpha > 0$ and $C > 0$, there exists $\epsilon_0 > 0$ depending on α and C, such that

$$|\hat{\Phi}(\bar{e} + \epsilon y + \epsilon^2 \eta) - \hat{\Phi}(\bar{e}) - \epsilon D\hat{\Phi}(\bar{e})(y) - \epsilon^2 D\hat{\Phi}(\bar{e})(\eta) - \frac{1}{2}\epsilon^2 D^2\hat{\Phi}(\bar{e})(y)| \leq \alpha\epsilon^2,$$

for all $\eta \in \mathcal{X}$ with $|\eta| \leq C$ and $\epsilon \in [0, \epsilon_0]$.

Lemma 4.1 *Assume that $\bar{e} \in E$ is a minimizer of (OP) with $\hat{\phi}_0(\bar{e}) = 0$, and that $(C5)$ holds. Let $y \in T_E^\flat(\bar{e})$ satisfy*

$$\begin{cases} D\hat{\phi}_i(\bar{e})(y) \leq 0,\ \textit{for}\ i \in \hat{I}_A, \\ D\hat{\psi}(\bar{e})(y) = 0, \quad T_E^{\flat(2)}(\bar{e}, y) \neq \emptyset, \end{cases} \tag{4.2}$$

where $D\hat{\psi}(\bar{e}) = \big(D\hat{\psi}_1(\bar{e}), \cdots, D\hat{\psi}_k(\bar{e})\big)^\top$ is a linear map from $\mathcal{X}$ to $\mathbb{R}^k$. Set

$$\begin{aligned} &\hat{I}_0' \equiv \hat{I}_N \cup \{i \in \hat{I}_A; D\hat{\phi}_i(\bar{e})(y) < 0\}, \\ &\hat{I}_0'' \equiv \{0, 1, \cdots, j\} \setminus \hat{I}_0'. \end{aligned}$$

Denote by

$$\mathcal{K} \equiv \{D\hat{\Phi}_{\hat{I}_0''}(\bar{e})(x) + \frac{1}{2}D^2\hat{\Phi}_{\hat{I}_0''}(\bar{e})(y); x \in T_E^{\flat(2)}(\bar{e}, y)\} \subset \mathbb{R}^{1+j+k} \tag{4.3}$$

and

$$\mathcal{K}^{\hat{\psi}} = \{D\hat{\psi}(\bar{e})(x) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y); x \in T_E^{\flat(2)}(\bar{e}, y)\} \subset \mathbb{R}^k.$$

Then $\mathcal{K}$ *and* $\mathcal{K}^{\hat{\psi}}$ *are both convex. Moreover, we set by* $Y \equiv (Y_0, Y_1, \cdots, Y_j)^\top$, *with*

$$Y_i = \begin{cases} D\hat{\phi}_i(\bar{e})(y), & i \in \hat{I}_A, \\ 0, & i \notin \hat{I}_A, \end{cases}$$

and by

$$Z \equiv (-\infty, 0)^{j+1} - \{\lambda(\hat{\phi}(\bar{e}) + Y);\ \lambda > 0\}, \tag{4.4}$$

where $\hat{\phi} = (\hat{\phi}_0, \hat{\phi}_1, \cdots, \hat{\phi}_j)^\top$. *If* $\mathcal{K}$ *and* $Z \times \{0\}$ *can not be seperated by any linear functional: there does not exist* $\ell \in \mathbb{R}^{1+j+k} \setminus \{0\}$ *such that*

$$\ell^\top(D\hat{\Phi}_{\hat{I}_0''}(\bar{e})(x) + \frac{1}{2}D^2\hat{\Phi}_{\hat{I}_0''}(\bar{e})(y)) \leq \ell^\top(z^\top, 0)^\top,\ \forall x \in T_E^{\flat(2)}(\bar{e}, y),\ \forall z \in Z, \tag{4.5}$$

then, aff$\mathcal{K}^{\hat{\psi}}$ *is a subspace of* $\mathbb{R}^k$, *where aff*$\mathcal{K}^{\hat{\psi}}$ *is the affine hull of* $\mathcal{K}^{\hat{\psi}}$ *(see [9, p.6]), and the dimension of aff*$\mathcal{K}^{\hat{\psi}}$ *is bigger than zero (see [9, p. 4]). We denote it by* $D(\mathcal{K}^{\hat{\psi}})$. *Moreover, there exist* $h_1, \cdots, h_{D(\mathcal{K}^{\hat{\psi}})+1} \in T_E^{\flat(2)}(\bar{e}, y)$ *and* $\delta_0 > 0$, *such that*

$$B_{aff\mathcal{K}^{\hat{\psi}}}(\delta_0) \subseteq Int\, co\{D\hat{\psi}(\bar{e})(h_l) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\}_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}, \tag{4.6}$$

$$D\hat{\phi}_i(\bar{e})(h_l) + \frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y) < 0,\ l = 1, \cdots, D(\mathcal{K}^{\hat{\psi}}) + 1,\ if\, i \in \hat{I}_0'', \tag{4.7}$$

where "$Int\, A$*" and "*$co\, A$*" are respectively the interior and the convex hull of a set* A, $B_{aff\mathcal{K}^{\hat{\psi}}}(\delta_0)$ *is the closed ball in subspace aff*$\mathcal{K}^{\hat{\psi}}$ *with center* $0 \in aff\mathcal{K}^{\hat{\psi}}$ *and radius* δ_0, *and* $D^2\hat{\psi}(\bar{e})(y) = \big(D^2\hat{\psi}_1(\bar{e})(y), \cdots, D^2\hat{\psi}_k(\bar{e})(y)\big)^\top$.

Proof. First, since E is convex, one can check by definition that $T_E^{\flat(2)}(\bar{e}, y)$ is convex, and consequently $\mathcal{K}$ and $\mathcal{K}^{\hat{\psi}}$ are convex.

Second, we claim that

$$0 \in \mathrm{ri}\, \mathcal{K}^{\hat{\psi}}, \tag{4.8}$$

where $\mathrm{ri}\, \mathcal{K}^{\hat{\psi}}$ is the interior of set $\mathcal{K}^{\hat{\psi}}$ relative to its affine hull (see [9, p.44]). Consequently, aff $\mathcal{K}^{\hat{\psi}}$ is a subspace of $\mathbb{R}^k$. By contradiction, we assume (4.8) were not true. Since the affine hull of $\mathcal{K}^{\hat{\psi}}$ is closed (see [9, p. 44]), and ri $\mathcal{K}^{\hat{\psi}}$ is not empty (by [9, Theorem 6.2, p. 45]), it follows from [2, Lemma 3.1] that, there exists $\xi \in \mathbb{R}^k \setminus \{0\}$ such that

$$\xi^\top(D\hat{\psi}(\bar{e})(x) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)) \leq 0, \quad \forall x \in T_E^{\flat(2)}(\bar{e}, y). \tag{4.9}$$

Consequently we have

$$(0,\xi^\top)(D\hat{\Phi}_{\hat{I}_0''}(\bar{e})(x)+\frac{1}{2}D^2\hat{\Phi}_{\hat{I}_0''}(\bar{e})(y))\le 0,\quad \forall x\in T_E^{\flat(2)}(\bar{e},y),\tag{4.10}$$

which contradicts the condition that $\mathcal{K}$ and $Z\times\{0\}$ are not separated by any linear functional.

Third, if $D(\mathcal{K}^{\hat{\psi}})\le 0$, then $D(\mathcal{K}^{\hat{\psi}})=0$, due to $\mathcal{K}^{\hat{\psi}}\neq\emptyset$. Consequently, we have $\mathcal{K}^{\hat{\psi}}=\{0\}$. Then, for any $\beta\in\mathbb{R}^k\setminus\{0\}$, (4.9) holds with ξ replaced by β, and consequently (4.10) holds with $(0,\xi^\top)$ replaced by $(0,\beta^\top)$. A contradiction follows.

Finally, there exist $x_1,\cdots,x_{D(\mathcal{K}^{\hat{\psi}})+1}\in T_E^{\flat(2)}(\bar{e},y)$ such that

$$0\in Int\,co\{D\hat{\psi}(\bar{e})(x_l)+\frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\}_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}.\tag{4.11}$$

According to [2, Lemma 3.1], $(Z\times\{0\})\cap\mathcal{K}\neq\emptyset$. Then, there exist $z_0,z_1,\cdots,z_j\in(-\infty,0)$, $\lambda>0$ and $\tilde{x}\in T_E^{\flat(2)}(\bar{e},y)$ such that

$$\begin{array}{ll} z_i=D\hat{\phi}_i(\bar{e})(\tilde{x})+\frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y), & \text{if } \hat{\phi}_i(\bar{e})=0,\ D\hat{\phi}_i(\bar{e})(y)=0;\\ z_i-\lambda\hat{\phi}_i(\bar{e})=0, & \text{if } \hat{\phi}_i(\bar{e})<0;\\ z_i-\lambda D\hat{\phi}_i(\bar{e})(y)=0, & \text{if } \hat{\phi}_i(\bar{e})=0,\ D\hat{\phi}_i(\bar{e})(y)<0;\end{array}\tag{4.12}$$

and $D\hat{\psi}(\bar{e})(\tilde{x})+\frac{1}{2}D^2\hat{\psi}(\bar{e})(y)=0$. Let $\eta\in(0,1)$ be such that

$$\begin{array}{rcl} 0 & > & (1-\eta)\Big(D\hat{\phi}_i(\bar{e})(x_l)+\frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big)+\eta\Big(D\hat{\phi}_i(\bar{e})(\tilde{x})+\frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big)\\ & = & D\hat{\phi}_i(\bar{e})((1-\eta)x_l+\eta\tilde{x})+\frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y),\end{array}$$

for all $l=1,\cdots,D(\mathcal{K}^{\hat{\psi}})+1$ and $i\in\{0,1,\cdots,j\}$ satisfying $\hat{\phi}_i(\bar{e})=0$ and $D\hat{\phi}_i(\bar{e})(y)=0$. Set $h_l=\eta\tilde{x}+(1-\eta)x_l$, for $l=1,\cdots,k+1$. Then (4.7) follows, and (4.6) follows from (4.11). $\Box$

The following results are respectively the first and second order necessary conditions of an minimizer of (OP), and the idea of proving them partially comes from [7, Theorem 4.1].

Theorem 4.1 *Assume that $(C5)$ holds, and that $\bar{e}\in E$ is a solution to problem (OP). Then, there exists $(\ell_{\hat{\phi}_0},\cdots,\ell_{\hat{\phi}_j},\ell_{\hat{\psi}}^\top)^\top\in\mathbb{R}^{1+j+k}\setminus\{0\}$ such that*

$$\ell_{\hat{\phi}_i}\in(-\infty,0],\ i=0,\cdots,j;\quad \ell_{\hat{\phi}_i}=0,\ \textit{if}\ i\in\hat{I}_N;\quad \sum_{i=0}^{j}\ell_{\hat{\phi}_i}\hat{\phi}_i(\bar{e})=0;\tag{4.13}$$

$$\sum_{i\in\hat{I}_A}\ell_{\hat{\phi}_i}D\hat{\phi}_i(\bar{e})(x)+\ell_{\hat{\psi}}^\top D\hat{\psi}(\bar{e})(x)\le 0,\ \forall x\in T_E^{\flat}(\bar{e}).\tag{4.14}$$

Theorem 4.2 *Assume that* $(C5)$ *holds, that* $\bar{e} \in E$ *is a solution to problem* (OP) *with* $\hat{\phi}_0(\bar{e}) = 0$*, and that* $y \in T_E^\flat(\bar{e})$ *satisfies (4.2). Then, there exists* $(\ell_{\hat{\phi}_0}, \ell_{\hat{\phi}_1}, \cdots, \ell_{\hat{\phi}_j}, \ell_{\hat{\psi}}^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ *satisfying (4.13), (4.14),*

$$\ell_{\hat{\phi}_i} = 0, \ \text{if}\, i \notin \hat{I}_0'', \tag{4.15}$$

and

$$\begin{array}{l} \sum_{i \in \hat{I}_0''} \ell_{\hat{\phi}_i} D\hat{\phi}_i(\bar{e})(x) + \ell_{\hat{\psi}}^\top D\hat{\psi}(\bar{e})(x) + \sum_{i \in \hat{I}_0''} \frac{1}{2}\ell_{\hat{\phi}_i} D^2\hat{\phi}_i(\bar{e})(y) \\ +\frac{1}{2}\ell_{\hat{\psi}}^\top D^2\hat{\psi}(\bar{e})(y) \le 0, \quad \forall\, x \in T_E^{\flat(2)}(\bar{e}, y). \end{array} \tag{4.16}$$

Remark 4.1 *If* $y \in T_E^\flat(\bar{e})$ *satisfies (4.2), it is easy to see that the first order necessary condition becomes trivial along the direction* y*: for any* $\ell = (\ell_{\hat{\phi}_0}, \cdots, \ell_{\hat{\phi}_j}, \ell_{\hat{\psi}}^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ *satisfying (4.13) and (4.14), it holds that* $\sum_{i \in \hat{I}_A} \ell_{\hat{\phi}_i} D\hat{\phi}_i(\bar{e})(y) + \ell_{\hat{\psi}}^\top D\hat{\psi}(\bar{e})(y) = 0$*. Thus Theorem 4.2 gives further information of* $\bar{e}$ *along direction* y*. When* $\bar{e} \in Int\,E$*,* $0 \in T_E^{\flat(2)}(\bar{e}, y)$*, and consequently Theorem 4.2 is consistent with [7, Theorem 4.1]. When* $\bar{e}$ *is on the boundary of* E*,* $0 \in T_E^{\flat(2)}(\bar{e}, y)$ *is not always true, thus, compared to [7, Theorem 4.1], the first two terms of the left hand side of (4.16) are extra terms.*

Since the proof of Theorem 4.1 is analogous to that of Theorem 4.2, we only prove Theorem 4.2 and give the key point of proving Theorem 4.1: The set $\{D\hat{\Phi}_{\hat{I}_A}(\bar{e})(x) | x \in T_E^\flat(\bar{e})\}$ is separated from $\Big((-\infty, 0)^{j+1} - \{\lambda(\hat{\phi}(\bar{e}); \lambda > 0\}\Big) \times \{0\}$.

Proof of Theorem 4.2

Step 1. We shall prove the case that $k > 0$.

First, we claim that there exists $\ell \equiv (\ell_{\hat{\phi}_0}, \cdots, \ell_{\hat{\phi}_j}, \ell_{\hat{\psi}}^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ such that (4.5) holds.

By contradiction, it follows from Lemma 4.1 that (4.6) and (4.7) hold. Fix h_l ($l = 1, \cdots, D(\mathcal{K}^{\hat{\psi}}) + 1$). Recall the definition of the second-order adjacent set (see Section 2.1). For any $\epsilon \to 0^+$, there exists $h_l^\epsilon \to h_l$ as $\epsilon \to 0^+$ such that $\bar{e} + \epsilon y + \epsilon^2 h_l^\epsilon \in E$. Then, there exists $\epsilon_0 > 0$ such that $\bar{e} + \epsilon y + \epsilon^2 h_l^\epsilon \in E$ for all $\epsilon \in [0, \epsilon_0]$ and $l = 1, \cdots, D(\mathcal{K}^{\hat{\psi}}) + 1$. By the convexity of E, for any $x = \sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1} \nu_l h_l \in co\{h_1, \cdots, h_{D(\mathcal{K}^{\hat{\psi}})+1}\}$ with $(\nu_1, \cdots, \nu_{D(\mathcal{K}^{\hat{\psi}})+1})$ satisfying

$$\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1} \nu_l = 1; \quad \nu_l \ge 0 \quad \text{for } l = 1, \cdots, D(\mathcal{K}^{\hat{\psi}}) + 1, \tag{4.17}$$

it holds that $\bar{e} + \epsilon y + \epsilon^2 \sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1} \nu_l h_l^\epsilon \in E$, for all $\epsilon \in [0, \epsilon_0]$.

Relation (4.2) and $(C5)$ imply that exists $\epsilon_1 \in (0, \epsilon_0]$ such that

$$\begin{aligned}
& \Big|\epsilon^{-2}\hat{\psi}(\bar{e}+\epsilon y+\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) - D\hat{\psi}(\bar{e})(\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l) - \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\Big| \\
\leq & \Big|\textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l D\hat{\psi}(\bar{e})(h_l^{\epsilon}-h_l)\Big| \\
& +\Big|\epsilon^{-2}\hat{\psi}(\bar{e}+\epsilon y+\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) - D\hat{\psi}(\bar{e})(\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) - \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\Big| \\
< & \ \delta_0, \ \forall \epsilon \in [0, \epsilon_1],
\end{aligned}$$

where $\nu_1, \cdots, \nu_{D(\mathcal{K}^{\hat{\psi}})+1}$ satisfy (4.17).

According to (4.2), (4.7) and $(C5)$, one can find $\epsilon_2 \in (0, \epsilon_1]$ such that, for any $\epsilon \in [0, \epsilon_2]$, the following relations hold: for $i \in \hat{I}_0''$,

$$\begin{aligned}
& \hat{\phi}_i(\bar{e}+\epsilon y+\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) \\
= & \ \epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l \Big(D\hat{\phi}_i(\bar{e})(h_l) + \frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big) \\
& +\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l D\hat{\phi}_i(\bar{e})(h_l^{\epsilon}-h_l) + \Big[\hat{\phi}_i(\bar{e}+\epsilon y+\epsilon^2 \sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) \\
& -\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l \Big(D\hat{\phi}_i(\bar{e})(h_l^{\epsilon}) + \frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big)\Big] \\
< & \ 0;
\end{aligned} \tag{4.18}$$

for $i \notin \hat{I}_0''$,

$$\begin{aligned}
& \hat{\phi}_i(\bar{e}+\epsilon y+\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) \\
= & \ \hat{\phi}_i(\bar{e}) + \epsilon D\hat{\phi}_i(\bar{e})(y) + \epsilon^2 \textstyle\sum_{l=1}^{D(K^{\hat{\psi}})+1}\nu_l \Big(D\hat{\phi}_i(\bar{e})h_l^{\epsilon} + \frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big) \\
& +\Big[\hat{\phi}_i(\bar{e}+\epsilon y+\epsilon^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon}) - \hat{\phi}_i(\bar{e}) - \epsilon D\hat{\phi}_i(\bar{e})(y) \\
& -\epsilon^2 \textstyle\sum_{l=1}^{D(K^{\hat{\psi}})+1}\nu_l \Big(D\hat{\phi}_i(\bar{e})h_l^{\epsilon} + \frac{1}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big)\Big] \\
< & \ 0.
\end{aligned} \tag{4.19}$$

Then, from the above relations and (4.6), we can define a map

$$\begin{aligned}
G : co\{D\hat{\psi}(\bar{e})(h_l) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\}_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1} \to \\
co\{D\hat{\psi}(\bar{e})(h_l) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\}_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}
\end{aligned}$$

by

$$\begin{aligned}
& G\Big(D\hat{\psi}(\bar{e})(\textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\Big) \\
= & \ -\epsilon_2^{-2}\hat{\psi}(\bar{e}+\epsilon_2 y+\epsilon_2^2 \textstyle\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l^{\epsilon_2}) + D\hat{\psi}(\bar{e})(\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l h_l) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y),
\end{aligned}$$

for all $\nu_1, \cdots, \nu_{D(\mathcal{K}^{\hat{\psi}})+1}$ satisfying (4.17). Obviously G is continuous and $co\{D\hat{\psi}(\bar{e})(h_l) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\}_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}$ is convex and compact. By Brouwer fixed point theorem, there exists $\nu_1^{\star}, \cdots, \nu_{D(\mathcal{K}^{\hat{\psi}})+1}^{\star}$ satisfying (4.17) such that $G\Big(D\hat{\psi}(\bar{e})\Big(\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l^{\star}h_l\Big) + \frac{1}{2}D^2\hat{\psi}(\bar{e})(y)\Big) =$

$D\hat{\psi}(\bar{e})(\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l^{\star}h_l)+\frac{1}{2}D^2\hat{\psi}(\bar{e})(y)$, which implies $\hat{\psi}(\bar{e}+\epsilon_2 y+\epsilon_2^2\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l^{\star}h_l^{\epsilon_2})=0$. Recalling (4.18) and (4.19), we obtain that $\bar{e}+\epsilon_2 y+\epsilon_2^2\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l^{\star}h_l^{\epsilon_2}$ satisfies (4.1) and $\hat{\phi}_0(\bar{e}+\epsilon_2 y+\epsilon_2^2\sum_{l=1}^{D(\mathcal{K}^{\hat{\psi}})+1}\nu_l^{\star}h_l^{\epsilon_2})<0$, which contradicts the optimality of $\bar{e}$.

Second, from (4.5) and the special structure of (4.4), we obtain (4.13), (4.15), and

$$\ell^{\top}(D\hat{\Phi}_{\hat{I}_0''}(\bar{e})(x)+\frac{1}{2}D^2\hat{\Phi}_{\hat{I}_0''}(\bar{e})(y))\leq\inf_{z\in Z}\ell^{\top}(z^{\top},0)^{\top}=0,\ \forall x\in T_E^{\flat(2)}(\bar{e},y),$$

which implies (4.16).

Finally, from [1, Proposition 4.2.1] and [6, Lemma 2.4] we have

$$T_E^{\flat(2)}(\bar{e},y)=T_E^{\flat(2)}(\bar{e},y)+T_E^{\flat}(\bar{e}). \tag{4.20}$$

If (4.14) were not true, there would exist $x_0\in T_E^{\flat}(\bar{e})$ such that $\sum_{i\in\hat{I}_A}\ell_{\hat{\phi}_i}D\hat{\phi}_i(\bar{e})(x_0)+\ell_{\hat{\psi}}^{\top}D\hat{\psi}(\bar{e})(x_0)>0$. Fix any $\sigma\in T_E^2(\bar{e},y)$. Choosing $\lambda>0$ big enough such that

$$\begin{aligned}&\textstyle\sum_{i\in\hat{I}_0''}\ell_{\hat{\phi}_i}D\hat{\phi}_i(\bar{e})(\sigma+\lambda x_0)+\ell_{\hat{\psi}}^{\top}D\hat{\psi}(\bar{e})(\sigma+\lambda x_0)+\sum_{i\in\hat{I}_0''}\frac{1}{2}\ell_{\hat{\phi}_i}D^2\hat{\phi}_i(\bar{e})(y)\\&+\tfrac{1}{2}\ell_{\hat{\psi}}^{\top}D^2\hat{\psi}(\bar{e})(y)>0,\end{aligned}$$

which contradicts (4.16), and the proof is concluded.

Step 2. For the case $k=0$, there exists $(\ell_{\hat{\phi}_0},\cdots,\ell_{\hat{\phi}_j})^{\top}\in\mathbb{R}^{1+j}\setminus\{0\}$ such that

$$\sum_{l=0}^{j}\ell_{\hat{\phi}_l}\beta_l\leq\sum_{l=0}^{j}\ell_{\hat{\phi}_l}z_l,\quad\forall(\beta_0,\cdots,\beta_j)^{\top}\in\mathcal{K},\ (z_0,\cdots,z_j)^{\top}\in Z.$$

If it were not true, by [2, Lemma 3.1] we would have $\mathcal{K}\cap Z\neq\emptyset$. Then, there exists $\tilde{x}\in T_E^{\flat(2)}(\bar{e},y)$, $\lambda>0$ and $(z_0,\cdots,z_j)^{\top}\in(-\infty,0)^{1+j}$ such that (4.12) holds. Thus, for any $i=0,\cdots,j$, we obtain from $(C5)$ and (4.12) that, there exists $\tilde{\epsilon}>0$ such that, for any $\epsilon\in[0,\tilde{\epsilon}]$ the following relation

$$\begin{aligned}&\hat{\phi}_i(\bar{e}+\epsilon y+\epsilon^2\tilde{x})\\=&\hat{\phi}_i(\bar{e})+\epsilon D\hat{\phi}_i(\bar{e})(y)+\epsilon^2D\hat{\phi}_i(\bar{e})(\tilde{x})+\frac{\epsilon^2}{2}D^2\hat{\phi}_i(\bar{e})(y)+\Big[\hat{\phi}_i(\bar{e}+\epsilon y+\epsilon^2\tilde{x})\\&-\hat{\phi}_i(\bar{e})-\epsilon D\hat{\phi}_i(\bar{e})(y)-\epsilon^2D\hat{\phi}_i(\bar{e})(\tilde{x})-\frac{\epsilon^2}{2}D^2\hat{\phi}_i(\bar{e})(y)\Big]\\<&0,\end{aligned}$$

holds for $i=0,1,\cdots,j$, which contradicts the optimality of $\bar{e}$. □

4.2 Proof of Theorem 2.2

We need the following lemmas.

Lemma 4.2 *Assume $U \subset \mathbb{R}^m$ is closed. Fix $\bar{u}(\cdot) \in \mathcal{U}$. Let $v(\cdot) \in L^1(0,T;\mathbb{R}^m)$ be such that $v(t) \in T_U^\flat(\bar{u}(t))$ a.e. $t \in [0,T]$. Assume there exist a positive constant ϵ_0 and $\ell(\cdot) \in L^i(0,T;\mathbb{R}^m)$ ($i = 1$ or 2) such that (2.21) holds. Fix any $\sigma(\cdot) \in L^1(0,T;\mathbb{R}^m)$ such that $\sigma(t) \in T_U^{\flat(2)}(\bar{u}(t), v(t))$ a.e. $t \in [0,T]$. Then, for any $\epsilon \in (0,\epsilon_0]$, there exists $\sigma_\epsilon(\cdot) \in L^i(0,T;\mathbb{R}^m)$ such that $u_\epsilon(t) := \bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma_\epsilon(t) \in U$ and $\lim_{\epsilon \to 0^+} \sigma_\epsilon(t) = \sigma(t)$ a.e. $t \in [0,T]$, and*

$$\|\sigma_\epsilon(\cdot)\|_{L^i(0,T;\mathbb{R}^m)} \leq \|\ell\|_{L^i(0,T;\mathbb{R}^m)} + 2\|\sigma\|_{L^i(0,T;\mathbb{R}^m)}. \tag{4.21}$$

Proof Since U is closed, by [1, Corollary 8.2.13, p. 317], for every $\epsilon > 0$, there exist measurable functions $\omega_\epsilon, z_\epsilon : [0,T] \to U$ such that

$$dist_U(\bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma(t)) = |\bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma(t) - \omega_\epsilon(t)| \quad a.e.\, t \in [0,T], \tag{4.22}$$

$$a_\epsilon(t) := dist_U(\bar{u}(t) + \epsilon v(t)) = |\bar{u}(t) + \epsilon v(t) - z_\epsilon(t)| \quad a.e.\, t \in [0,T]. \tag{4.23}$$

Set $\sigma_\epsilon(t) = \frac{1}{\epsilon^2}\big(\omega_\epsilon(t) - \bar{u}(t) - \epsilon v(t)\big)$ for $t \in [0,T]$. Then, we have $\bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma_\epsilon(t) \in U$ a.e $t \in [0,T]$. Since $\sigma(t) \in T_U^{\flat(2)}(\bar{u}(t), v(t))$ a.e. $t \in [0,T]$, recalling (4.22) and (2.4), we have

$$0 = \lim_{\epsilon \to 0^+} \frac{1}{\epsilon^2} |\bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma(t) - \omega_\epsilon(t)| = \lim_{\epsilon \to 0^+} |\sigma_\epsilon(t) - \sigma(t)|, \quad a.e.\, t \in [0,T]. \tag{4.24}$$

Applying (4.22), (4.23) and (2.21), we have, for $\epsilon \in (0,\epsilon_0]$,

$$\begin{aligned}
&\epsilon^2(|\sigma_\epsilon(t)| - |\sigma(t)|) \leq |\epsilon^2 \sigma(t) - \epsilon^2 \sigma_\epsilon(t)| = |\bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma(t) - \omega_\epsilon(t)| \\
&\leq |\bar{u}(t) + \epsilon v(t) + \epsilon^2 \sigma(t) - z_\epsilon(t)| \leq a_\epsilon(t) + \epsilon^2 |\sigma(t)| \leq \epsilon^2(\ell(t) + |\sigma(t)|), \quad a.e.\, t \in [0,T],
\end{aligned}$$

which implies that, $\sigma_\epsilon(\cdot) \in L^i(0,T;\mathbb{R}^m)$ and (4.21), if $\ell(\cdot) \in L^i(0,T;\mathbb{R}^m)$ ($i = 1,2$). $\square$

Lemma 4.3 *Assume $U \subset \mathbb{R}^m$ is closed. Fix $\bar{u}(\cdot) \in L^2(0,T;\mathbb{R}^m) \cap \mathcal{U}$. Let $v(\cdot) \in L^2(0,T;\mathbb{R}^m)$ be such that $v(t) \in T_U^\flat(\bar{u}(t))$ a.e. in $[0,T]$. Then, $v(\cdot) \in T^\flat_{L^2(0,T;\mathbb{R}^m)\cap\mathcal{U}}(\bar{u}(\cdot))$. Moreover, for any $\sigma(\cdot) \in L^2(0,T;\mathbb{R}^m)$ such that $\sigma(t) \in T_U^{\flat(2)}(\bar{u}(t), v(t))$ a.e. in $[0,T]$ and (2.21) holds for some $\ell(\cdot) \in L^2(0,T;\mathbb{R}^m)$ and $\epsilon_0 > 0$, it holds that $\sigma(\cdot) \in T^{\flat(2)}_{\mathcal{U}\cap L^2(0,T;\mathbb{R}^m)}(\bar{u}(\cdot), v(\cdot))$.*

Proof. First, by [1, Corollary 8.2.13, p. 317], for each $\epsilon > 0$, there exists a measurable map $v_\epsilon : [0,T] \to \mathbb{R}^m$ such that $v_\epsilon(t) \in \frac{U - \{\bar{u}(t)\}}{\epsilon}$ and $dist_{\frac{U-\{\bar{u}(t)\}}{\epsilon}} v(t) = |v(t) - v_\epsilon(t)|$ for almost every $t \in [0,T]$. It follows from (2.3) that $\lim_{\epsilon \to 0^+} |v(t) - v_\epsilon(t)| = 0$ a.e. $t \in [0,T]$. Then, we have

$$|v_\epsilon(t)| - |v(t)| \leq |v_\epsilon(t) - v(t)| \leq |v(t) - \frac{1}{\epsilon}(\bar{u}(t) - \bar{u}(t))| = |v(t)|.$$

We obtain from Lebesgue's dominated convergence theorem that $\lim_{\epsilon \to 0^+} v_\epsilon(\cdot) = v(\cdot)$ in $L^2(0,T;\mathbb{R}^m)$, consequently we have $v(\cdot) \in T^\flat_{\mathcal{U}\cap L^2(0,T;\mathbb{R}^m)}(\bar{u}(\cdot))$.

Then, it follows from Lemma 4.2 that, for any $\epsilon > 0$, there exists $\sigma_\epsilon(\cdot) \in L^2(0,T;\mathbb{R}^m)$ such that $\lim_{\epsilon\to 0^+} \sigma_\epsilon(\cdot) = \sigma(\cdot)$ in $L^2(0,T;\mathbb{R}^m)$ and $\bar{u}(\cdot)+\epsilon v(\cdot)+\epsilon^2\sigma_\epsilon(\cdot) \in \mathcal{U}\cap L^2(0,T;\mathbb{R}^m)$, and consequently $\sigma(\cdot) \in T^{\flat(2)}_{\mathcal{U}\cap L^2(0,T;\mathbb{R}^m)}(\bar{u}(\cdot), v(\cdot))$. The proof is concluded. $\square$

Then, we are going to prove Theorem 2.2.

Proof of Theorem 2.2 First, we shall transform problem (OCP) to an optimization problem. Assume $(\bar{u}(\cdot), \bar{y}(\cdot))$ is an optimal pair for problem (OCP) with $\bar{u}(\cdot) \in L^2(0,T;\mathbb{R}^m)$. For any $(Y, u(\cdot)) \in T_{\bar{y}(0)}M \times L^2(0,T;\mathbb{R}^m)$, set

$$\begin{aligned}
&\hat{\phi}_i(Y,u(\cdot)) \equiv \phi_i\left(y_u(0; exp_{\bar{y}(0)}Y), y_u(T; exp_{\bar{y}(0)}Y)\right),\ i = 1,\cdots,j,\\
&\hat{\phi}_0(Y,u(\cdot)) \equiv \phi_0\left(y_u(0; exp_{\bar{y}(0)}Y), y_u(T; exp_{\bar{y}(0)}Y)\right) - \phi_0(\bar{y}(0),\bar{y}(T)),\\
&\hat{\psi}(Y,u(\cdot)) \equiv \psi\left(y_u(0; exp_{\bar{y}(0)}Y), y_u(T; exp_{\bar{y}(0)}Y)\right),
\end{aligned} \tag{4.25}$$

where $y_u(\cdot; x)$ is the solution to (1.2) with initial state $x \in M$ and control $u(\cdot)$, and $exp_x\cdot$ is the exponential map at x (see Section [3, Section 2.1]).

We obtain from the optimality of $(\bar{u}(\cdot), \bar{y}(\cdot))$ for problem (OCP) that, $(0, \bar{u}(\cdot)) \in T_{\bar{y}(0)}M \times \mathcal{U}$ is the solution to the following optimization problem

$(\widetilde{OCP})$ Find $(Y, u(\cdot)) \in T_{\bar{y}(0)}M \times \left(L^2(0,T;\mathbb{R}^m) \cap \mathcal{U}\right)$ minimizes $\hat{\phi}_0(Y,u(\cdot))$ subject to $\hat{\phi}_i(Y,u(\cdot)) \le 0$ for $i = 1,\cdots,j$, $\hat{\psi}(Y,u(\cdot)) = 0$ and $(Y,u(\cdot)) \in T_{\bar{y}(0)}M \times (\mathcal{U} \cap L^2(0,T;\mathbb{R}^m))$.

Second, we shall check that condition $(C5)$ holds. Fix $(V, v(\cdot)) \in T_{\bar{y}(0)}M \times L^2(0,T;\mathbb{R}^m)$. For $\epsilon > 0$, we denote by $y(\cdot; \exp_{\bar{y}(0)} \epsilon V, \bar{u}(\cdot) + \epsilon v(\cdot))$ the solution to (1.2) corresponding to the initial state $\exp_{\bar{y}(0)} \epsilon V$ and the control $\bar{u}(\cdot) + \epsilon v(\cdot)$. For $i = 0, 1, \cdots, j$, we obtain from Proposition 3.1 that

$$\begin{aligned}
&\hat{\phi}_i(\epsilon V, \bar{u}(\cdot) + \epsilon v(\cdot)) - \hat{\phi}_i(0, \bar{u}(\cdot))\\
=&\epsilon\Big[\nabla_1\phi_i(\bar{y}(0),\bar{y}(T))(V) + \nabla_2\phi_i(\bar{y}(0),\bar{y}(T))(X_{v,V}(T))\Big] + o(\epsilon),
\end{aligned}$$

where $X_{v,V}(\cdot)$ is the solution to (2.19) with $X_{v,V}(0) = V$. This implies that $\hat{\phi}_i$ is Fréchet differentiable at $(0, \bar{u}(\cdot))$, and its Fréchet derivative is as follows

$$D\hat{\phi}_i(0,\bar{u}(\cdot))(V,v(\cdot)) = \nabla_1\phi_i(\bar{y}(0),\bar{y}(T))(V) + \nabla_2\phi_i(\bar{y}(0),\bar{y}(T))(X_{v,V}(T)). \tag{4.26}$$

Similarly we can show that $\hat{\psi}$ is Fréchet differentiable at $(0, \bar{u}(\cdot))$, and its Fréchet derivative is given by

$$D\hat{\psi}(0,\bar{u}(\cdot))(V,v(\cdot)) = \nabla_1\psi(\bar{y}(0),\bar{y}(T))(V) + \nabla_2\psi(\bar{y}(0),\bar{y}(T))(X_{v,V}(T)), \tag{4.27}$$

where $\nabla_i\psi$ $(i = 1, 2)$ is defined in (2.1).

Fix any $(W, \sigma(\cdot)) \in T_{\bar{y}(0)}M \times L^2(0,T;\mathbb{R}^m)$. Denote by $X_{\sigma,W}(\cdot)$ the solution to (2.19) with $v(\cdot)$ replaced by $\sigma(\cdot)$ and $X_{\sigma,W}(0) = W$. For any $\epsilon > 0$, we denote by $y_\epsilon(\cdot)$ the solution to (1.2) with initial state $\exp_{\bar{y}(0)}(\epsilon V+\epsilon^2 W)$ and control $\bar{u}(\cdot)+\epsilon v(\cdot)+\epsilon^2\sigma(\cdot)$. Denote by $Y_{\sigma W}^{X_{v,V}}(\cdot)$ the solution to (3.1) with $(\sigma_\epsilon(\cdot), X_v(\cdot))$ replaced by $(\sigma(\cdot), X_{v,V}(\cdot))$. We employ the notations $V_\epsilon(\cdot)$ and $\beta(\cdot;t)$ ($t\in[0,T]$) given respectively by (3.3) and (3.9). Note that (3.14) still holds. Fix $\alpha>0$. It follows from Proposition 3.1 that, there exists $\epsilon_1>0$ such that

$$V_\epsilon(t)=\epsilon X_{v,V}(t)+\epsilon^2 Y_{\sigma W}^{X_{v,V}}(t)+\gamma_\epsilon(t),\ \forall\, t\in[0,T],\ \epsilon\in[0,\epsilon_1], \tag{4.28}$$

with

$$|\gamma_\epsilon(t)|\le \frac{\alpha}{2K}\epsilon^2,\quad \forall\,\epsilon\in[0,\epsilon_1], \tag{4.29}$$

where constant K is given in condition $(C2)$.

Set by

$$\begin{aligned}
&D^2\hat{\phi}_i(0,\bar{u}(\cdot))(V,v(\cdot))\\
=\ &\nabla_1^2\phi_i(\bar{y}(0),\bar{y}(T))(V,V)+2\nabla_2\nabla_1\phi_i(\bar{y}(0),\bar{y}(T))(V,X_{v,V}(T))\\
&+\nabla_2^2\phi_i(\bar{y}(0),\bar{y}(T))(X_{v,V}(T),X_{v,V}(T))+2\nabla_2\phi_i(\bar{y}(0),\bar{y}(T))(Y_{00}^{X_{v,V}}(T)),
\end{aligned} \tag{4.30}$$

where $Y_{00}^{X_{v,V}}(\cdot)$ is the solution to (3.1) with $\sigma(\cdot)=0$ and $W=0$, and with $X_v(\cdot)$ replaced by $X_{v,V}(\cdot)$. It is easy to check that

$$Y_{\sigma W}^{X_{v,V}}(t)=Y_{00}^{X_{v,V}}(t)+X_{\sigma,W}(t),\quad \forall\, t\in[0,T], \tag{4.31}$$

where $X_{\sigma,W}(\cdot)$ is the solution to (2.19) with initial state $X_{\sigma,W}(0)=W$, and with $v(\cdot)$ replaced by $\sigma(\cdot)$.

For $i=0,1,\cdots,j$, we obtain by Newton-Leibniz formula, exchange of integral variables, (4.28) and (4.31) that

$$\begin{aligned}
&\hat{\phi}_i(\epsilon V+\epsilon^2 W,\bar{u}(\cdot)+\epsilon v(\cdot)+\epsilon^2\sigma(\cdot))-\hat{\phi}_i(0,\bar{u}(\cdot))-\epsilon D\hat{\phi}_i(0,\bar{u}(\cdot))(V,v(\cdot))\\
&-\epsilon^2\Big[D\hat{\phi}_i(0,\bar{u}(\cdot))(W,\sigma(\cdot))+\frac{1}{2}D^2\hat{\phi}_i(0,\bar{u}(\cdot))(V,v(\cdot))\Big]=L_i^\epsilon,
\end{aligned}$$

where

$$\begin{aligned}
L_i^\epsilon=\int_0^1&\Big[\nabla_1^2\phi_i(\beta(\tau;0),\beta(\tau;T))\Big(\frac{\partial}{\partial\tau}\beta(\tau;0),\frac{\partial}{\partial\tau}\beta(\tau;0)\Big)-\nabla_1^2\phi_i(\bar{y}(0),\bar{y}(T))(V,V)\epsilon^2\\
&+2\nabla_2\nabla_1\phi_i(\beta(\tau;0),\beta(\tau;T))\Big(\frac{\partial}{\partial\tau}\beta(\tau;0),\frac{\partial}{\partial\tau}\beta(\tau;T)\Big)-2\epsilon^2\nabla_2\nabla_1\phi_i(\bar{y}(0),\bar{y}(T))(V,\\
&X_{v,V}(T))+\nabla_2^2\phi_i(\beta(\tau;0),\beta(\tau;T))\Big(\frac{\partial}{\partial\tau}\beta(\tau;T),\frac{\partial}{\partial\tau}\beta(\tau;T)\Big)\\
&-\epsilon^2\nabla_2^2\phi_i(\bar{y}(0),\bar{y}(T))(X_{v,V}(T),X_{v,V}(T))\Big](1-\tau)d\tau+\nabla_2\phi_i(\bar{y}(0),\bar{y}(T))(\gamma_\epsilon(T)).
\end{aligned}$$

By (3.14) and (4.28), we have

$$\begin{aligned}L_i^\epsilon =&\epsilon^2\int_0^1\Big[\nabla_1^2\phi_i(\beta(\tau;0),\beta(\tau;T))(L_{\bar y(0)\beta(\tau;0)}V,L_{\bar y(0)\beta(\tau;0)}V)-\nabla_1^2\phi_i(\bar y(0),\bar y(T))(V,V)\\&+2\nabla_2\nabla_1\phi_i(\beta(\tau;0),\beta(\tau;T))(L_{\bar y(0)\beta(\tau;0)}V,L_{\bar y(T)\beta(\tau;T)}X_{v,V}(T))\\&-2\nabla_2\nabla_1\phi_i(\bar y(0),\bar y(T))(V,X_{v,V}(T))+\nabla_2^2\phi_i(\beta(\tau;0),\beta(\tau;T))(L_{\bar y(T)\beta(\tau;T)}X_{v,V}(T),\\&L_{\bar y(T)\beta(\tau;T)}X_{v,V}(T))-\nabla_2^2\phi_i(\bar y(0),\bar y(T))(X_{v,V}(T),X_{v,V}(T))\Big](1-\tau)d\tau+o(\epsilon^2)\\&+\nabla_2\phi_i(\bar y(0),\bar y(T))(\gamma_\epsilon(T)).\end{aligned}$$

Applying Lebesgue's dominated convergence theorem, $(C2)$, $(C3)$, (4.26) and (4.31) to the above identity, we obtain that, there exists $\tilde\epsilon_0\in(0,\epsilon_1]$ such that $|L_i^\epsilon|\le\alpha\epsilon^2$ for all $\epsilon\in[0,\tilde\epsilon_0]$. Similarly one can show that, there exists $\epsilon_0\in(0,\tilde\epsilon_0]$ such that

$$\begin{aligned}&\Big|\hat\psi(\epsilon V+\epsilon^2W,\bar u(\cdot)+\epsilon v(\cdot)+\epsilon^2\sigma(\cdot))-\hat\psi(0,\bar u(\cdot))-\epsilon D\hat\psi(0,\bar u(\cdot))(V,v(\cdot))\\&-\epsilon^2[D\hat\psi(0,\bar u(\cdot))(W,\sigma(\cdot))+\frac12D^2\hat\psi(0,\bar u(\cdot))(V,v(\cdot))\Big]\Big|\le\alpha\epsilon^2,\end{aligned}$$

for all $\epsilon\in[0,\epsilon_0]$, where

$$\begin{aligned}&D^2\hat\psi(0,\bar u(\cdot))(V,v(\cdot))\\=&\nabla_1^2\psi(\bar y(0),\bar y(T))(V,V)+2\nabla_2\nabla_1\psi(\bar y(0),\bar y(T))(V,X_{v,V}(T))\\&+\nabla_2^2\psi(\bar y(0),\bar y(T))(X_{v,V}(T),X_{v,V}(T))+2\nabla_2\psi(\bar y(0),\bar y(T))(Y_{00}^{X_{v,V}}(T)).\end{aligned}\tag{4.32}$$

Thus, for problem $(\widetilde{OCP})$, condition $(C5)$ holds.

Third, we shall use Theorem 4.1 to prove Theorem 2.1. Recalling (2.10), (2.11), (2.20) and (4.25), we have $I_A=\{i\in\{1,\cdots,j\}|\,\hat\phi_i(0,\bar u(\cdot))=0\}\cup\{0\}$ and $I_N=\{i\in\{1,\cdots,j\}|\,\hat\phi_i(0,\bar u(\cdot))<0\}$. Applying Theorem 4.1 to problem $(\widetilde{OCP})$, we can find $\ell=(\ell_{\phi_0},\cdots,\ell_{\phi_j},\ell_\psi)\in\mathbb{R}^{1+j+k}\setminus\{0\}$ satisfying (2.14) and

$$\begin{aligned}&\textstyle\sum_{i\in I_A}\ell_{\phi_i}\left(\nabla_1\phi_i(\bar y(0),\bar y(T))(Y)+\nabla_2\phi_i(\bar y(0),\bar y(T))(X_{w,Y}(T))\right)\\&+\ell_\psi^\top\left(\nabla_1\psi(\bar y(0),\bar y(T))(Y)+\nabla_2\psi(\bar y(0),\bar y(T))(X_{w,Y}(T))\right)\le0,\end{aligned}\tag{4.33}$$

for all $(Y,w(\cdot))\in T_{\bar y(0)}M\times T^\flat_{\mathcal{U}\cap L^2(0,T;\mathbb{R}^m)}(\bar u(\cdot))$, where $X_{w,Y}(\cdot)$ is the solution to (2.19) with $X_{w,Y}(0)=Y$, and we have used (4.26) and (4.27). Let $p^\ell(\cdot)$ be the solution to (2.12). Inserting (2.12) into (4.33) and integrating by parts, we can obtain from Lemma 4.3 that, $\int_0^T\nabla_uH[t,\ell](w(t))\le0$ holds for all $w(\cdot)\in L^2(0,T;\mathbb{R}^m)$ with $w(t)\in T_U^\flat(\bar u(t))$ a.e. $t\in[0,T]$, and that (2.16) stands. Applying needle variation to this inequality, we obtain (2.15). Thus, ℓ is a Lagrange multiplier in the sense of convex variation.

Finally, we shall employ Theorem 4.2 to prove Theorem 2.2. Assume that $v(\cdot)\in L^2(0,T;\mathbb{R}^m)$ is a singular direction in the sense of convex variation, with $X_v(\cdot)$ satisfying (2.18) and (2.19), and that (2.21) holds for some $\epsilon_0>0$ and $\ell(\cdot)\in L^2(0,T;\mathbb{R}^m)$. Recall

(4.30) and (4.32). It follows from Theorem 4.2 and Lemma 4.3 that, there exist a Lagrange mulitplier in the sense of convex variation $\ell = (\ell_{\phi_0}, \ell_{\phi_1}, \cdots, \ell_{\phi_j}, \ell_\psi^\top)^\top \in \mathbb{R}^{1+j+k} \setminus \{0\}$ satisfying (2.22) and

$$
\begin{aligned}
&\sum_{i\in I_0''} \ell_{\phi_i}\Big(\nabla_1\phi_i(\bar y(0),\bar y(T))(W) + \nabla_2\phi_i(\bar y(0),\bar y(T))(X_{\sigma,W}(T) + Y_{00}^{X_v}(T))\Big)\\
&+\ell_\psi^\top\Big(\nabla_1\psi(\bar y(0),\bar y(T))(W) + \nabla_2\psi(\bar y(0),\bar y(T))(X_{\sigma,W}(T) + Y_{00}^{X_v}(T))\Big)\\
&+\tfrac12\sum_{i\in I_0''} \ell_{\phi_i}\Big(\nabla_1^2\phi_i(\bar y(0),\bar y(T))(X_v(0),X_v(0)) + 2\nabla_2\nabla_1\phi_i(\bar y(0),\bar y(T))(X_v(0),\\
&X_v(T)) + \nabla_2^2\phi_i(\bar y(0),\bar y(T))(X_v(T),X_v(T))\Big) + \tfrac12\ell_\psi^\top\Big(\nabla_1^2\psi(\bar y(0),\bar y(T))(X_v(0),X_v(0))\\
&+2\nabla_2\nabla_1\psi(\bar y(0),\bar y(T))(X_v(0),X_v(T)) + \nabla_2^2\psi(\bar y(0),\bar y(T))(X_v(T),X_v(T)) \le 0,
\end{aligned}
\tag{4.34}
$$

for all $(W,\sigma) \in T_{\bar y(0)}M \times L^2(0,T;\mathbb{R}^m)$ with $\sigma(t) \in T_U^{\flat(2)}(\bar u(t), v(t))$ a.e. $t \in [0,T]$, where $Y_{00}^{X_v}(\cdot)$ is the solution to (3.1) with $\sigma(\cdot) = 0$ and $W = 0$.

Recall that $p^\ell(\cdot)$ solves (2.12) with initial data (2.16). We obtain from (2.12), (3.1), (4.31) and integration by parts over $[0,T]$ that

$$
\begin{aligned}
0 \ \ \ge\ \ & -p^\ell(0)(W) + p^\ell(T)(Y_{\sigma W}^{X_{v,V}}(T)) + \tfrac12\Big(\nabla_2^2\mathcal{L}(\bar y(0),\bar y(T),\hat\ell)(X_{v,V}(T),X_{v,V}(T))\\
& +2\nabla_1\nabla_2\mathcal{L}(\bar y(0),\bar y(T);\ell)(X_{v,V}(T),V) + \nabla_1^2\mathcal{L}(\bar y(0),\bar y(T);\ell)(V,V)\Big)\\
=\ \ & \int_0^T \Big(p^\ell(t)(Y_{\sigma W}^{X_{v,V}}(t))\Big)' dt + I\\
=\ \ & \int_0^T \nabla_u H[t,\ell](\sigma(t))dt + \tfrac12\int_0^T \Big(\nabla_x^2 H[t,\ell](X_{v,V}(t),X_{v,V}(t))\\
& +2\nabla_u\nabla_x H[t,\ell](X_{v,V}(t),v(t)) + \nabla_u^2 H[t,\ell](v(t),v(t))\\
& -R(\tilde p^\ell(t),X_{v,V}(t),f[t],X_{v,V}(t))\Big)dt + I,
\end{aligned}
$$

where

$$
\begin{aligned}
I = \ \ & \tfrac12\nabla_1^2\mathcal{L}(\bar y(0),\bar y(T);\ell)(V,V) + \nabla_1\nabla_2\mathcal{L}(\bar y(0),\bar y(T);\ell)(X_{v,V}(T),V)\\
& +\tfrac12\nabla_2^2\mathcal{L}(\bar y(0),\bar y(T);\ell)(X_{v,V}(T),X_{v,V}(T)),
\end{aligned}
$$

and thus (2.23) follows. □

References

[1] J. P. Aubin and H. Frankowska, *Set-valued analysis*, Systems & Control: Foundations & Applications, vol. 2, Birkhäuser Boston, Inc., Boston, MA, 1990. MR 1048347

[2] D. S. Bernstein, *A systematic approach to higher-order necessary conditions in optimization theory*, SIAM J. Control Optim. **22** (1984), no. 2, 211–238. MR 732425

[3] Q. Cui, L. Deng, and X. Zhang, *Second order optimality conditions for optimal control problems on Riemannian manifolds*, ESAIM Control Optim. Calc. Var. **25** (2019). MR 4003463

[4] L. Deng and X. Zhang, *A survey of optimal control problems with states constrained to riemannian manifolds*, In progress.

[5] ———, *Second order necessary conditions for endpoints-constrained optimal control problems on riemannian manifolds*, J. Differential Equations **272** (2021), 854–910.

[6] H. Frankowska, D. Hoehener, and D. Tonon, *A second-order maximum principle in optimal control under state constraints*, Serdica Math. J. **39** (2013), no. 3-4, 233–270. MR 3203187

[7] E. G. Gilbert and D. S. Bernstein, *Second-order necessary conditions in optimal control: accessory-problem results without normality conditions*, J. Optim. Theory Appl. **41** (1983), no. 1, 75–106. MR 718039

[8] P. Petersen, *Riemannian geometry*, second ed., Graduate Texts in Mathematics, vol. 171, Springer, New York, 2006. MR 2243772

[9] R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR 0274683

[10] J. Warga, *A second-order condition that strengthens Pontryagin's maximum principle*, J. Differential Equations **28** (1978), no. 2, 284–307. MR 0513535