Hi Urszula and Heinz!

Let me tell you how I see things now. I think I have made some progress, and we have material for a nice piece of joint work, maybe a major paper or series of papers. I still have a lot of stuff that needs to be checked, but what I have so far definitely looks promising, provided that it turns out to be right and that there isn't a mistake somewhere that invalidates the whole thing.

I include below my notes with what I have so far. In these comments, I will be referring always to the main theorem of AEMP (that is, $p = 2$ only; I do not want to get involved with the case of a $p > 2$ until I have thoroughly gone over the $p = 2$ case). And it's always for the fixed terminal state problem, although my guess is that what I say, if it is right, will work for more general terminal conditions.

It seems that, using the methods I know (Brouwer fixed point theorem stuff, no Lyusternik or Avakov or Dubovitskii-Milyutin) I can prove your theorem WITH $\rho = 0$ AND WITHOUT THIRD DERIVATIVES.

This is not what I had originally set out to prove, but it is what seems to come out naturally by applying the methods I know and like.

My original idea was that by doing things my way I would either

a. reach a point where I would be stuck and would not be able to go any further, and then I would have (reluctantly) to bring in the Avakov+Dubovitskii-Milyutin machinery in order to get unstuck,

or

b. end up proving your result with these other methods, in which case I would finally get to understand in my own terms why ρ is needed and why one needs those third derivatives which seem so out of place in a second-order result; in particular, at some point in my pursuit the new multiplier would pop out, and I would finally see why it's got to be there.

To my surprise, what I seem to have found is that the proof can indeed be pursued all the way to the end, and I can get your result by doing things my way, but ρ is still nowhere to be seen. If my theorem is true and my proof is right, then your theorem can be derived from it by just taking $\rho = 0$.

When I realized this, I was incredulous at first, because my theorem sort of contradicts yours. (Formally, there is no contradiction. Your theorem would follow from my theorem by just taking $\rho = 0$. So if my theorem is true then your theorem is also true, and everything is fine. Yet, it seems clear to me that one of the main points of your theorem is the new multiplier ρ, and that if ρ can always be taken to be zero then we are back to the old stuff of a single multiplier, and a lot of what makes the new result surprising is lost.) So I thought that there must be something wrong with my proof, and went back to it trying to find a mistake. I didn't find anything, so I decided to ask you to show me an example proving that ρ cannot always be taken to be zero. Such an example would have persuaded me that I was on the wrong track, and that my proof had to be wrong even if I could not find a mistake.

That's why I sent you the message with Question 2, thinking that you would be able to come up with an example. When you replied that you did *not* have an example, I regained confidence in my own proof, and now I am beginning to believe that it could be right, after all. I have written it down, trying to be very careful, and I have gone over it quite a few times. I think it is right, it feels right, I do not see how it could be wrong. But I have reached the point of diminishing returns by now. I think it would be useless for me to check again. Now it's your turn to do some checking.

So I am sending what I have to you, hoping that you will check it quite thouroughly. I am sure you will.

The theorem that you should compare with your result is Theorem 4.1. You may read it by thinking that $L = 1$, that is, that there is only one pair (h, v). (I did it with several such pairs because I wanted something that would also include the Goh and Legendre-Clebsch conditions. Theorem 4.1 does the job, but it would take some extra work to prove it. I do not want to do this extra work until I am sure that Theorem 4.1 is right. That's why I would ask you to try to go over my proof as fast as you can.) The "abstract" necessary condition for a minimum from which that theorem follows is Theorem 3.1. To make things clearer, I have also included (in Theorem 3.2) a restatement of Theorem 3.1 just for the case of a single $\hat{v}$. You may find it useful to look at Theorem 3.2 first. I am sure that you will recognize it as the abstract result that suffices to get your theorem with $\rho = 0$. In my way of doing things, a necessary condition such as the one of Theorem 3.1 is just a special case of a necessary condition for separation. So I am getting Theorem 3.1 as a corollary of the separation result, namely, Theorem 2.2 and its "multiplier" version, Corollary 2.3. The crux of the matter is the proof of Theorem 2.2. Technically, the main point of the proof is the use of the Brouwer fixed point theorem, which appears near the end of §2. This theorem plays in my framework the same role as the Lyusternik and Avakov theorems play in yours, namely, that of allowing us to solve *exactly* certain equations that we know how to solve approximately to first or second order.

You may also notice that in my approach one does not need to work with Banach spaces. Normed spaces are enough. (Of course, some completeness is needed somewhere, but in my way of doing things the only completeness one uses is that of *finite-dimensional spaces*. That completeness is undoubtedly an essential ingredient of the Brouwer fixed point theorem, but it's not needed otherwise.) I like to work with normed spaces that are not necessarily complete, because, for example, sometimes it may be technically convenient to use L^∞ controls but endow the space of controls with some much weaker norm (so that, for example, every control will be a limit of bang-bang controls).

Also, the convex set $\mathcal{U}$ over which we carry out the constrained minimization in the "abstract" setting is *not* assumed to have nonempty interior. Notice, in particular, that my set of "second order feasible directions" (called $var_{2,1}(u_*, v, \mathcal{U})$ in my notes) is more general than your set, and in particular can be nonempty even if the set $\mathcal{U}$ has empty interior.

If it turns out that I am right, then I suggest we write a joint paper with this stuff plus some more things that you may want to put in (e.g. $p > 2$, the "global" rather than "local" maximum principle, Goh and Legendre-Clebsch, whatever). It could be a very nice paper, showing how lots of things such as your extended maximum principle and many other results can be incorporated into one general unified framework. (It would be a *single multiplier* framework, naturally.) I am sure we can write it in a very positive, upbeat tone. The only non-upbeat thing that I would like to put in the paper, if you don't mind, would be a little remark to the effect that the Brouwer fixed point methods have proved themselves to be superior to the Dubovitskii-Milyutin approach, at least for this specific finite-dimensional type of problem.

I look forward to receiving your comments. Please do not hesitate to be critical, even harsh if necessary. As I said, I have checked and rechecked my proof many times, and I don't see how it could be wrong, but maybe you will see things that I do not see. And, if it turns out that I was wrong, only you and I will know that I made this mistake. And you are good friends, so I know that I have nothing to worry about.

Cordially,

Héctor

A NOTE ON SECOND-ORDER CONDITIONS

April 18, 2001

NOTE: this note is a very preliminary tentative draft. It's probably full of minor mistakes such as typos, and perhaps even major mistakes, such as, for example, the possibility that the whole thing is completely wrong. The author has checked the proofs to the best of his ability, and since he has worked so damn hard on this, he flatly refuses to accept any responsibility for any errors that might still be there. All such errors should be blamed on bad luck, the weather, the author's wife who repeatedly distracted him, or whatever, but the author is not responsible because he has done his best, and if he made any mistakes, he definitely didn't mean to, so these mistakes do not count (or, at least, should still entitle him to a lot of partial credit), as the author's undergraduate students have explained to him many times.

The only people in the whole universe who are authorized to read this note are Urszula Ledzewicz, Heinz Schättler, and the author (who reserves the right to re-read this any time he wants to). Anyone else reading this stuff will be indicted, prosecuted, convicted, drawn and quartered and sentenced to roast in Hell for all eternity.

Nobody, not even Heinz and Urszula, is allowed to quote from this text, except possibly in conversations between Heinz and Urszula and between one or both of them and the author.

Until the validity of the results and the proofs have been thouroughly checked, the author wishes to remain anonymous, which is why his name does not appear anywhere in the text.

1. Preliminaries

If A is a set and j is a positive integer, we use A^j to denote the Cartesian product of j copies of A. If a is any object and j is a positive integer, we use $a^{<j>}$ to denote the j-tuple $(a, a, \dots, a)$, consisting of j items all of them equal to a.

If $\mathcal{X}$, $\mathcal{Y}$ are normed spaces, and k is a nonnegative integer, then a *polynomial map of degree* $\le k$ *from* $\mathcal{X}$ *to* $\mathcal{Y}$ is a map

$$\mathcal{X} \ni x \mapsto \mathcal{P}(x) = P_0 + \sum_{j=1}^{k} \frac{1}{j!} P_j(x^{<j>}) \in \mathcal{Y} \tag{1.1}$$

such that

1. $P_0 \in \mathcal{Y}$,
2. for each $j \in \{1, \dots, k\}$,

$$\mathcal{X}^j \ni (x_1, \dots, x_j) \mapsto P_j(x_1, \dots, x_j) \in \mathcal{Y}$$

   is a symmetric multilinear map.

The maps P_j are the *Taylor coefficients* of $\mathcal{P}$. If $P_0 = 0$ then we say that $\mathcal{P}$ is a polynomial map *without constant term.* If $P_k \neq 0$, then we say that $\mathcal{P}$ is a polynomial map *of degree* k.

It is easy to see that

FACT 1.1. *The Taylor coefficients of a map* $\mathcal{P}$ *given by (1.1) are uniquely determined by the map* $\mathcal{P}$*,* $\diamondsuit$

and

FACT 1.2. *The map* $\mathcal{P}$ *given by (1.1) is continuous if and only if all the Taylor coefficients* P_j *are bounded, that is, if and only if there exists a constant* $C \in \mathbb{R}$ *such that*

$$\|P_j(x_1,\ldots,x_j)\| \leq C\prod_{i=1}^{j}\|x_i\| \text{ whenever } j \in \{1,\ldots,k\} \text{ and } (x_1,\ldots,x_j) \in \mathcal{X}^j\,. \quad \diamondsuit$$

DEFINITION 1.3. Assume that

1. $\mathcal{X}$, $\mathcal{Y}$ are normed spaces,
2. $\mathcal{S}$ is a subset of $\mathcal{X}$,
3. $x_* \in \mathcal{S}$,
4. $\mathcal{F}$ is a map from $\mathcal{S}$ to $\mathcal{Y}$,
5. k is a positive integer.

We say that $\mathcal{F}$ *is of class* D^k *at* x_* *in the direction of* $\mathcal{S}$ if

1. $\mathcal{F}$ is continuous on $N \cap \mathcal{S}$ for some neighborhood N of x_* in $\mathcal{X}$,
2. there exists a continuous polynomial map $\mathcal{P}$ of degree $\leq k$ without constant term such that
$$\lim_{x\to x_*,x\in\mathcal{S},x\neq x_*}\frac{\mathcal{F}(x)-\mathcal{F}(x_*)-\mathcal{P}(x-x_*)}{\|x-x_*\|^k} = 0\,. \quad \diamondsuit$$

DEFINITION 1.4. Let $\mathcal{X}$ be a normed space, let $\mathcal{S}$ be a subset of $\mathcal{X}$, and let $x_* \in \mathcal{S}$. A *variation of* x_* *in* $\mathcal{S}$ is a continuous curve

$$[0,\bar{\varepsilon}_\eta] \ni \varepsilon \mapsto \eta(t) \in \mathcal{S}\,, \tag{1.2}$$

defined on the interval $[0,\bar{\varepsilon}_\eta]$ for some positive real number $\bar{\varepsilon}_\eta$, having the property that $\eta(0) = x_*$. $\diamondsuit$

We use $Var(x_*,\mathcal{S})$ to indicate the set of all variations of x_* in $\mathcal{S}$.

DEFINITION 1.5. Let $\mathcal{X}$, $\mathcal{S}$, x_* be as in Definition 1.4. A *first-order variational direction of* $\mathcal{S}$ *at* x_* is a vector $v \in \mathcal{X}$ such that there exists a variation $\eta \in Var(x_*,\mathcal{S})$ with the property that

$$\eta(\varepsilon) = x_* + \varepsilon v + o(\varepsilon) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

A *limiting first-order variational direction of* $\mathcal{S}$ *at* x_* is a vector $v \in \mathcal{X}$ which is the limit of a sequence $\{v_j\}_{j\in\mathbb{N}}$ of first-order variational directions of $\mathcal{S}$ at x_*. $\diamondsuit$

We use $var_1(x_*,\mathcal{S})$, $\overline{var}_1(x_*,\mathcal{S})$, to denote, respectively, the set of all first-order variational directions of $\mathcal{S}$ at x_*. and the set of all limiting first-order variational directions of $\mathcal{S}$ at x_*. Clearly, $\overline{var}_1(x_*,\mathcal{S})$ is the closure of $var_1(x_*,\mathcal{S})$ in $\mathcal{X}$.

DEFINITION 1.6. Let $\mathcal{X}$ be a normed space. A *cone* in $\mathcal{X}$ is a nonempty subset C of $\mathcal{X}$ such that $r \cdot x \in C$ whenever $x \in C$, $r \in \mathbb{R}$, and $r \geq 0$. $\diamondsuit$

Definition 1.7. Let $\mathcal{X}$ be a normed space, let $\mathcal{S}$ be a subset of $\mathcal{X}$, and let $x_* \in \mathcal{S}$. The *Bouligand tangent cone* to $\mathcal{S}$ at x_* is the set of all vectors $v \in \mathcal{X}$ such that there exist sequences $\{v_j\}_{j\in\mathbb{N}}$, $\{\alpha_j\}_{j\in\mathbb{N}}$ for which

1. $v_j \in \mathcal{X}$ for every j,
2. $\alpha_j \in \mathbb{R}$ and $\alpha_j > 0$ for every j,
3. $x_* + \alpha_j v_j \in \mathcal{S}$ for every j,
4. $\lim_{j\to\infty} \alpha_j = 0$,
5. $\lim_{j\to\infty} v_j = v$. ♢

If $\mathcal{X}$, $\mathcal{S}$, x_* are as in Definition 1.7, we use $Boul(x_*, \mathcal{S})$ to denote the Bouligand tangent cone to $\mathcal{S}$ at x_*.

Fact 1.8. *Let $\mathcal{X}$ be a normed space, let $\mathcal{S}$ be a convex subset of $\mathcal{X}$, and let $x_* \in \mathcal{S}$. Then*

$$\overline{var}_1(x_*, \mathcal{S}) = Boul(x_*, \mathcal{S})\,.$$

Proof. Let $v \in \overline{var}_1(x_*, \mathcal{S})$, $j \in \mathbb{N}$. Then we can find $w_j \in var(x_*, \mathcal{S})$ such that $\|v - w_j\| \le 2^{-j}$. Since $w_j \in var(x_*, \mathcal{S})$, there exist a positive $\bar\varepsilon_j$ and a curve $\eta_j : [0, \bar\varepsilon_j] \mapsto \mathcal{S}$ such that $\eta_j(\varepsilon) = x_* + \varepsilon w_j + o(\varepsilon)$ as $\varepsilon \downarrow 0$. Choose α_j such that $0 < \alpha_j < \min(\bar\varepsilon_j, 2^{-j})$ and $\|\eta_j(\alpha_j) - x_* - \alpha_j w_j\| \le \alpha_j 2^{-j}$. Then $\lim_{j\to\infty} \alpha_j = 0$.

Let

$$v_j = \frac{\eta_j(\alpha_j) - x_*}{\alpha_j}\,.$$

Then $x_* + \alpha_j v_j = \eta_j(\alpha_j) \in \mathcal{S}$. In addition,

$$\|v_j - w_j\| = \left\|\frac{\eta_j(\alpha_j) - x_*}{\alpha_j} - w_j\right\| = \frac{1}{\alpha_j}\|\eta_j(\alpha_j) - x_* - \alpha_j w_j\| \le 2^- j\,.$$

Therefore $v_j - w_j \to 0$ as $j \to \infty$. So $\lim_{j\to\infty} v_j = v$. This proves that v belongs to $Boul(x_*, \mathcal{S})$.

Conversely, suppose that $v \in Boul(x_*, \mathcal{S})$. Let $\{v_j\}_{j\in\mathbb{N}}$, $\{\alpha_j\}_{j\in\mathbb{N}}$ be sequences having the five properties of Definition 1.7. Define a map $\eta_j : [0, \alpha_j] \mapsto \mathcal{X}$ by letting

$$\eta_j(\varepsilon) = x_* + \varepsilon v_j \tag{1.3}$$

for $\varepsilon \in [0, \alpha_j]$. Then $\eta_j(\varepsilon)$ is, for each ε, a convex combination of x_* and $x_* + \alpha_j v_j$. Since $x_* \in \mathcal{S}$, $x_* + \alpha_j v_j \in \mathcal{S}$, and $\mathcal{S}$ is convex, we can conclude that $\eta_j(\varepsilon) \in \mathcal{S}$, so η_j is a curve in $\mathcal{S}$. Clearly,

$$\eta_j(\varepsilon) = x_* + \varepsilon v_j + o(\varepsilon) \qquad \text{as} \qquad \varepsilon \downarrow 0\,,$$

because the much stronger condition (1.3) holds. So $v_j \in var_1(x_*, \mathcal{S})$ for every $j \in \mathbb{N}$, and then $v \in \overline{var}_1(x_*, \mathcal{S})$. ♢

Fact 1.9. *Let $\mathcal{X}$, $\mathcal{S}$, x_* be as in Definition 1.4. Then the sets $var_1(x_*, \mathcal{S})$ and $\overline{var}_1(x_*, \mathcal{S})$ are cones in $\mathcal{X}$, and $\overline{var}_1(x_*, \mathcal{S})$ is closed. If $\mathcal{S}$ is convex, then $var_1(x_*, \mathcal{S})$ and $\overline{var}_1(x_*, \mathcal{S})$ are convex cones.*

Proof. It is clear that $0 \in var(x_*, \mathcal{S})$, since $x_* \in \mathcal{S}$. Assume that v belongs to $var_1(x_*, \mathcal{S})$ and $\alpha > 0$. Let $\eta \in Var(x_*, \mathcal{S})$ be such that $\eta(\varepsilon) = x_* + \varepsilon v + o(\varepsilon)$ as $\varepsilon \downarrow 0$. Let $\tilde\eta$ be the reparametrized curve

$$\left[0, \frac{\bar\varepsilon_\eta}{\alpha}\right] \ni \varepsilon \mapsto \eta(\alpha\varepsilon) \stackrel{\text{def}}{=} \tilde\eta(\varepsilon) \in \mathcal{S}\,.$$

Then

$$\tilde{\eta}(\varepsilon) = x_* + \varepsilon \cdot (\alpha v) + o(\varepsilon) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

Therefore $\alpha v \in var_1(x_*, \mathcal{S})$. So $var_1(x_*, \mathcal{S})$ is a cone. The other statements are obvious. $\diamondsuit$

DEFINITION 1.10. Let $\mathcal{X}$, $\mathcal{S}$, x_* be as in Definition 1.4. A *second-order variational direction of $\mathcal{S}$ at x_** is an ordered pair (v, w) of vectors in $\mathcal{X}$ such that there exists a variation $\eta \in Var(x_*, \mathcal{S})$ with the property that

$$\eta(\varepsilon) = x_* + \varepsilon v + \varepsilon^2 w + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,. \qquad \diamondsuit$$

We use $var_2(x_*, \mathcal{S})$ to denote the set of all second-order variational directions of $\mathcal{S}$ at x_*. In addition, if $v_* \in \mathcal{X}$, we use $var_{2,1}(x_*, v_*, \mathcal{S})$ to denote the set of all $w \in \mathcal{X}$ such that $(v_*, w) \in var_2(x_*, \mathcal{S})$.

FACT 1.11. *Let $\mathcal{X}$, $\mathcal{S}$, x_* be as in Definition 1.4. Then*

1. $(0, 0) \in var_2(x_*, \mathcal{S})$,

and

2. $var_2(x_*, \mathcal{S})$ *has the* dilation invariance property*:*
   (DIP) $(\rho v, \rho^2 w) \in var_2(x_*, \mathcal{S})$ *whenever* $(v, w) \in var_2(x_*, \mathcal{S})$ *and* $\rho > 0$.

If, in addition, $\mathcal{S}$ is convex, then

3. $var_2(x_*, \mathcal{S})$ *is convex,*
4. $var_2(x_*, \mathcal{S})$ *satisfies:*
   (4.#) $(v, \rho w) \in var_2(x_*, \mathcal{S})$ *whenever* $(v, w) \in var_2(x_*, \mathcal{S})$ *and* $\rho \geq 1$,
5. *if* $v_* \in \mathcal{X}$, *then*
   (5.#) $var_{2,1}(x_*, v_*, \mathcal{S})$ *is convex,*
   (5.*) $\rho w \in var_{2,1}(x_*, v_*, \mathcal{S})$ *whenever* $w \in var_{2,1}(x_*, v_*, \mathcal{S})$ *and* $\rho \geq 1$,
   (5.&) *if* $0 \in var_{2,1}(x_*, v_*, \mathcal{S})$ *then* $var_{2,1}(x_*, v_*, \mathcal{S})$ *is a convex cone.*

PROOF. The fact that $(0, 0) \in var_2(x_*, \mathcal{X})$ is obvious.

Assume that $(v, w) \in var_2(x_*, \mathcal{S})$ and $\rho > 0$. Let $\eta \in Var(x_*, \mathcal{S})$ be such that $\eta(\varepsilon) = x_* + \varepsilon v + \varepsilon^2 w + o(\varepsilon^2)$ as $\varepsilon \downarrow 0$. Let $\tilde{\eta}$ be the reparametrized curve

$$\Big[0, \frac{\bar{\varepsilon}_\eta}{\rho}\Big] \ni \varepsilon \mapsto \eta(\rho\varepsilon) \stackrel{\text{def}}{=} \tilde{\eta}(\varepsilon) \in \mathcal{S}\,.$$

Then

$$\tilde{\eta}(\varepsilon) = x_* + \varepsilon \cdot (\rho v) + \varepsilon^2 \rho^2 w + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

Therefore $(\rho v, \rho^2 w) \in var_2(x_*, \mathcal{S})$. So (DIP) holds.

Now assume that the set $\mathcal{S}$ is convex. Pick pairs $(v_1, w_1) \in var_2(x_*, \mathcal{S})$, $(v_2, w_2) \in var_2(x_*, \mathcal{S})$, and α such that $0 \leq \alpha \leq 1$. Let $\eta_1 \in Var(x_*, \mathcal{S})$, $\eta_2 \in Var(x_*, \mathcal{S})$, be such that, for $i = 1, 2$,

$$\eta_i(\varepsilon) = x_* + \varepsilon v_i + \varepsilon^2 w_i + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

Let $\eta(\varepsilon) = \alpha\eta_1(\varepsilon) + (1 - \alpha)\eta_2(\varepsilon)$.. Then $\eta(\varepsilon) \in \mathcal{S}$ if ε is small enough, and

$$\eta(\varepsilon) = x_* + \varepsilon(\alpha v_1 + (1 - \alpha)v_2) + \varepsilon^2(\alpha w_1 + (1 - \alpha)w_2) + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

Therefore $\alpha(v_1, w_1) + (1 - \alpha)(v_2, w_2) \in var_2(x_*, \mathcal{S})$. This proves that $var_2(x_*, \mathcal{S})$ is convex.

If $(v,w) \in var_2(x_*,\mathcal{S})$, and $\rho \geq 1$, then $(\rho v,\rho^2 w) \in var_2(x_*,\mathcal{S})$ by dilation invariance. It follows that

$$(v,\rho w) = \frac{1}{\rho}(\rho v,\rho^2 w) = \Big(1-\frac{1}{\rho}\Big)(0,0) + \frac{1}{\rho}(\rho v,\rho^2 w) \in var_2(x_*,\mathcal{S})\,,$$

because $(0,0) \in var_2(x_*,\mathcal{S})$, $0 \leq \frac{1}{\rho} \leq 1$, and $var_2(x_*,\mathcal{S})$ is convex. So (4.#) holds.

Now pick $v_* \in \mathcal{X}$, and let $W = var_{2,1}(x_*,v_*,\mathcal{S})$. Let $w_1 \in W$, $w_2 \in W$. Let α be such that $0 \leq \alpha \leq 1$. Then

$$(v_*, \alpha w_1 + (1-\alpha)w_2) = \alpha(v_*,w_1) + (1-\alpha)(v_*,w_2) \in var_2(x_*,v_*,\mathcal{S})\,,$$

because $var_2(x_*,v_*,\mathcal{S})$ is convex. So $\alpha w_1 + (1-\alpha)w_2 \in W$. This proves that W is convex.

Property (5.*) follows easily from (4.#). Finally, we assume that $0 \in W$ and show that W is a convex cone. We already know that W is convex and is closed under multiplication by real numbers ρ such that $\rho \geq 1$. So it suffices to show that $\rho w \in W$ whenever $w \in W$ and $0 \leq \rho \leq 1$. But, if $w \in W$ and $0 \leq \rho \leq 1$, then $\rho w = (1-\rho)\cdot 0 + \rho w$, which belongs to W because $0 \in W$ and W is convex. So W is indeed a convex cone. $\diamondsuit$

We let $\pi^{\mathcal{X}}_{2,1}$ be the projection map

$$\mathcal{X} \times \mathcal{X} \ni (v,w) \mapsto \pi^{\mathcal{X}}_{2,1}(v,w) \stackrel{\text{def}}{=} v \in \mathcal{X}\,.$$

Then the following observation is trivial:

FACT 1.12. *If $\mathcal{X}$, $\mathcal{S}$, x_* are as in Definition 1.4, then*

$$\pi^{\mathcal{X}}_{2,1}\Big(var_2(x_*,\mathcal{S})\Big) \subseteq var_1(x_*,\mathcal{S})\,. \qquad \diamondsuit$$

2. The separation theorem

We assume that

(A1) *$\mathcal{U}^a$ (the "ambient source space") and Y (the "target space") are normed spaces, Y is finite-dimensional, $\mathcal{U}$ is a convex subset of $\mathcal{U}^a$, u_* is a point of $\mathcal{U}$, $\mathcal{F} : \mathcal{U} \mapsto Y$ is a map, $\omega \in Y$ and $\omega \neq 0$.*

DEFINITION 2.1. Let $\mathcal{X}$ be a normed space, let $\mathcal{S}$ be a subset of $\mathcal{X}$, and let $\mathcal{F}$ be a map from $\mathcal{S}$ to another normed space $\mathcal{Y}$. We say that a subset $\mathcal{S}_0$ of $\mathcal{S}$ is $(\mathcal{S},\mathcal{F})$-*abundant* if there exists a sequence $\{\sigma_j\}_{j\in\mathbb{N}}$ of maps $\sigma_j : \mathcal{S} \mapsto \mathcal{S}_0$ such that

1. the map $\mathcal{F} \circ \sigma_j$ is continuous for each j,
2. $\mathcal{F} \circ \sigma_j$ converges to $\mathcal{F}$ uniformly on compact subsets of $\mathcal{S}$ as $j \to \infty$. $\diamondsuit$

We assume that

(A2) *$\mathcal{A}$ is an $(\mathcal{U},\mathcal{F})$-abundant subset of $\mathcal{U}$ such that $u_* \in \mathcal{A}$.*

We write

$$y_* = \mathcal{F}(u_*)\,, \tag{2.1}$$

and let Λ be the closed half-line

$$\Lambda \stackrel{\text{def}}{=} \{y_* + t\omega : t \in \mathbb{R}, t \geq 0\}\,. \tag{2.2}$$

Clearly, $y_* \in \mathcal{F}(\mathcal{A}) \cap \Lambda$. We will look for second-order necessary conditions for the following *separation property*:

(SEP) $$\mathcal{F}(\mathcal{A}) \cap \Lambda = \{y_*\}\,.$$

We assume that

(A3) *$\mathcal{F}$ is of class D^2 at u_*.*

We let $\mathcal{V}$ be the set of all "first-order variational directions." That is, we let

$$\mathcal{V} = var_1(u_*, \mathcal{U})\,. \tag{2.3}$$

We consider, in addition, a finite set $\hat{\mathcal{V}}$ of "independent first-order variational directions with second-order effects." This means, to begin with, that

(A4) $\hat{\mathcal{V}}$ is a finite subset of $\mathcal{V}$.

We will need additional assumptions on $\hat{\mathcal{V}}$. In order to state these assumptions, we use (A3) and choose P_1, P_2 such that

(A5) *P_1 is a continuous linear map from $\mathcal{U}^a$ to Y, P_2 is a continuous bilinear symmetric map from $\mathcal{U}^a \times \mathcal{U}^a$ to Y, and*

$$\mathcal{F}(u_* + v) = \mathcal{F}(u_*) + P_1 \cdot v + \frac{1}{2}P_2(v,v) + o(\|v\|^2) \text{ as } v \to 0\,,\ u_* + v \in \mathcal{U}\,.$$

We define

$$\Lambda_0 = \{t\omega : t \geq 0\}\,, \tag{2.4}$$

so that

$$\Lambda = y_* + \Lambda_0 = \{y_* + z : z \in \Lambda_0\}\,.$$

We then assume that

(A6) *the vector $P_1 \cdot v$ belongs to Λ_0 whenever $v \in \hat{\mathcal{V}}$,*

as well as the *cross term condition*:

(A7) *whenever $v_1, v_2 \in \hat{\mathcal{V}}$ are such that $v_1 \neq v_2$, the vector $P_2(v_1, v_2)$ belongs to Λ_0.*

Finally, we need the "independence condition." The purpose of this condition is to be able to add variations

$$\varepsilon \mapsto \eta_v(\varepsilon) = u_* + \tilde{\eta}_v(\varepsilon) = u_* + \varepsilon v + o(\varepsilon)$$

of u_* in $\mathcal{U}$ corresponding to different members v of $\hat{\mathcal{V}}$, and still obtain a variation of u_* in $\mathcal{U}$. We will guarantee this by requiring that the values $\tilde{\eta}_v(\varepsilon)$ belong to sets $\mathcal{U}_v^0$ such that $u_* + \sum_\ell \mathcal{U}_v^0 \subseteq \mathcal{U}$, from which it will follow that $u_* + \sum_v \tilde{\eta}_v(\varepsilon) \in \mathcal{U}$.

The sets $\mathcal{U}_v^0$ will actually play a role in the statement of Theorem 2.2. Therefore, rather than merely assume that they exist, we will actually regard them as additional data for our separation problem. If we write $\mathcal{U}_v = \mathcal{U}_v^0 + u_*$, the "independence condition" becomes:

(A8) *$\{\mathcal{U}_v\}_{v \in \hat{\mathcal{V}}}$ is a family of convex subsets of $\mathcal{U}^a$ such that*

1. *$u_* \in \mathcal{U}_v$ for every $v \in \hat{\mathcal{V}}$,*
2. *$u_* + \sum_{v \in \hat{\mathcal{V}}} \Big(\mathcal{U}_v - u_*\Big) \subseteq \mathcal{U}$,*
3. *$v \in var_1(u_*, \mathcal{U}_v)$ for each $v \in \hat{\mathcal{V}}$.*

We will prove that if a *second-order surjectivity condition*—called (SUR-2)—holds, then (SEP) fails. So, for (SEP) to hold, it will be a necessary condition that (SUR-2) not be true. Rewriting this in terms of multipliers will give us a second-order Maximum Principle.

We define the cone $C_1(\mathcal{V})$ to be the smallest closed convex cone in Y that contains all the vectors $P_1 \cdot v$, for $v \in \mathcal{V}$.

If $v \in \hat{\mathcal{V}}$, we define

$$\mathcal{W}(v) \stackrel{\text{def}}{=} \{w \in \mathcal{U}^a : (v, w) \in var_2(u_*, \mathcal{U}_v)\} ,$$

that is,

$$\mathcal{W}(v) \stackrel{\text{def}}{=} var_{2,1}(u_*, v, \mathcal{U}_v) . \tag{2.5}$$

We then let $C_2(\mathcal{V}, \hat{\mathcal{V}})$ be the smallest closed convex cone in Y that contains $C_1(\mathcal{V})$ as well as all the vectors $\frac{1}{2}P_2(v, v) + P_1 \cdot w$, for all (v, w) such that $v \in \hat{\mathcal{V}}$ and $w \in \mathcal{W}(v)$. The *second-order surjectivity condition* then says that

(SUR-2) *ω is an interior point of the cone $C_2(\mathcal{V}, \hat{\mathcal{V}})$.*

We are going to prove the following result.

THEOREM 2.2. *Assume that the data $\mathcal{U}^a$, Y, $\mathcal{U}$, u_*, $\mathcal{F}$, ω, $\mathcal{A}$, $\hat{\mathcal{V}}$, P_1, P_2, $\{\mathcal{U}_v\}_{v\in\hat{\mathcal{V}}}$ satisfy Assumptions (A1)-(A8), where y_*, Λ, $\mathcal{V}$, Λ_0 are defined by (2.1), (2.2), (2.3), (2.4). Then the second-order surjectivity condition (SUR-2) implies that the separation property (SEP) does not hold.*

An obvious consequence of Theorem 2.2 is the following "multiplier" result:

COROLLARY 2.3. *Assume that the hypotheses of Theorem 2.2 hold. Then a necessary condition for the separation property (SEP) is that there exist a nonzero covector $\bar{\psi} \in Y^*$ such that*

$$\langle \bar{\psi}, P_1 \cdot v \rangle \;\; \geq \;\; 0 \text{ for all } v \in \overline{var}_1(u_*, \mathcal{U}) , \tag{2.6}$$

$$\langle \bar{\psi}, \frac{1}{2}P_2(v, v) + P_1 \cdot w \rangle \;\; \geq \;\; 0 \text{ for all } v \in \hat{\mathcal{V}},\, w \in var_{2,1}(x_*, v, \mathcal{U}_v) , \tag{2.7}$$

$$\langle \bar{\psi}, \omega \rangle \;\; \leq \;\; 0 . \tag{2.8}$$

Condition (2.7) implies, in particular, that

$$\langle \bar{\psi}, P_1 \cdot w \rangle \;\; \geq \;\; 0 \qquad \text{whenever} \qquad v \in \hat{\mathcal{V}}, w \in var_{2,1}(x_*, v, \mathcal{U}_v) . \tag{2.9}$$

PROOF THAT THEOREM 2.2 IMPLIES COROLLARY 2.3. First, we observe that (2.9) is a consequence of (2.7) because, if $v \in \hat{\mathcal{V}}$, $w \in \mathcal{W}(v)$, then $\rho \cdot w \in \mathcal{W}(v)$ whenever $\rho \geq 1$, in view of condition (5.*) of Fact 1.11, so that

$$\langle \bar{\psi}, \tfrac{1}{2}P_2(v, v) + \rho.P_1 \cdot w \rangle \;\; \geq \;\; 0 \qquad \text{whenever} \qquad \rho \geq 1 ,$$

and then, dividing by ρ and letting $\rho \uparrow +\infty$, (2.9) follows.

Write

$$C = C_2(\mathcal{V}, \hat{\mathcal{V}}) , \qquad \overset{\circ}{C} \stackrel{\text{def}}{=} \text{Interior}(C) . \tag{2.10}$$

If $\overset{\circ}{C} = \emptyset$ then the linear subspace of Y spanned by C is of positive codimension in Y, so there exists a nonzero covector $\bar{\psi} \in Y^*$ such that $\langle \bar{\psi}, v \rangle = 0$ for all $v \in C$.

After replacing $\bar{\psi}$ by $-\bar{\psi}$, if necessary, we may assume that $\langle\bar{\psi},\omega\rangle \le 0$, and then $\bar{\psi}$ satisfies all our conditions.

If $\overset{o}{C} \neq \emptyset$, then $\overset{o}{C}$ and Λ_0 are disjoint convex subsets of Y, $\overset{o}{C}$ is open, and Λ_0 is closed. The Hahn-Banach theorem then implies that there exist $\bar{\psi} \in Y^*\backslash\{0\}$ and $\alpha \in \mathbb{R}$ such that $\langle\bar{\psi}, t\omega\rangle \le \alpha$ for all $t \in [0,+\infty[$, and $\langle\bar{\psi}, z\rangle \ge \alpha$ for all $z \in \overset{o}{C}$. The first condition tells us in particular that $\langle\bar{\psi},\omega\rangle \le \alpha$ and $0 \le \alpha$. The second condition implies that $\langle\bar{\psi}, z\rangle \ge \alpha$ for all $z \in C$, since C is the closure of $\overset{o}{C}$, thanks to the fact that $\overset{o}{C} \neq \emptyset$. Since $0 \in C$, we have $\alpha \le 0$. Then $\alpha = 0$, and our conclusion follows. (The fact that the inequality of (2.6) holds for all $v \in \overline{var}_1(u_*,\mathcal{U})$ follows because any such v is the limit of a sequence $\{v_j\}_{j\in\mathbb{N}}$ of members of $\mathcal{V}$, so that $P_1 \cdot v_j \in C$ for every j, and then $\langle\bar{\psi}, P_1\cdot v_j\rangle \ge 0$, so $\langle\bar{\psi}, P_1 \cdot v\rangle \ge 0$.) ♢

Proof of Theorem 2.2. Use C, $\overset{o}{C}$ as in (2.10). Assume that (SUR-2) holds, so that $\omega \in \overset{o}{C}$. Let $\mathcal{C}$ be the set consisting of all vectors $P_1 \cdot v$, $v \in \mathcal{V}$, together with all the vectors $\frac{1}{2}P_2(v,v) + P_1 \cdot w$, for (v,w) such that $v \in \hat{\mathcal{V}}$ and $w \in \mathcal{W}(v)$. Then every member of C is a limit of finite linear combinations of members of $\mathcal{C}$ with nonnegative coefficients.

Enumerate the members of $\hat{\mathcal{V}}$, by writing

$$\hat{\mathcal{V}} = \{\hat{v}_1,\ldots,\hat{v}_L\}\,, \qquad \text{with} \qquad v_\ell \neq v_{\ell'} \qquad \text{whenever} \qquad \ell \neq \ell'\,.$$

Also, write

$$\mathcal{U}_\ell \stackrel{\text{def}}{=} \mathcal{U}_{\hat{v}_\ell}\,, \qquad \mathcal{U}_\ell^0 \stackrel{\text{def}}{=} \mathcal{U}_\ell - u_*\,.$$

Let $n = \dim Y$. Let $\tilde{e}_1,\ldots,\tilde{e}_n$ be a basis of Y, and define $\tilde{e}_0 = -(\tilde{e}_1+\ldots+\tilde{e}_n)$. Since $\omega \in \overset{o}{C}$, we can choose $\bar{s}$ such that $\bar{s} > 0$ and the vectors $\omega + \bar{s}\tilde{e}_i$ belong to $\overset{o}{C}$ for $i = 0,\ldots,n$. Let $\hat{e}_i = \bar{s}\tilde{e}_i$ for $i = 0,\ldots,n$. Then the vectors $\hat{e}_1,\ldots,\hat{e}_n$ constitute a basis of Y, $\hat{e}_0 = -(\hat{e}_1+\ldots+\hat{e}_n)$, and

$$\omega + \hat{e}_i \in \overset{o}{C} \ \text{ for } \ i = 0,\ldots,n\,.$$

Let $\hat{\omega}_i = \omega + \hat{e}_i$, for $i = 0,\ldots,n$. Then each $\hat{\omega}_i$ is the limit of a sequence $\{\hat{\omega}_i^j\}_{j\in\mathbb{N}}$ of finite linear combinations of members of $\mathcal{C}$ with nonnegative coefficients. Write

$$\Omega = (\omega,1)\,, \quad \hat{\omega}_i^\infty = \hat{\omega}_i\,, \quad \Omega_i = \Omega_i^\infty = (\hat{\omega}_i,1)\,, \quad \Omega_i^j = (\hat{\omega}_i^j,1)\,.$$

Then Ω and the Ω_i^j—for $i = 0,\ldots,n$, $j \in \bar{\mathbb{N}} \stackrel{\text{def}}{=} \mathbb{N}\cup\{+\infty\}$—are members of $Y\times\mathbb{R}$. Moreover, $\Omega_0,\Omega_1,\ldots,\Omega_n$ are linearly independent. (Reason: if $\sum_{i=0}^n r_i\Omega_i = 0$ then $\sum_{i=0}^n r_i\hat{\omega}_i = 0$ and $\sum_{i=0}^n r_i = 0$, so

$$\sum_{i=0}^n r_i\hat{e}_i = \sum_{i=0}^n r_i(\omega+\hat{e}_i) = \sum_{i=0}^n r_i\hat{\omega}_i = 0\,,$$

and then $\sum_{i=1}^n (r_i - r_0)\hat{e}_i = 0$, since $\hat{e}_0 = -(\hat{e}_1+\ldots+\hat{e}_n)$. But then the linear independence of the vectors $\hat{e}_1,\ldots,\hat{e}_n$ implies that

$$r_1 - r_0 = r_2 - r_0 = \ldots = r_n - r_0 = 0\,.$$

Therefore $0 = \sum_{i=0}^n r_i = (n+1)r_0$, so $r_0 = 0$, and then $r_1 = r_2 = \ldots = r_n = 0$.)

It follows that the $n+1$-tuple $(\Omega_0^j,\Omega_1^j,\ldots,\Omega_n^j)$ is linearly independent for large j, so that we may assume, after passing to a subsequence if necessary, that $(\Omega_0^j,\Omega_1^j,\ldots,\Omega_n^j)$ is a linearly independent $n+1$-tuple for all $j \in \bar{\mathbb{N}}$. Therefore, if

$j \in \bar{\mathbb{N}}$ then every vector $\xi \in Y \times \mathbb{R}$ can be expressed in a unique way as a linear combination $\sum_{i=0}^{n} \tilde{s}_i^j(\xi)\Omega_i^j$ of $\Omega_0^j, \Omega_1^j, \dots, \Omega_n^j$. Clearly, the coefficients $\tilde{s}_i^j(\xi)$ depend continuously on ξ for each i, j and, moreover, $\lim_{j\to\infty} \tilde{s}_i^j(\xi) = \tilde{s}_i^\infty(\xi)$ for each i, ξ, uniformly with respect to ξ for ξ in a compact subset of $Y \times \mathbb{R}$. If $y \in Y$, define $s_i^j(y) = \tilde{s}_i^j(y, 1)$. It follows that

$$\sum_{i=0}^{n} s_i^j(y)\hat{\omega}_i^j = y \qquad \text{and} \qquad \sum_{i=0}^{n} s_i^j(y) = 1 \qquad \text{whenever} \quad y \in Y\,.$$

Moreover,

$$\omega = \frac{1}{n+1}(\hat{\omega}_0 + \hat{\omega}_1 + \ldots + \hat{\omega}_n)\,,$$

so that

$$s_0^\infty(\omega) = s_1^\infty(\omega) = \ldots + s_n^\infty(\omega) = \frac{1}{n+1}\,.$$

It follows that

$$s_0^j(\omega) > 0\,,\; s_1^j(\omega) > 0\,,\; \ldots\; s_n^j(\omega) > 0\,, \tag{2.11}$$

if j is large enough.

Now fix a $j \in \mathbb{N}$ such that (2.11) holds. Write

$$\hat{\omega}_i = \hat{\omega}_i^j \qquad \text{for} \qquad i = 0, \dots, n\,.$$

Then every $y \in Y$ has a unique expression as a linear combination of $\hat{\omega}_0, \dots, \hat{\omega}_n$ with coefficients adding up to 1, and the coefficients of this combination are the numbers $\hat{s}_i(y) \stackrel{\text{def}}{=} s_i^j(y)$. These coefficients depend continuously on y, and $\hat{s}_i(\omega) > 0$ for $i = 0, \dots, n$. Moreover, each $\hat{\omega}_i$ is a finite linear combination with nonnegative coefficients of members of $\mathcal{C}$. It follows (using the fact that the sets $\mathcal{W}(v)$ are convex for each $v \in \hat{\mathcal{V}}$) that there exist

1. a finite subset $\{v_1, \dots, v_J\}$ of $\mathcal{V}$,
2. members $w_{i\ell}$ of $\mathcal{W}(\hat{v}_\ell)$ for $i \in \{0, \dots, n\}$, $\ell \in \{1, \dots, L\}$,
3. nonnegative real numbers α_{ij}, for $i = 0, \dots, n$, $j = 1, \dots, J$,
4. nonnegative real numbers $\hat{\beta}_{i\ell}$, for $i = 0, \dots, n$, $\ell = 1, \dots, L$,

such that

$$\hat{\omega}_i = \sum_{j=1}^{J} \alpha_{ij} P_1 \cdot v_j + \sum_{\ell=1}^{L} \hat{\beta}_{i\ell}\Big(P_1 \cdot w_{i\ell} + \frac{1}{2} P_2(\hat{v}_\ell, \hat{v}_\ell)\Big) \text{ for } i = 0, \dots, n\,.$$

Now, if $\delta > 0$, define

$$\omega_i^\delta = \sum_{j=1}^{J} \alpha_{ij} P_1 \cdot v_j + \sum_{\ell=1}^{L} (\hat{\beta}_{i\ell} + \delta)\Big((1-\delta)P_1 \cdot w_{i\ell} + \frac{(1-\delta)^2}{2} P_2(\hat{v}_\ell, \hat{v}_\ell)\Big)\,,$$

for $i = 0, \dots, n$. Then $\omega_i^\delta \to \hat{\omega}_i$ as $\delta \downarrow 0$. It follows that, if δ is sufficiently small, then

(&) every $y \in Y$ has a unique expression as a linear combination of $\omega_0^\delta, \dots, \omega_n^\delta$ with coefficients adding up to 1, and the coefficients $s_{i,\delta}(y)$ of this combination depend continuously on y, and are such that $s_{i,\delta}(\omega) > 0$ for $i = 0, \dots, n$.

Fix a δ for which (&) happens, such that $0 < \delta < 1$. Call it $\bar{\delta}$, and write

$$\omega_i = \omega_i^{\bar{\delta}}\,, \qquad s_i(y) = s_{i,\bar{\delta}}(y)\,, \qquad \beta_{i\ell} = \hat{\beta}_{i\ell} + \bar{\delta}\,.$$

Then

$$\omega_i = \sum_{j=1}^{J} \alpha_{ij} P_1 \cdot v_j + \sum_{\ell=1}^{L} \beta_{i\ell}\Big((1-\bar{\delta})P_1 \cdot w_{i\ell} + \frac{(1-\bar{\delta})^2}{2} P_2(\hat{v}_\ell, \hat{v}_\ell)\Big)\,, \tag{2.12}$$

for $i = 0, \ldots, n$, and there exist continuous functions $Y \ni y \mapsto s_i(y) \in \mathbb{R}$, for $i = 0, \ldots, n$, such that

$$y = \sum_{i=0}^{n} s_i(y)\omega_i \ \text{ and } \ \sum_{i=0}^{n} s_i(y) = 1 \ \text{ for all } \ y \in Y\,. \tag{2.13}$$

Moreover, the numbers α_{ij} are nonnegative, the $\beta_{i\ell}$ are strictly positive, and

$$s_0(\omega) > 0\,,\ s_1(\omega) > 0\,,\ \ldots\ s_n(\omega) > 0\,. \tag{2.14}$$

Let Σ be the convex hull of the vectors ω_i, $i = 0, \ldots, n$, so Σ is the simplex

$$\Sigma = \Big\{ y \in Y : (\exists s_0, \ldots, s_n \in \mathbb{R})\Big(y = \sum_{i=0}^{n} s_i\omega_i\,,\ s_i \geq 0 \text{ for } i = 0, \ldots, n\,,\ \sum_{i=0}^{n} s_i = 1\Big)\Big\},$$

that is,

$$\Sigma = \Big\{\, y \in Y : s_i(y) \geq 0 \text{ for } i = 0, \ldots, n \Big\}\,.$$

Then the following are true:

1. ω is an interior point of Σ,
2. every $y \in \Sigma$ has an expression

$$y = \sum_{i=0}^{n} s_i(y)\omega_i\,,$$

where the functions $\Sigma \ni y \mapsto s_i(y) \in \mathbb{R}$ are continuous, nonnegative, and such that $\sum_{i=0}^{n} s_i(y) = 1$ for every $y \in \Sigma$.

We now fix a ρ such that

$$\rho > 0 \qquad \text{and} \qquad \mathcal{B} \subseteq \Sigma\,, \qquad \text{where} \qquad \mathcal{B} = \{y \in Y : \|y - \omega\| \leq \rho\}\,.$$

We then use (A6) and (A7) to pick nonnegative numbers τ_ℓ, $\theta_{\ell,\ell'}$ such that

$$P_1 \cdot \hat{v}_\ell = \tau_\ell\omega \ \text{ for } \ell = 1, \ldots, L\,, \tag{2.15}$$

$$P_2(\hat{v}_\ell, \hat{v}_{\ell'}) = \theta_{\ell,\ell'}\omega \ \text{ for } \ell, \ell' = 1, \ldots, L\,,\ \ell \neq \ell'\,. \tag{2.16}$$

Lemma 2.4. *Let δ be such that $0 < \delta < 1$. Then there exist an $\bar{\varepsilon}$ such that $\bar{\varepsilon} > 0$, and continuous curves*

$$\eta^y : [0, \bar{\varepsilon}] \mapsto \mathcal{U}\,,$$

such that

1. *the map*

$$[0, \bar{\varepsilon}] \times \Sigma \ni (\varepsilon, y) \mapsto \eta^y(\varepsilon) \in \mathcal{U}$$

is continuous,

2. *as* $\varepsilon \downarrow 0$,

$$\begin{aligned}\eta^y(\varepsilon) &= u_* + \varepsilon\left((1-\delta)\sum_{\ell=1}^{L}\left(\sqrt{\sum_{i=0}^{n} s_i(y)\beta_{i\ell}}\right)\hat{v}_\ell\right) + \varepsilon^2\left(\sum_{j=1}^{J}\sum_{i=0}^{n} s_i(y)\alpha_{ij}v_j\right)\\ &\quad +\varepsilon^2(1-\delta)\left(\sum_{\ell=1}^{L}\sum_{i=0}^{n} s_i(y)\beta_{i\ell}w_{i\ell}\right) + o(\varepsilon^2)\,,\end{aligned} \tag{2.17}$$

uniformly with respect to $y \in \Sigma$.

Proof of Lemma 2.4. The fact that the v_j belong to $\mathcal{V}$ implies that there exist continuous curves $\tilde{\zeta}_j : [0, \tilde{\varepsilon}_j] \mapsto \mathcal{U}$ such that $\tilde{\zeta}_j(\varepsilon) = u_* + \varepsilon v_j + o(\varepsilon)$. Let

$$\bar{\alpha} = \max\{\alpha_{ij} : i = 0, \ldots, n\}\,.$$

Define

$$\zeta_j^y(\varepsilon) = \tilde{\zeta}_j\Big(\frac{J\varepsilon^2}{\delta}\sum_{i=0}^{n} s_i(y)\alpha_{ij}\Big)\,.$$

Then the curves ζ_j^y are certainly defined for $\varepsilon \in [0, \bar{\varepsilon}_1]$, where

$$\bar{\varepsilon}_1 = \sqrt{\frac{\delta\min\{\tilde{\varepsilon}_j : j = 1, \ldots, J\}}{1 + J\bar{\alpha}}}\,.$$

Moreover, the ζ_j^y are $\mathcal{U}$-valued, and satisfy

$$\zeta_j^y(\varepsilon) = u_* + \frac{J\varepsilon^2}{\delta}\Big(\sum_{i=0}^{n} s_i(y)\alpha_{ij}\Big)v_j + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,, \tag{2.18}$$

the various $o(\varepsilon^2)$s being uniform with respect to y.

Then define

$$\zeta^y(\varepsilon) = \frac{1}{J}\Big(\sum_{j=1}^{J}\zeta_j^y(\varepsilon)\Big)\,.$$

The map $[0, \bar{\varepsilon}_1] \times \Sigma \ni (\varepsilon, y) \mapsto \zeta^y(\varepsilon)$ is clearly continuous and $\mathcal{U}$-valued. Moreover,

$$\zeta^y(\varepsilon) = u_* + \frac{\varepsilon^2}{\delta}\Big(\sum_{i=0}^{n}\sum_{j=1}^{J} s_i(y)\alpha_{ij}v_j\Big) + o(\varepsilon^2)\,.$$

Next we use the fact that the pairs $(\hat{v}_\ell, w_{i\ell})$ belong to $var_2(u_*, \mathcal{U}_\ell)$ to conclude that there exist continuous curves $\hat{\theta}_{i\ell} : [0, \hat{\varepsilon}_{i\ell}] \mapsto \mathcal{U}_\ell$, defined on intervals $[0, \hat{\varepsilon}_{i\ell}]$, with $\hat{\varepsilon}_{i\ell} > 0$, such that

$$\hat{\theta}_{i\ell}(\varepsilon) = u_* + \varepsilon\hat{v}_\ell + \varepsilon^2 w_{i\ell} + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

Define

$$\tilde{\theta}_\ell^y(\varepsilon) = \sum_{i=0}^{n}\frac{s_i(y)\beta_{i\ell}}{\sum_{i'=0}^{n} s_{i'}(y)\beta_{i'\ell}}\hat{\theta}_{i\ell}^y(\varepsilon)\,. \tag{2.19}$$

Then

$$\tilde{\theta}_\ell^y(\varepsilon) = u_* + \varepsilon\hat{v}_\ell + \varepsilon^2\sum_{i=0}^{n}\frac{s_i(y)\beta_{i\ell}w_{i\ell}}{\sum_{i'=0}^{n} s_{i'}(y)\beta_{i'\ell}} + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0\,.$$

Define

$$\theta_\ell^y(\varepsilon) = \tilde{\theta}_\ell^y\left(\varepsilon\sqrt{\sum_{i=0}^n s_i(y)\beta_{i\ell}}\right).$$

Then all the curves θ_ℓ^y are defined on the interval $[0, \bar{\varepsilon}_2]$, where

$$\bar{\varepsilon}_2 = \frac{\min\{\hat{\varepsilon}_{i\ell} : i = 0, \dots, n, \ell = 1 \dots, L\}}{\sqrt{\max\{\beta_{i\ell} : i = 0, \dots, n, \ell = 1 \dots, L\}}}.$$

Moreover,

$$\theta_\ell^y(\varepsilon) = u_* + \left(\varepsilon\sqrt{\sum_{i=0}^n s_i(y)\beta_{i\ell}}\right)v_\ell + \varepsilon^2\Big(\sum_{i=0}^n s_i(y)\beta_{i\ell}w_{i\ell}\Big) + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0.$$

Now, the sets $\mathcal{U}_\ell$ are convex, the curves $\hat{\theta}_{i\ell}$ are $\mathcal{U}_\ell$-valued, the vectors $\bar{\theta}_\ell^y(\varepsilon)$ are, for each y, ε, convex combinations of the $\hat{\theta}_{i\ell}(\varepsilon)$, and the curves θ_ℓ^y are just reparametrized versions of the $\bar{\theta}_\ell^y$. This implies that the curves θ_ℓ^y are $\mathcal{U}_\ell$-valued. It follows that

$$\theta_\ell^y(\varepsilon) - u_* \in \mathcal{U}_\ell^0 \quad \text{whenever} \quad \varepsilon \in [0, \bar{\varepsilon}_2],\ y \in \Sigma.$$

Define

$$\theta^y(\varepsilon) = u_* + \sum_{\ell=1}^L \Big(\theta_\ell^y(\varepsilon) - u_*\Big).$$

Then the curves θ^y take values in the set $u_* + \mathcal{U}_1^0 + \mathcal{U}_2^0 + \cdots + \mathcal{U}_L^0$, which by our assumptions is a subset of $\mathcal{U}$. So the θ^y are $\mathcal{U}$-valued curves. Moreover,

$$\theta^y(\varepsilon) = u_* + \left(\varepsilon\sum_{\ell=1}^L\sqrt{\sum_{i=0}^n s_i(y)\beta_{i\ell}}\right)v_\ell + \varepsilon^2\Big(\sum_{\ell=1}^L\sum_{i=0}^n s_i(y)\beta_{i\ell}w_{i\ell}\Big) + o(\varepsilon^2) \quad \text{as} \quad \varepsilon \downarrow 0.$$

It is clear from our construction that the map $[0, \bar{\varepsilon}_2] \times \Sigma \ni (\varepsilon, y) \mapsto \theta^y(\varepsilon)$ is continuous. (The only step which in principle could have caused some problems is the definition of $\tilde{\theta}_\ell^y(\varepsilon)$ in (2.19), where the denominator $\sum_{i'=0}^n s_{i'}(y)\beta_{i'\ell}$ occurs. Notice, however, that we have taken the precaution to make sure that all the numbers $\beta_{i\ell}$ are strictly positive, and this ensures that the denominator never vanishes, since $\sum_{i'=0}^n s_{i'}(y) = 1$.)

Finally, we let

$$\eta^y(\varepsilon) = \delta\zeta^y(\varepsilon) + (1-\delta)\theta^y(\varepsilon) \quad \text{for} \quad y \in \Sigma,\, \varepsilon \in [0, \bar{\varepsilon}],$$

where $\bar{\varepsilon} = \min(\bar{\varepsilon}_1, \bar{\varepsilon}_2)$.

Then the η^y satisfy all the desired conditions, and the proof of Lemma 2.4 is complete. ♢

Continuation of the proof of Theorem 2.2. We apply Lemma 2.4, taking δ to be the number $\bar{\delta}$ introduced earlier. The lemma allows us to choose a family $\{\eta^y\}_{y\in\Sigma}$ of continuous $\mathcal{U}$-valued curves on an interval $[0, \bar{\varepsilon}]$, $\bar{\varepsilon} > 0$, such that $\eta^y(\varepsilon)$ is jointly continuous in ε, y, and (2.17) holds.

We then define, for $\varepsilon \in [0,\bar{\varepsilon}]$, $y \in \Sigma$, a vector $\mu_\varepsilon(y) \in Y$ by letting

$$\mu_\varepsilon(y) = \frac{1}{\varepsilon^2}\left(\mathcal{F}\Big(\eta^y(\varepsilon)\Big) - \mathcal{F}(u_*) - \Big(\varepsilon^2 + \xi_\varepsilon(y)\Big)\omega\right),$$

where

$$\begin{aligned}\xi_\varepsilon(y) \quad \stackrel{\text{def}}{=} \quad & \left(\varepsilon(1-\delta)\sum_{\ell=1}^{L}\tau_\ell\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell}}\right) \\ & +\left(\frac{\varepsilon^2(1-\delta)^2}{2}\sum_{1\le\ell,\ell'\le L,\ell\ne\ell'}\theta_{\ell,\ell'}\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell}}\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell'}}\right).\end{aligned}$$

Then μ_ε is a continuous map from Σ to Y, and ξ_ε is a continuous map from Σ to $\mathbb{R}$ with nonnegative values.

The next step is to use (2.17) to expand $\mu_\varepsilon(y)$ to second powers of ε. For this purpose, we first expand $\mathcal{F}\Big(\eta^y(\varepsilon)\Big)$. The result is as follows.

$$\begin{aligned}\mathcal{F}\Big(\eta^y(\varepsilon)\Big) \quad = \quad & \mathcal{F}(u_*) + \varepsilon\left((1-\delta)\sum_{\ell=1}^{L}\left(\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell}}\right)P_1\cdot\hat{v}_\ell\right) \\ & +\varepsilon^2\left(\sum_{j=1}^{J}\sum_{i=0}^{n}s_i(y)\alpha_{ij}P_1\cdot v_j\right) + \varepsilon^2(1-\delta)\left(\sum_{\ell=1}^{L}\sum_{i=0}^{n}s_i(y)\beta_{i\ell}P_1\cdot w_{i\ell}\right) \\ & +\frac{\varepsilon^2}{2}\left((1-\delta)^2\sum_{\ell=1}^{L}\sum_{\ell'=1}^{L}\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell}}\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell'}}\right)P_2(\hat{v}_\ell,\hat{v}_{\ell'}) \\ & +o(\varepsilon^2).\end{aligned}$$

We now use (2.15) and (2.16) to replace $P_1\cdot\hat{v}_\ell$ by $\tau_\ell\omega$ and $P_2(\hat{v}_\ell,\hat{v}_{\ell'})$ by $\theta_{\ell,\ell'}\omega$ when $\ell\ne\ell'$. This yields:

$$\begin{aligned}\mathcal{F}\Big(\eta^y(\varepsilon)\Big) \quad = \quad & \mathcal{F}(u_*) + \left(\varepsilon(1-\delta)\sum_{\ell=1}^{L}\tau_\ell\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell}}\right)\omega \\ & +\varepsilon^2\left(\sum_{j=1}^{J}\sum_{i=0}^{n}s_i(y)\alpha_{ij}P_1\cdot v_j\right) + \varepsilon^2(1-\delta)\left(\sum_{\ell=1}^{L}\sum_{i=0}^{n}s_i(y)\beta_{i\ell}P_1\cdot w_{i\ell}\right) \\ & +\left(\frac{\varepsilon^2(1-\delta)^2}{2}\sum_{\ell=1}^{L}\Big(\sum_{i=0}^{n}s_i(y)\beta_{i\ell}\Big)\right)P_2(\hat{v}_\ell,\hat{v}_\ell) \\ & +\left(\frac{\varepsilon^2(1-\delta)^2}{2}\sum_{1\le\ell,\ell'\le L,\ell\ne\ell'}\theta_{\ell,\ell'}\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell}}\sqrt{\sum_{i=0}^{n}s_i(y)\beta_{i\ell'}}\right)\omega \\ & +o(\varepsilon^2).\end{aligned}$$

It follows that

$$\begin{aligned}\mathcal{F}\Big(\eta^y(\varepsilon)\Big) &= \mathcal{F}(u_*) + \xi_\varepsilon(y)\omega + o(\varepsilon^2) \\ &\quad + \varepsilon^2 \sum_{i=0}^{n} s_i(y) \Bigg(\sum_{j=1}^{J} \alpha_{ij} P_1 \cdot v_j + (1-\delta)\Big(\sum_{\ell=1}^{L} \beta_{i\ell} P_1 \cdot w_{i\ell} \Big) \\ &\quad + \frac{(1-\delta)^2}{2} \Big(\sum_{\ell=1}^{L} \beta_{i\ell} P_2(\hat{v}_\ell, \hat{v}_\ell \Big) \Bigg) \\ &= \mathcal{F}(u_*) + \xi_\varepsilon(y)\omega + \varepsilon^2 \sum_{i=0}^{n} s_i(y)\omega_i + o(\varepsilon^2) \\ &= \mathcal{F}(u_*) + \xi_\varepsilon(y)\omega + \varepsilon^2 y + o(\varepsilon^2)\,. \end{aligned}$$

Therefore,

$$\begin{aligned}\mu_\varepsilon(y) &= \frac{1}{\varepsilon^2} \Bigg(\mathcal{F}\Big(\eta^y(\varepsilon)\Big) - \mathcal{F}(u_*) - (\varepsilon^2 + \xi_\varepsilon(y))\omega \Bigg) \\ &= \frac{1}{\varepsilon^2} \Big(\varepsilon^2 (y-\omega) + o(\varepsilon^2) \Big) \\ &= y - \omega + o(1)\,, \end{aligned}$$

as $\varepsilon \downarrow 0$, the $o(1)$ being uniform with respect to $y \in \Sigma$.

Fix an ε such that $\varepsilon > 0$ and

$$\|\mu_\varepsilon(y) - (y-\omega)\| \leq \frac{\rho}{2} \qquad \text{for all} \qquad y \in \Sigma\,. \tag{2.20}$$

Let $K = \{\eta^y(\varepsilon) : y \in \Sigma\}$ so K is a compact subset of $\mathcal{U}$. Let $\{\sigma_j\}_{j\in\mathbb{N}}$ be a sequence of maps from $\mathcal{U}$ to $\mathcal{A}$ such that $\mathcal{F} \circ \sigma_j$ is continuous for each j and $\mathcal{F} \circ \sigma_j \to \mathcal{F}$ uniformly on compact subsets of $\mathcal{U}$. Pick j such that

$$\|\mathcal{F}(\sigma_j(u)) - \mathcal{F}(u)\| \leq \frac{\varepsilon^2 \rho}{2} \qquad \text{whenever} \qquad u \in K\,.$$

Define

$$\mu_\varepsilon^j(y) = \frac{1}{\varepsilon^2} \Bigg(\mathcal{F}\Big(\sigma_j(\eta^y(\varepsilon))\Big) - \mathcal{F}(u_*) - (\varepsilon^2 + \xi_\varepsilon(y))\omega \Bigg)\,.$$

Then $\|\mu_\varepsilon^j(y) - \mu_\varepsilon(y)\| \leq \frac{\rho}{2}$ for all $y \in \Sigma$. Therefore

$$\|\mu_\varepsilon^j(y) - (y-\omega)\| \leq \rho \qquad \text{for all} \qquad y \in \Sigma\,. \tag{2.21}$$

Now, if $y \in \mathcal{B}$, we have

$$\|y - \mu_\varepsilon^j(y) - \omega\| = \|\mu_\varepsilon^j(y) - (y-\omega)\| \leq \rho\,,$$

so that $y - \mu_\varepsilon(y)^j \in \mathcal{B}$. Therefore the continuous map

$$\mathcal{B} \ni y \mapsto y - \mu_\varepsilon^j(y) \in Y$$

is in fact a map from $\mathcal{B}$ to $\mathcal{B}$. By the Brouwer fixed point theorem, this map has a fixed point. Therefore, there exists a $y \in \mathcal{B}$ such that $y - \mu_\varepsilon^j(y) = y$, and then this y satisfies $\mu_\varepsilon^j(y) = 0$. We have thus shown that *for every sufficiently small positive* ε *there exists a* $j(\varepsilon) \in \mathbb{N}$ *such that the equation* $\mu_\varepsilon^j(y) = 0$ *has a solution* $y^{j,\varepsilon} \in \Sigma$ *for every* $j \in \mathbb{N}$ *such that* $j \geq j(\varepsilon)$.

Now, if $y^{j,\varepsilon}$ is a solution for an ε such that $\varepsilon > 0$ and a $j \in \mathbb{N}$, then the fact that $\mu^j_\varepsilon(y^{j,\varepsilon}) = 0$ says that

$$\mathcal{F}(\sigma_j(\eta^{y^{j,\varepsilon}}(\varepsilon))) = y_* + (\varepsilon^2 + \xi_\varepsilon(y^{j,\varepsilon}))\omega\,.$$

So

$$\mathcal{F}(\sigma_j(\eta^{y^{j,\varepsilon}}(\varepsilon))) \in \mathcal{F}(\mathcal{A}) \cap \Lambda\,.$$

Moreover,

$$\mathcal{F}(\sigma_j(\eta^{y^{j,\varepsilon}}(\varepsilon))) \neq y_*\,,$$

because $\varepsilon > 0$, $\xi_\varepsilon(y^{j,\varepsilon}) \geq 0$, and $\omega \neq 0$. Therefore (SEP) fails to hold.

Our proof of Theorem 2.2 is thus complete. $\diamondsuit$

3. Second-order necessary condition for a minimum

THEOREM 3.1. *Assume that*

1. $\mathcal{U}^a$ *and* Z *are normed spaces,*
2. Z *is finite-dimensional,*
3. $\mathcal{U}$ *is a convex subset of* $\mathcal{U}^a$*,*
4. $u_* \in \mathcal{U}$*,*
5. $\mathcal{E} : \mathcal{U} \mapsto Z$ *is a map,*
6. $\mathcal{I} : \mathcal{U} \mapsto \mathbb{R}$ *is a function,*
7. $\mathcal{A}$ *is an* $(\mathcal{U}, \mathcal{F})$*-abundant subset of* $\mathcal{U}$ *such that* $u_* \in \mathcal{A}$*, where* $\mathcal{F}$ *is the map*
$$\mathcal{U} \ni u \mapsto \mathcal{F}(u) \stackrel{\text{def}}{=} (\mathcal{E}(u), \mathcal{I}(u)) \in Y \stackrel{\text{def}}{=} Z \times \mathbb{R}\,.$$
8. $\mathcal{E}$ *and* $\mathcal{I}$ *are of class* D^2 *at* u_**.*
9. P_1*,* p_1 *are continuous linear maps from* $\mathcal{U}^a$ *to* Z*,* $\mathbb{R}$*, respectively,* P_2*,* p_2 *are continuous bilinear symmetric maps from* $\mathcal{U}^a \times \mathcal{U}^a$ *to* Z*,* $\mathbb{R}$*,*
$$\mathcal{E}(u_* + v) = \mathcal{E}(u_*) + P_1 \cdot v + \frac{1}{2}P_2(v, v) + o(\|v\|^2) \text{ as } v \to 0\,,\ u_* + v \in \mathcal{U}\,,$$
and
$$\mathcal{I}(u_* + v) = \mathcal{I}(u_*) + p_1 \cdot v + \frac{1}{2}p_2(v, v) + o(\|v\|^2) \text{ as } v \to 0\,,\ u_* + v \in \mathcal{U}\,,$$
10. $\hat{\mathcal{V}}$ *is a finite subset of* $var_1(u_*, \mathcal{U})$*.*
11. $P_1 \cdot v = 0$ *and* $p_1 \cdot v \leq 0$ *whenever* $v \in \hat{\mathcal{V}}$*,*
12. $P_2(v_1, v_2) = 0$ *and* $p_2(v_1, v_2) \leq 0$ *whenever* $v_1, v_2 \in \hat{\mathcal{V}}$ *are such that* $v_1 \neq v_2$*,*
13. $\{\mathcal{U}_v\}_{v \in \hat{\mathcal{V}}}$ *is a family of convex subsets of* $\mathcal{U}^a$ *such that*
    i. $u_* \in \mathcal{U}_v$ *for every* $v \in \hat{\mathcal{V}}$*,*
    ii. $u_* + \sum_{v \in \hat{\mathcal{V}}} \Big(\mathcal{U}_v - u_*\Big) \subseteq \mathcal{U}$*,*
    iii. $v \in var_1(u_*, \mathcal{U}_v)$ *for each* $v \in \hat{\mathcal{V}}$*.*

Let

$$z_* = \mathcal{E}(u_*)\,. \tag{3.1}$$

Assume that

$$\mathcal{I}(u_*) \leq \mathcal{I}(u) \qquad \text{for all} \qquad u \in \mathcal{A} \qquad \text{such that} \qquad \mathcal{E}(u) = z_*\,. \tag{3.2}$$

Then there exist a covector $\psi \in Z^*$ *and a real number* ψ_0 *such that*

a. $\psi_0 \geq 0$*,*

b. $\|\psi\| + \psi_0 > 0$,
c. $\langle\psi, P_1 \cdot v\rangle + \psi_0 p_1 \cdot v \geq 0$ *for all* $v \in \overline{var}_1(u_*, \mathcal{U})$,
d. $\langle\psi, P_1 \cdot w\rangle + \psi_0 p_1 \cdot w + \frac{1}{2}\Big(\langle\psi, P_2(v,v)\rangle + \psi_0 p_2(v,v)\Big) \geq 0$ *for all* (v, w) *such that* $v \in \hat{\mathcal{V}}$ *and* $w \in var_{2,1}(u_*, v, \mathcal{U}_v)$.
e. $\langle\psi, P_1 \cdot w\rangle + \psi_0 p_1 \cdot w$ *for all* (v, w) *such that* $v \in \hat{\mathcal{V}}$ *and* $w \in var_{2,1}(u_*, v, \mathcal{U}_v)$.

Proof. Define

$$c_* = \mathcal{I}(u_*)\,, \qquad y_* = (z_*, c_*) = \mathcal{F}(u_*)\,.$$

Let

$$\omega = (0, -1) \in Z \times \mathbb{R} = Y\,.$$

Let Λ be the closed half-line

$$\Lambda \stackrel{\mathrm{def}}{=} \{y_* + t\omega : t \in \mathbb{R}, t \geq 0\}\,. \tag{3.3}$$

Then $y_* \in \mathcal{F}(\mathcal{A}) \cap \Lambda$. Assumption (3.2) says precisely that $\mathcal{F}(\mathcal{A}) \cap \Lambda = \{y_*\}$, i.e., that the separation condition of the previosu section is satisfied.

If we then apply Corollary 2.3, we get a nonzero covector $\bar{\psi}$ that satisfies the inequalities of that corollary. Write $\bar{\psi} = (\psi, \psi_0)$. Then the inequality $\langle\bar{\psi}, \omega\rangle \leq 0$ implies that $\psi_0 \geq 0$. All the other desired conclusions follow immediately from the properties of $\bar{\psi}$ stated in Corollary 2.3. $\diamondsuit$

We now restate Theorem 3.2 for the case when $\hat{\mathcal{V}}$ consists of a single vector.

Theorem 3.2. *Assume that*

1. $\mathcal{U}^a$ *and* Z *are normed spaces,*
2. Z *is finite-dimensional,*
3. $\mathcal{U}$ *is a convex subset of* $\mathcal{U}^a$,
4. $u_* \in \mathcal{U}$,
5. $\mathcal{E} : \mathcal{U} \mapsto Z$ *is a map,*
6. $\mathcal{I} : \mathcal{U} \mapsto \mathbb{R}$ *is a function,*
7. $\mathcal{A}$ *is an* $(\mathcal{U}, \mathcal{F})$*-abundant subset of* $\mathcal{U}$ *such that* $u_* \in \mathcal{A}$, *where* $\mathcal{F}$ *is the map*
$$\mathcal{U} \ni u \mapsto \mathcal{F}(u) \stackrel{\mathrm{def}}{=} (\mathcal{E}(u), \mathcal{I}(u)) \in Y \stackrel{\mathrm{def}}{=} Z \times \mathbb{R}\,.$$
8. $\mathcal{E}$ *and* $\mathcal{I}$ *are of class* D^2 *at* u_*.
9. P_1, p_1 *are continuous linear maps from* $\mathcal{U}^a$ *to* Z, $\mathbb{R}$, *respectively,* P_2, p_2 *are continuous bilinear symmetric maps from* $\mathcal{U}^a \times \mathcal{U}^a$ *to* Z, $\mathbb{R}$,
$$\mathcal{E}(u_* + v) = \mathcal{E}(u_*) + P_1 \cdot v + \frac{1}{2}P_2(v,v) + o(\|v\|^2) \ \text{ as } \ v \to 0\,, \ u_* + v \in \mathcal{U}\,,$$
and
$$\mathcal{I}(u_* + v) = \mathcal{I}(u_*) + p_1 \cdot v + \frac{1}{2}p_2(v,v) + o(\|v\|^2) \ \text{ as } \ v \to 0\,, \ u_* + v \in \mathcal{U}\,,$$
10. $\hat{v}$ *is a member of* $var_1(u_*, \mathcal{U})$,
11. $P_1 \cdot \hat{v} = 0$ *and* $p_1 \cdot \hat{v} \leq 0$.

Let

$$z_* = \mathcal{E}(u_*)\,. \tag{3.4}$$

Assume that

$$\mathcal{I}(u_*) \le \mathcal{I}(u) \quad \text{for all} \quad u \in \mathcal{A} \quad \text{such that} \quad \mathcal{E}(u) = z_* \,. \tag{3.5}$$

Then there exist a covector $\psi \in Z^*$ and a real number ψ_0 such that

a. $\psi_0 \ge 0$,
b. $\|\psi\| + \psi_0 > 0$,
c. $\langle \psi, P_1 \cdot v\rangle + \psi_0 p_1 \cdot v \ge 0$ *for all* $v \in \overline{var}_1(u_*, \mathcal{U})$,
d. $\langle \psi, P_1 \cdot w\rangle + \psi_0 p_1 \cdot w + \frac{1}{2}\Big(\langle \psi, P_2(\hat{v}, \hat{v})\rangle + \psi_0 p_2(\hat{v}, \hat{v})\Big) \ge 0$ *for all vectors* $w \in var_{2,1}(u_*, \hat{v}, \mathcal{U})$,
e. $\langle \psi, P_1 \cdot w\rangle + \psi_0 p_1 \cdot w$ *for all* $w \in var_{2,1}(u_*, v, \mathcal{U}_v)$.

Proof. It suffices to apply Theorem 3.2 with $\hat{\mathcal{V}} = \{\hat{v}\}$, and to observe that the cross-term condition 12 of the statement of Theorem 3.2 is vacuously true, and the independence condition 13 can be satisfied by just taking $\mathcal{U}_{\hat{v}} = \mathcal{U}$. ♢

4. Optimal control

We consider a control system

$$\dot{x} = f(x, u, t)\,, \quad x \in \mathbb{R}^n\,, \quad u \in U\,, a \le t \le b\,, \tag{4.1}$$

and a Lagrangian

$$\mathbb{R}^n \times U \times [a, b] \ni (x, u, t) \mapsto L(x, u, t) \in \mathbb{R}\,. \tag{4.2}$$

We assume that

B1. $m \in \mathbb{N}$ *and* U *is a closed convex subset of* $\mathbb{R}^m$,
B2. *the partial maps*

$$\mathbb{R}^n \times U \ni (x, u) \mapsto (f(x, u, t), L(x, u, t)) \in \mathbb{R}^{n+1}$$

are of class C^2 *for every* $t \in [a, b]$,
B3. *the partial maps*

$$[a, b] \ni t \mapsto (f(x, u, t), L(x, u, t)) \in \mathbb{R}^{n+1}$$

are measurable for every $(x, u) \in \mathbb{R}^n \times U$,
B4. *the maps of condition B2 are bounded, together with their first and second derivatives with respect to* (x, u)*, on every compact subset of* $\mathbb{R}^n \times U \times [a, b]$.

Warning!. As stated, the conditions do no guarantee that trajectories do not explode. So, strictly speaking, the trajectory map defined below need not be everywhere defined, and the same is true, naturally, of $\mathcal{E}$ and $\mathcal{I}$. This does not cause any problems, but in the final writing of this note this issue has to be taken care of. At this point, we may, if we wish to be absolutely sure that everything is O.K., assume that f_x is *globally bounded.* ♢

We let $\mathcal{U}^a = L^\infty([a, b], \mathbb{R}^m)$, and let $\mathcal{U}$ be the set of those controls $u(\cdot) \in \mathcal{U}^a$ that are U-valued.

A *trajectory-control pair* is a pair $(x(\cdot), u(\cdot))$ consisting of an absolutely continuous curve $x(\cdot) : [a, b] \mapsto \mathbb{R}^n$ and a control $u(\cdot) : [a, b] \mapsto U$ such that

$$\dot{x}(t) = f(x(t), u(t), t) \qquad \text{for almost all} \quad t \in [a, b]\,.$$

We fix an initial state $\bar{x}$ and a terminal state $\bar{x}$, and let $TCP(\bar{x}, \hat{x})$ denote the set of all trajectory-control pairs $(x(\cdot), u(\cdot))$ such that $x(a) = \bar{x}$ and $x(b) = \hat{x}$.

We assume that we are given a pair $(x_*(\cdot), u_*(\cdot))$ such that

B5. $(x_*(\cdot), u_*(\cdot)) \in TCP(\bar{x}, \hat{x})$, *and* $(x_*(\cdot), u_*(\cdot))$ *minimizes the functional*

$$TCP(\bar{x}, \hat{x}) \ni (x(\cdot), u(\cdot)) \mapsto J(x(\cdot), u(\cdot)) \stackrel{\text{def}}{=} \int_a^b L(x(t), u(t), t)dt\,.$$

We assume that we are also given pairs $(h_1, v_1), \ldots, (h_L, v_L)$ that are trajectory-control pairs of the linearized equation about $(x_*(\cdot), u_*(\cdot))$, supported by pairwise disjoint subintervals $I_1, \ldots, I_\ell$. More precisely, we assume that

B6. $v_1, \ldots, v_L \in \mathcal{U}^a$, $h_1, \ldots, h_L$ *are absolutely continuous maps from* $[a, b]$ *to* $\mathbb{R}^n$ *such that*

$$h_1(a) = \ldots = h_L(a) = 0$$

and

$$\dot{h}_\ell(t) = \frac{\partial f}{\partial x}(x_*(t), u_*(t), t) \cdot h_\ell(t) + \frac{\partial f}{\partial u}(x_*(t), u_*(t), t) \cdot v_\ell(t)$$

for a.e. $t \in [a, b]$, *for* $\ell = 1, \ldots, L$.

Moreover, we assume that

B7. $h_1(b) = \ldots = h_L(b) = 0$,
B8. $\int_a^b L_x(x_*(t), v_*(t), t) \cdot h_\ell(t) + \int_a^b L_u(x_*(t), v_*(t), t) \cdot v_\ell(t) \leq 0$ *for* $\ell = 1, \ldots, L$,
B9. $I_1, \ldots, I_\ell$ *are pairwise disjoint subintervals of* $[a, b]$ *such that*

$$\{t \in [a, b] : v_\ell(t) \neq 0\} \subseteq I_\ell \qquad for \qquad \ell = 1, \ldots, L\,. \tag{4.3}$$

We remark that conditions B6, B7 and B9 imply that

$$\{t \in [a, b] : h_\ell(t) \neq 0\} \subseteq I_\ell \qquad \text{for} \qquad \ell = 1, \ldots, L\,. \tag{4.4}$$

We let $\mathcal{T}$ be the *trajectory map*, i.e., the map that assigns to each $u \in \mathcal{U}$ the curve $\mathcal{T}(u)$ which is the unique solution $x(\cdot)$ of the initial value problem

$$\begin{cases} \dot{x}(t) &= f(x(t), u(t), t) \quad \text{for a.e.} \quad t \in [a, b]\,, \\ x(a) &= \bar{x}\,. \end{cases}$$

Then $\mathcal{T}$, regarded as a map from $\mathcal{U}$ to the space $C^0([a, b], \mathbb{R}^n)$, is twice Fréchet differentiable. The derivatives of $\mathcal{T}$ at u_* are easily computed by observing that

$$\mathcal{T}(u_* + \varepsilon v) = x_* + \varepsilon \mathcal{T}'(u_*) \cdot v + \frac{\varepsilon^2}{2} \mathcal{T}''(u_*)(v, v) + o(\varepsilon^2)\,.$$

If we let

$$\mathcal{T}(u_* + \varepsilon v) = x^\varepsilon\,, \quad y^\varepsilon(t) = \frac{d}{d\varepsilon}\Big(x^\varepsilon(t)\Big)\,, \quad z^\varepsilon(t) = \frac{d}{d\varepsilon}\Big(y^\varepsilon(t)\Big)\,,$$

we have

$$x^\varepsilon(t) = \bar{x} + \int_a^t f(x^\varepsilon(s), u_*(s) + \varepsilon v(s), s)\, ds\,,$$

so that

$$y^\varepsilon(t) = \int_a^t f_x(x^\varepsilon(s), u_*(s) + \varepsilon v(s), s) \cdot y^\varepsilon(s)\, ds + \int_a^t f_u(x^\varepsilon(s), u_*(s) + \varepsilon v(s), s) \cdot v(s)\, ds\,,$$

and

$$\begin{aligned} z^\varepsilon(t) &= \int_a^t f_x(x^\varepsilon(s), u_*(s)+\varepsilon v(s), s)\cdot z^\varepsilon(s)\,ds \\ &\quad + \int_a^t f_{xx}(x^\varepsilon(s), u_*(s)+\varepsilon v(s), s)(y^\varepsilon(s), y^\varepsilon(s))\,ds \\ &\quad +2\int_a^t f_{xu}(x^\varepsilon(s), u_*(s)+\varepsilon v(s), s)(y^\varepsilon(s), v(s))\,ds \\ &\quad + \int_a^t f_{uu}(x^\varepsilon(s), u_*(s)+\varepsilon v(s), s)(v(s), v(s))\,ds\,. \end{aligned}$$

It is clear that $\mathcal{T}'(u_*)\cdot v = y^0$, and $\mathcal{T}''(u_*)(v,v) = z^0$. Therefore

$$\begin{aligned} \Big(\mathcal{T}'(u_*)\cdot v\Big)(t) &= \int_a^t f_x(x_*(s), u_*(s), s)\cdot\Big(\mathcal{T}'(u_*)\cdot v\Big)(s)\,ds \\ &\quad + \int_a^t f_u(x_*(s), u_*(s), s)\cdot v(s)\,ds\,, \end{aligned}$$

and

$$\begin{aligned} &\Big(\mathcal{T}''(u_*)(v,v)\Big)(t) = \int_a^t f_x(x_*, u_*(s), s)\cdot\Big(\mathcal{T}''(u_*)(v,v)\Big)(s)\,ds \\ &+\int_a^t f_{xx}(x_*(s), u_*(s), s)\Big(\Big(\mathcal{T}'(u_*)\cdot v\Big)(s), \Big(\mathcal{T}'(u_*)\cdot v\Big)(s)\Big)\,ds \\ &+2\int_a^t f_{xu}(x_*(s), u_*(s), s)\Big(\Big(\mathcal{T}'(u_*)\cdot v\Big)(s), v(s)\Big)\,ds \\ &+\int_a^t f_{uu}(x_*(s), u_*(s), s)(v(s), v(s))\,ds\,. \end{aligned}$$

It then follows easily that the symmetric bilinear form $(v,\tilde v)\mapsto \mathcal{T}''(u_*)(v,\tilde v)$ satisfies

$$\begin{aligned} &\Big(\mathcal{T}''(u_*)(v,\tilde v)\Big)(t) = \int_a^t f_x(x_*, u_*(s), s)\cdot\Big(\mathcal{T}''(u_*)(v,\tilde v)\Big)(s)\,ds \\ &+\int_a^t f_{xx}(x_*(s), u_*(s), s)\Big(\Big(\mathcal{T}'(u_*)\cdot v\Big)(s), \Big(\mathcal{T}'(u_*)\cdot \tilde v\Big)(s)\Big)\,ds \\ &+\int_a^t f_{xu}(x_*(s), u_*(s), s)\Big(\Big(\mathcal{T}'(u_*)\cdot v\Big)(s), \tilde v(s)\Big)\,ds \\ &+\int_a^t f_{xu}(x_*(s), u_*(s), s)\Big(\Big(\mathcal{T}'(u_*)\cdot \tilde v\Big)(s), v(s)\Big)\,ds \\ &+\int_a^t f_{uu}(x_*(s), u_*(s), s)(v(s), \tilde v(s))\,ds\,. \end{aligned}$$

In particular,

$$\Big(\mathcal{T}''(u_*)(v_\ell, v_{\ell'})\Big)(t) = \int_a^t f_x(x_*, u_*(s), s) \cdot \Big(\mathcal{T}''(u_*)(v_\ell, v_{\ell'})\Big)(s)\, ds$$
$$+ \int_a^t f_{xx}(x_*(s), u_*(s), s)\Big(h_\ell(s), h_{\ell'}(s)\Big)\, ds$$
$$+ \int_a^t f_{xu}(x_*(s), u_*(s), s)\Big(h_\ell(s), v_{\ell'}(s)\Big)\, ds$$
$$+ \int_a^t f_{xu}(x_*(s), u_*(s), s)\Big(h_{\ell'}(s), v_\ell(s)\Big)\, ds$$
$$+ \int_a^t f_{uu}(x_*(s), u_*(s), s)(v_\ell(s), v_{\ell'}(s))\, ds\,.$$

If $\ell \neq \ell'$, then the last four integrals vanish because of (4.3), (4.4), and the fact that $I_\ell \cap I_{\ell'} = \emptyset$. It follows that

$$\mathcal{T}''(u_*)(v_\ell, v_{\ell'}) = 0 \qquad \text{if} \qquad \ell \neq \ell'\,. \tag{4.5}$$

We let $\mathcal{C}$ be the *running cost map*, that is, the map $u \mapsto \zeta_u$, where

$$\zeta_u(t) = \int_a^t L(\mathcal{T}(u)(s), u(s), s)\, ds\,.$$

Then a similar calculation shows that

$$\mathcal{C}''(u_*)(v_\ell, v_{\ell'}) = 0 \qquad \text{if} \qquad \ell \neq \ell'\,. \tag{4.6}$$

We now let $\mathcal{E}$, $\mathcal{I}$ be, respectively, the *terminal point map* and the *terminal cost map*, i.e., the maps defined by

$$\mathcal{E}(u) = \mathcal{T}(u)(b)\,, \qquad \mathcal{I}(u) = \mathcal{C}(u)(b)\,.$$

Then all the conditions of Theorem 3.1 are satisfied, and we obtain the following result.

THEOREM 4.1. *If B1-B9 hold, then there exist an absolutely comtinuous field of covectors $[a, b] \ni t \mapsto \psi(t) \in \mathbb{R}_n$ (where we are writing $\mathbb{R}_n$ to denote the space of real n-dimensional row vectors) and a real number ψ_0 such that*

1. $\psi_0 \geq 0$,
2. $\psi_0 + \|\psi(b)\| > 0$,
3. $\dot\psi(t) = -\psi(t) \cdot f_x(x_*(t), u_*(t), t) - \psi_0 L_x(x_*(t), u_*(t), t)$ *for almost all* $t \in [a, b]$,
4. *for almost all $t \in [a, b]$, the inequality*

$$\Big(\psi_0 L_u(x_*(t), u_*(t), t) + \psi(t) \cdot f_u(x_*(t), u_*(t), t)\Big) \cdot (u - u_*(t)) \geq 0$$

   holds for all $u \in U$,
5. *for every $\ell \in \{1, \dots, L\}$, the inequalities*

$$\int_a^b \Big(\psi_0 L_u(x_*(t), u_*(t), t) + \psi(t) \cdot f_u(x_*(t), u_*(t), t)\Big) \cdot w(t)\, dt \geq 0$$

and

$$\int_a^b \Big(\psi_0 L_u(x_*(t),u_*(t),t)+\psi(t)\cdot f_u(x_*(t),u_*(t),t)\Big)\cdot w(t)\,dt$$
$$+\frac{1}{2}\Bigg(\int_a^b \Big[\psi(t)\cdot f_{xx}(x_*(t),u_*(t),t)\Big(h_\ell(t),h_\ell(t)\Big)$$
$$+\psi_0\cdot L_{xx}(x_*(t),u_*(t),t)\Big(h_\ell(t),h_\ell(t)\Big)\Big]\,dt$$
$$+2\int_a^b \Big[\psi(t)\cdot f_{xu}(x_*(t),u_*(t),t)\Big(h_\ell(t),v_\ell(t)\Big)$$
$$+\psi_0\cdot L_{xu}(x_*(t),u_*(t),t)\Big(h_\ell(t),v_\ell(t)\Big)\Big]\,dt$$
$$+\int_a^b \Big[\psi(t)\cdot f_{uu}(x_*(t),u_*(t),t)\Big(v_\ell(t),v_\ell(t)\Big)$$
$$+\psi_0\cdot L_{uu}(x_*(t),u_*(t),t)\Big(v_\ell(t),v_\ell(t)\Big)\Big]\,dt\Bigg)\ge 0$$

hold for every $w\in var_{2,1}(u_*,v_\ell,\mathcal{U})$ *such that* $\{t : w(t)\neq 0\}\subseteq I_\ell$.

Proof. We apply Theorem 3.1 to the maps $\mathcal{E}$ and $\mathcal{I}$, observing that

1. the cross-term condition (Hypothesis 12 of the statement of Theorem 3.1) is automatically satisfied, because of (4.5) and (4.6),
2. the independence condition (Hypothesis 13 of the statement of Theorem 3.1) is satisfied if we take

$$\mathcal{U}_\ell=\{u\in\mathcal{U} : u(t)=u_*(t) \text{ whenever } t\notin I_\ell\}\,. \tag{4.7}$$

More precisely: if we define $\mathcal{U}_\ell$ by (4.7), then either $v_\ell\in var_1(u_*,\mathcal{U})$, in which case Hypothesis 13 of the statement of Theorem 3.1 is valid, or $v_\ell\notin var_1(u_*,\mathcal{U})$, in which case the set $var_{2,1}(u_*,v_\ell,\mathcal{U})$ is empty, and we can simply omit v_ℓ.

The conclusions follow by applying the formulas for $\mathcal{T}'$, $\mathcal{T}''$, $\mathcal{I}'$ and $\mathcal{I}''$ obtained earlier, except for the fact that condition 4 is obtained in an integrated form, as the statement that

$$\int_a^b \Big(\psi_0 L_u(x_*(t),u_*(t),t)+\psi(t)\cdot f_u(x_*(t),u_*(t),t)\Big)\cdot\Big(u(t)-u_*(t)\Big)\,dt\ge 0$$

for all $u(\cdot)\in\mathcal{U}$. To go from this to the form given in our conclusion no. 4, it suffices to take t to be a Lebesgue point of the functions

$$s\mapsto \psi_0 L_u(x_*(s),u_*(s),s)+\psi(s)\cdot f_u(x_*(s),u_*(s),s)$$

and

$$s\mapsto \Big(\psi_0 L_u(x_*(s),u_*(s),s)+\psi(s)\cdot f_u(x_*(s),u_*(s),s)\Big)\cdot u_*(s)\,,$$

and let $u(\cdot)$ be an arbitrary constant control. ♢

5. Comparison with the result of Ledzewicz and Schättler

If we take $\ell=1$, then Theorem 4.1 contains as a special case Theorem 4.1 of [**1**], with the following additional features:

1. the additional multiplier ρ of Theorem 4.1 of [**1**] can be taken to be zero,
2. in Hypothesis (A1) of Page 166 of [**1**], the requirement that f be of class C^3 in x and u can be weakened by changing "C^3" to "C^2,"
3. Hypothesis (c) in Page 168 of [**1**] is not needed,
4. the "second-order feasible set" $FC^{(2)}(\mathcal{U}, u_*, v)$ of [**1**] is replaced by the set $var_{2,1}(u_*, v, \mathcal{U})$, which in general is much larger,
5. Hypothesis (e) in Page 168 of [**1**], on the convexity of $FC^{(2)}(\mathcal{U}, u_*, v)$, is not needed.

6. Things to be done

1. Show in detail how the Goh and Legendre-Clebsch conditions follow from our theorem,
2. bring in the "abundant set" stuff to show, via relaxed controls, that Theorem 4.1 yields a *global* as well as a local maximum principle,
3. take care of the possibility of explosions mentioned in the "Warning" in §5; a related issue is that of taking the state space to be an open subset of $\mathbb{R}^n$ rather than all of $\mathbb{R}^n$, or a manifold; these things are of course trivial, but writing them down will take some work,
4. discuss in detail the case of a nonconvex control set,
5. bring in more general terminal conditions; the appropriate setting should be a terminal condition $x(b) \in S$, where S is a fairly general set having a tangent cone (in some adequate sense, e.g., Boltyanskii approximating cone) C at $x_*(b)$; the conclusion would then contain the transversality condition $\langle \psi(b), v \rangle \leq 0$ for all $v \in C$,
6. do the higher-order case, $p > 2$.

7. A big mystery

Will the high-order multipliers also go away for $p > 2$? This author's guess is "yes." But it's only a guess.

Stay tuned.

References

[1] Ledzewicz, U., and H. Schättler, "An extended maximum principle." *Nonlineanr Analysis, Theory, Methods and Applications* **29**, 1997, pp. 159-183.