IL PRINCIPIO DEL MASSIMO DI PONTRYAGIN

GIORGIO STEFANI

SOMMARIO. In questa nota studiamo il Principio del Massimo di Pontryagin, così come presentato nel corso di *Calcolo delle Variazioni* tenuto dal Professor Franco Rampazzo durante il secondo semestre dell'a.a. 2014 - 2015.

1. Preliminari

In questi appunti studiamo il Principio del Massimo di Pontryagin, nel caso non vincolato e nel caso vincolato, così come l'abbiamo incontrato durante il corso di *Calcolo delle Variazioni*. In questa sezione raccogliamo alcuni risultati ausiliari di cui avremo bisogno durante le dimostrazioni dei teoremi principali.

1.1. **Dualità per problemi di Cauchy lineari.** Il seguente risultato è una semplice relazione tra l'equazione differenziale lineare $\dot{v} = Av$ e la sua duale $\dot{p} = -pA$.

Lemma 1.1 (dualità per problemi di Cauchy lineari)**.** *Sia $A\colon [a,b] \to M_{n\times n}(\mathbb{R})$ una funzione a valori matriciali tale che $\int_a^b \|A(t)\|_{op}\, dt < \infty$. Sia $v \in AC([a,b];\mathbb{R}^n)$ la soluzione del problema di Cauchy lineare (ai valori iniziali)*

$$\begin{cases} \dot{v}(t) = A(t)v(t), & t \in (a,b) \\ v(a) = \bar{v} \end{cases}$$

e sia $p \in AC([a,b];\mathbb{R}^n)$ la soluzione del problema di Cauchy lineare duale (ai valori finali)

$$\begin{cases} \dot{p}(t) = -p(t)A(t), & t \in (a,b) \\ p(b) = \bar{p}. \end{cases}$$

Allora $t \mapsto p(t)v(t)$ è una funzione costante su $[a,b]$.

Dimostrazione. Infatti $t \mapsto p(t)v(t)$ è assolutamente continua su $[a,b]$ perché prodotto di funzioni assolutamente continue, e

$$\frac{d}{dt}\Big(p(t)v(t)\Big) = \dot{p}(t)v(t) + p(t)\dot{v}(t) = -p(t)A(t)v(t) + p(t)A(t)v(t) = 0.$$

□

1.2. **Il Lemma di Grönwall.** Ricordiamo il seguente risultato sulle disuguaglianze differenziali.

Lemma 1.2 (Grönwall)**.** *Siano $u \in AC([a,b];\mathbb{R})$ e $\alpha, \beta \in L^1([a,b];\mathbb{R})$ tali che $u, \alpha, \beta \geq 0$ q.o. in $[a,b]$. Supponiamo che*

$$\dot{u}(t) \leq \alpha(t)u(t) + \beta(t)$$

per q.o. $t \in [a,b]$. Allora

$$u(t) \leq u(a)e^{\int_a^t \alpha(s)ds} + \int_a^t \beta(s)e^{\int_s^t \alpha(\sigma)d\sigma}\, ds$$

per ogni $t \in [a,b]$.

Date: 27 maggio 2015.

1.3. **Il Teorema di differenziazione di Lebesgue.** Nella dimostrazione del Principio del Massimo avremo bisogno del seguente classico risultato, dovuto a H. Lebesgue.

Teorema 1.3 (Teorema di Differenziazione di Lebesgue)**.** *Sia $f\colon \mathbb{R}^n \to \mathbb{R}$ una funzione localmente integrabile, $f \in L^1_{loc}(\mathbb{R}^n)$. Allora, per q.o. $x \in \mathbb{R}^n$, si ha*[1]

$$\lim_{r\to 0} \fint_{B(x,r)} |f(y) - f(x)| \, dy = 0. \tag{1.1}$$

Un punto $x \in \mathbb{R}^n$ che soddisfa la (1.1) si dice *punto di Lebesgue* per la funzione f; indichiamo con $LP(f) \subset \mathbb{R}^n$ l'insieme dei punti di Lebesgue della funzione f. Il Teorema 1.3, in altre parole, afferma che l'insieme $LP(f)$ ha misura piena nel dominio della funzione f.

2. Il caso non vincolato

In questa sezione studiamo il problema di minimizzazione non vincolato

$$\begin{cases} \inf\limits_{(x,u)} \int_a^b l(t,x(t),u(t)) \, dt + g(x(b)) \\ \dot{x}(t) = f(t,x(t),u(t)), \ t \in (a,b) \\ x(a) = x_a \end{cases} \tag{P-nv}$$

dove $u\colon [a,b] \to U$ è detto *controllo* e U è detto *insieme dei controlli.*

2.1. **Notazioni per il caso non vincolato.** Fissiamo le notazioni. Diciamo che la coppia controllo-traiettoria (u,x) è *ammissibile* per il problema (P-nv) se risolve il problema di Cauchy

$$\begin{cases} \dot{x}(t) = f(t,x(t),u(t)), \ t \in (a,b) \\ x(a) = x_a. \end{cases}$$

La coppia controllo-traiettoria (u^*,x^*) è *ottimale* se è ammissibile ed è un minimizzante per il problema (P-nv), ovvero

$$\int_a^b l(t,x^*(t),u^*(t)) \, dt + g(x^*(b)) = \inf_{(u,x)} \int_a^b l(t,x(t),u(t)) \, dt + g(x(b))$$

dove l'estremo inferiore è fatto su tutte le coppie (u,x) ammissibili. Poniamo infine[2]

$$\begin{aligned} H(\lambda,t,x,u) &= \begin{pmatrix} -1 & \lambda \end{pmatrix} \begin{bmatrix} l(t,x,u) & f(t,x,u) \end{bmatrix} = \\ &= -l(t,x,u) + \lambda f(t,x,u) \end{aligned} \tag{2.1}$$

la *Hamiltoniana (non massimizzata)* del problema non vincolato (P-nv).

2.2. **Teorema del Massimo nel caso non vincolato.** Enunciamo e dimostriamo il risultato principale di questa sezione.

Teorema 2.1 (Principio del Massimo di Pontryagin[3], non vincolato)**.** *Sia dato il problema di minimizzazione non vincolato* (P-nv) *e sia (u^*,x^*) una coppia controllo-traiettoria ottimale. Facciamo le seguenti ipotesi.*

(i) La funzione $g\colon \mathbb{R}^n \to \mathbb{R}$ è di classe C^1.

[1] Si ricordi che $\fint_A = \frac{1}{|A|} \int_A$ per ogni $A \subset \mathbb{R}^n$ misurabile.

[2] Le matrici sono indicate con parentesi quadre, mentre i vettori riga e i vettori colonna con parentesi tonde.

[3] Lev Semyonovich Pontryagin (1908 - 1988) fu un matematico Sovietico. Nacque a Mosca e divenne cieco all'età di 14 anni. A dispetto della sua menomazione divenne uno dei più importanti matematici del 20^o secolo, grazie anche all'aiuto di sua madre, che gli leggeva libri e articoli di matematica.

(ii) Per ogni $(x,u) \in \mathbb{R}^n \times U$ fissato, la funzione

$$t \mapsto (l,f)(t,x,u)$$

è misurabile su $[a,b]$.

(iii) Per ogni $(t,u) \in [a,b] \times \mathbb{R}^n$ fissato, la funzione

$$x \mapsto (l,f)(t,x,u)$$

è di classe C^1 su $\mathbb{R}^n$.

(iv) Per ogni insieme compatto $Q \subset \mathbb{R}^n$ fissato, esiste una funzione $k \in L^1([a,b])$ tale che, per ogni $(x,u) \in Q \times U$,

$$\max\left\{ |(l,f)(t,x,u)|, \left|\frac{\partial(l,f)}{\partial x}(t,x,u)\right| \right\} \leq k(t) \quad \forall t \in [a,b].$$

Allora esiste una funzione $\lambda \in AC([a,b];\mathbb{R}^n)$ tale che

(1) (equazione aggiunta) *per q.o. $t \in [a,b]$ si ha*

$$\begin{aligned}\dot\lambda(t) &= -\begin{pmatrix}-1 & \lambda(t)\end{pmatrix}\begin{bmatrix}\frac{\partial l}{\partial x}(t,x^*(t),u^*(t)) & \frac{\partial f}{\partial x}(t,x^*(t),u^*(t))\end{bmatrix} = \\ &= \frac{\partial l}{\partial x}(t,x^*(t),u^*(t)) - \lambda(t)\frac{\partial f}{\partial x}(t,x^*(t),u^*(t));\end{aligned}$$

(2) (relazione di massimo **debole**) *per ogni $u \in U$ esiste un insieme misurabile $E_u \subset [a,b]$ di misura piena, $|[a,b] \smallsetminus E_u| = 0$, tale che*

$$H(\lambda(t),t,x^*(t),u^*(t)) \geq H(\lambda(t),t,x^*(t),u)$$

per ogni $t \in E_u$;

(3) (condizione di trasversalità) *si ha*

$$\lambda(b) = -\nabla g(x^*(b)).$$

Se, in aggiunta alle ipotesi precedenti, supponiamo che l'insieme dei controlli U sia un sottoinsieme di $\mathbb{R}^m$ e che, per ogni $t \in [a,b]$ fissato, si abbia

$$(x,u) \mapsto (l,f)(t,x,u) \in C(\mathbb{R}^n \times U), \tag{$\diamond$}$$

allora la (2) può essere migliorata come segue:

(2$\star$) (relazione di massimo **forte**) *per q.o. $t \in [a,b]$ si ha*

$$H(\lambda(t),t,x^*(t),u^*(t)) = \max_{u\in U} H(\lambda(t),t,x^*(t),u). \tag{$\star$}$$

Dimostrazione. Procediamo per step. Per il momento assumiamo solo le ipotesi (i), (ii), (iii) e (iv). L'ipotesi aggiuntiva ($\diamond$) verrà assunta più avanti, nello Step 9.

Step 1: problema equivalente. Il problema (P-nv) è equivalente al problema

$$\text{(P-nv equiv.)} \qquad \begin{cases}\inf\limits_{(x_0,x)} x_0(b) + g(x(b)) \\ \dot x_0(t) = l(t,x(t),u(t)),\ t \in (a,b) \\ x_0(a) = 0 \\ \dot x(t) = f(t,x(t),u(t)),\ t \in (a,b) \\ x(a) = x_a\end{cases}$$

Infatti basta osservare che

$$x_0(b) = x_0(a) + \int_a^b l(t,x(t),u(t))\,dt = \int_a^b l(t,x(t),u(t))\,dt.$$

In particolare, osserviamo che la soluzione $(x_0, x)\colon [a,b] \to \mathbb{R} \times \mathbb{R}^n$ esiste, è unica ed è assolutamente continua, perché le ipotesi (ii), (iii) e (iv) permettono di applicare il Teorema di esistenza e unicità al problema di Cauchy in (P-nv equiv.).

Step 2: definizione della funzione λ. Definiamo la funzione $(\lambda_0, \lambda) \in AC([a,b]; \mathbb{R} \times \mathbb{R}^n)$ come la soluzione del problema di Cauchy (lineare e ai valori terminali)

$$\begin{cases} \dot\lambda_0(t) = 0 \\ \dot\lambda(t) = -\begin{pmatrix}\lambda_0(t) & \lambda(t)\end{pmatrix} \begin{bmatrix} \dfrac{\partial l}{\partial x}(t, x^*(t), u^*(t)) & \dfrac{\partial f}{\partial x}(t, x^*(t), u^*(t)) \end{bmatrix}, \ t \in (a,b) \\ \begin{pmatrix}\lambda_0(b) & \lambda(b)\end{pmatrix} = -\begin{pmatrix} 1 & \nabla g(x^*(b))\end{pmatrix}. \end{cases}$$

Ricaviamo immediatamente che $\lambda_0(t) \equiv -1$, per cui il sistema si riduce a

$$\text{(PC-}\lambda\text{)} \qquad \begin{cases} \dot\lambda(t) = -\begin{pmatrix}1 & \lambda(t)\end{pmatrix} \begin{bmatrix} \dfrac{\partial l}{\partial x}(t, x^*(t), u^*(t)) & \dfrac{\partial f}{\partial x}(t, x^*(t), u^*(t)) \end{bmatrix}, \ t \in (a,b) \\ \lambda(b) = -\nabla g(x^*(b)). \end{cases}$$

La funzione $\lambda \in AC([a,b]; \mathbb{R}^n)$ soddisfa per definizione l'equazione aggiunta (1) e la condizione di trasversalità (3). Dobbiamo verificare che λ soddisfa anche la relazione di massimo debole (2) e, sotto l'ipotesi ($\diamond$), la relazione di massimo forte ($\star$).

Step 3: definizione della variazione $(x_0^\varepsilon, x^\varepsilon)$ **e costruzione della *needle variation*.** Definiamo la seguente nozione di *variazione* della traiettoria (x_0^*, x^*) in un punto.

Definizione 2.2 (variazione)**.** Sia $\hat t \in (a,b)$ fissato e sia $\bar\varepsilon > 0$. Diciamo che una famiglia di funzioni

$$(x_0^\varepsilon, x^\varepsilon)\colon [a,b] \to \mathbb{R} \times \mathbb{R}^n, \ \varepsilon \in [0, \bar\varepsilon]$$

è una *variazione* della traiettoria (x_0^*, x^*) nel punto $\hat t \in (a,b)$ se si ha che

(i) $(x_0^\varepsilon, x^\varepsilon)\big|_{\varepsilon=0} = (x_0^*, x^*)$;
(ii) per ogni $\varepsilon \in [0, \bar\varepsilon]$ esiste un controllo $u^\varepsilon\colon [a,b] \to U$ tale che la coppia controllo-traiettoria $(u^\varepsilon, x^\varepsilon)$ è ammissibile per il problema (P-nv equiv.), cioè $(x_0^\varepsilon, x^\varepsilon)$ risolve il problema

$$\text{(PC-}\varepsilon\text{)} \qquad \begin{cases} \dot x_0^\varepsilon(t) = l(t, x^\varepsilon(t), u^\varepsilon(t)), \ t \in (a,b) \\ x_0^\varepsilon(a) = 0 \\ \dot x^\varepsilon(t) = f(t, x^\varepsilon(t), u^\varepsilon(t)), \ t \in (a,b) \\ x^\varepsilon(a) = x_a; \end{cases}$$

(iii) per ogni $\varepsilon \in [0, \bar\varepsilon]$ vale $u^\varepsilon(t) = u^*(t)$ per ogni $t \in [\hat t, b]$;
(iv) per ogni $t \in [a,b]$ fissato, il cammino $\varepsilon \mapsto (x_0^\varepsilon(t), x^\varepsilon(t))$ è differenziabile per $\varepsilon = 0$.

Costruiamo ora un esempio concreto di variazione secondo la Definizione 2.2, detta *needle variation* o *variazione di Weierstrass*, che useremo da qui in poi nella dimostrazione. Siano $\hat t \in (a,b)$ e $\hat u \in U$ fissati a piacere. Definiamo il controllo u^ε variato al tempo $\hat t$ con valore $\hat u$ ponendo

$$u^\epsilon(t) = \begin{cases} u^*(t) & t \in [a, \hat t - \varepsilon) \\ \hat u & t \in [\hat t - \varepsilon, \hat t) \\ u^*(t) & t \in [\hat t, b], \end{cases} \tag{2.2}$$

dove $\varepsilon \geq 0$ varia in modo che $\hat t - \varepsilon \in (a, \hat t]$. Definiamo la *needle variation* $(x_0^\varepsilon, x^\varepsilon)$ della traiettoria ottimale (x_0^*, x^*) nel punto $\hat t$ con scelta $\hat u$ come la famiglia di soluzioni del problema (PC-ε), dove il controllo u^ε è definito sopra, al variare di ε (le ipotesi (ii), (iii) e (iv) assicurano, similmente a quanto visto in precedenza, che, per ogni scelta opportuna di ε, la soluzione esista, sia unica e sia assolutamente continua). Dobbiamo verificare che la *needle variation* $(x_0^\varepsilon, x^\varepsilon)$ rispetta la Definizione 2.2: ci basta verificare che, per ogni $t \in [a,b]$ fissato, il cammino $\varepsilon \mapsto (x_0^\varepsilon(t), x^\varepsilon(t))$ è differenziabile per $\varepsilon = 0$.

Step 4: il vettore tangente di $\varepsilon \mapsto (x_0^\varepsilon(t), x^\varepsilon(t))$. Dobbiamo calcolare il vettore

$$\left.\frac{d(x_0^\varepsilon(t), x^\varepsilon(t))}{d\varepsilon}\right|_{\varepsilon=0}$$

al variare di $t \in [a,b]$. Vale il seguente risultato.

Proposizione 2.3 (vettore tangente, versione debole)**.** *Assumiamo le ipotesi* (i)*,* (ii)*,* (iii) *e* (iv)*. Sia $\hat{u} \in U$. Allora esiste un insieme misurabile $E_{\hat{u}} \subset [a,b]$, di misura piena, tale che, per ogni $t \in E_{\hat{u}}$, la* needle variation *$(x_0^\varepsilon, x^\varepsilon)$ della traiettoria ottimale (x_0^*, x^*) nel punto t con scelta $\hat{u}$ verifica*

$$\left.\frac{d(x_0^\varepsilon(t), x^\varepsilon(t))}{d\varepsilon}\right|_{\varepsilon=0} = \begin{pmatrix} l(t, x^*(t), \hat{u}) \\ f(t, x^*(t), \hat{u}) \end{pmatrix} - \begin{pmatrix} l(t, x^*(t), u^*(t)) \\ f(t, x^*(t), u^*(t)) \end{pmatrix} \tag{2.3}$$

Dimostrazione della Proposizione 2.3. Possiamo ragionare per componenti, ad esempio la seconda. Fissiamo $t \in [a,b]$ e sia u^ε il controllo variato al tempo t con valore $\hat{u}$ come nella (2.2). Allora

$$x^\varepsilon(t) = x^\varepsilon(t-\varepsilon) + \int_{t-\varepsilon}^t f(s, x^\varepsilon(s), u^\varepsilon(s))\ ds = x^*(t-\varepsilon) + \int_{t-\varepsilon}^t f(s, x^\varepsilon(s), \hat{u})\ ds$$

e, per definizione,

$$x^*(t) = x^*(t-\varepsilon) + \int_{t-\varepsilon}^t f(s, x^*(s), u^*(s))\ ds.$$

Allora possiamo scrivere

$$\begin{aligned} \left.\frac{d}{d\varepsilon} x^\varepsilon(t)\right|_{\varepsilon=0} &= \lim_{\varepsilon\to 0} \frac{x^\varepsilon(t) - x^*(t)}{\varepsilon} = \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^t f(s, x^\varepsilon(s), \hat{u}) - f(s, x^*(s), u^*(s))\ ds = \\ &= \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^t f(s, x^\varepsilon(s), \hat{u}) - f(s, x^*(s), \hat{u}))\ ds + && =: I_1 \\ &\quad + \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^t f(s, x^*(s), \hat{u}) - f(s, x^*(s), u^*(s))\ ds. && =: I_2 \end{aligned}$$

Per I_1 osserviamo che, per l'ipotesi (iv),

$$\begin{aligned} \left| \int_{t-\varepsilon}^t f(s, x^\varepsilon(s), \hat{u}) - f(s, x^*(s), \hat{u}))\ ds \right| &\le \int_{t-\varepsilon}^t |f(s, x^\varepsilon(s), \hat{u}) - f(s, x^*(s), \hat{u}))|\ ds \le \\ &\le \int_{t-\varepsilon}^t k(s)|x^\varepsilon(s) - x^*(s)|\ ds \le \sup_{a\le s\le b} |x^\varepsilon(s) - x^*(s)| \int_{t-\varepsilon}^t k(s)\ ds, \end{aligned}$$

per una opportuna funzione $k \in L^1([a,b])$ indipendente[4] da ε. Vogliamo stimare la funzione $s \mapsto |x^\varepsilon(s) - x^*(s)|$ usando la disuguaglianza di Grönwall, Lemma 1.2. Osserviamo che, per q.o.

[4]È sufficiente infatti scegliere, in riferimento all'ipotesi (iv), un compatto Q tale che $x^\varepsilon([a,b]) \subset Q$ per ogni $\varepsilon \in [0, \bar{\varepsilon}]$; in altre parole, Q contiene tutte le curve immagini delle traiettorie variate x^ε.

$s \in [t-\varepsilon, t]$, si ha

$$\begin{aligned}\frac{d}{ds}|x^\varepsilon(s) - x^*(s)| &\le |\dot{x}^\varepsilon(s) - \dot{x}^*(s)| = |f(s, x^\varepsilon(s), \hat{u}) - f(s, x^*(s), u^*(s))| \le \\ &\le |f(s, x^\varepsilon(s), \hat{u}) - f(s, x^*(s), \hat{u}))| + |f(s, x^*(s), \hat{u}) - f(s, x^*(s), u^*(s))| \le \\ &\le k(s)|x^\varepsilon(s) - x^*(s)| + h(s)\end{aligned}$$

dove abbiamo posto

$$h(s) = |f(s, x^*(s), \hat{u}) - f(s, x^*(s), u^*(s))|.$$

Osserviamo che, per l'ipotesi (iv), $h \in L^1([a,b])$. Per il Lemma 1.2 di Grönwall, allora, troviamo che

$$\begin{aligned}|x^\varepsilon(s) - x^*(s)| &\le |x^\varepsilon(t-\varepsilon) - x^*(t-\varepsilon)| e^{\int_{t-\varepsilon}^s k(\sigma)d\sigma} + \int_{t-\varepsilon}^s h(\sigma) e^{\int_s^t k(\tau)d\tau} d\sigma \le \\ &\le \int_{t-\varepsilon}^t h(\sigma) e^{\int_a^b k(\tau)d\tau} d\sigma = e^{\|k\|_{L^1([a,b])}} \int_{t-\varepsilon}^t h(\sigma) d\sigma,\end{aligned}$$

perché $x^\varepsilon(t-\varepsilon) = x^*(t-\varepsilon)$ per la (2.2). Affermiamo ora che, pur di scegliere $\bar{\varepsilon} > 0$ sufficientemente piccolo, esiste una costante $M > 0$ indipendente da ε tale che

$$\sup_{a \le s \le b} |x^\varepsilon(s) - x^*(s)| \le M\varepsilon \quad \forall \varepsilon \in [0, \bar{\varepsilon}].$$

Infatti abbiamo mostrato che

$$\sup_{a \le s \le b} |x^\varepsilon(s) - x^*(s)| \le e^{\|k\|_{L^1([a,b])}} \int_{t-\varepsilon}^t h(\sigma) d\sigma.$$

Per il Teorema 1.3 di differenziazione di Lebesgue applicato alla funzione $s \mapsto h(s)$, esiste un insieme $A_1^{\hat{u}}$ (che dipende dalla scelta di $\hat{u} \in U$) di misura piena in $[a,b]$ tale che, se si sceglie $t \in A_1^{\hat{u}}$, si ha

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^t h(\sigma) d\sigma = h(t) \ge 0.$$

Allora, scelto $C = h(t) + 1$, esiste $\bar{\varepsilon} > 0$ tale che

$$\frac{1}{\varepsilon} \int_{t-\varepsilon}^t h(\sigma) d\sigma \le C \quad \forall \varepsilon \in [0, \bar{\varepsilon}].$$

Quindi

$$\sup_{a \le s \le b} |x^\varepsilon(s) - x^*(s)| \le e^{\|k\|_{L^1([a,b])}} \cdot C\varepsilon = M\varepsilon$$

dove $M = Ce^{\|k\|_{L^1([a,b])}}$. Pertanto possiamo stimare

$$|I_1| \le \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \sup_{a \le s \le b} |x^\varepsilon(s) - x^*(s)| \int_{t-\varepsilon}^t k(s)\ ds \le \lim_{\varepsilon \to 0} M \int_{t-\varepsilon}^t k(s)\ ds = 0.$$

Per I_2, invece, scriviamo

$$\begin{aligned}I_2 = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} &\int_{t-\varepsilon}^t f(s, x^*(s), \hat{u}) - f(t, x^*(t), \hat{u})\ ds + && =: I_{21} \\ &+ f(t, x^*(t), \hat{u}) - f(t, x^*(t), u^*(t)) + \\ &+ \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^t f(t, x^*(t), u^*(t)) - f(s, x^*(s), u^*(s))\ ds. && =: I_{22}\end{aligned}$$

Per il Teorema 1.3 di differenziazione di Lebesgue applicato alla funzione $s \mapsto f(s, x^*(s), u^*(s))$, esiste un insieme A^* (che non dipende dalla scelta di $\hat{u} \in U$) di misura piena in $[a, b]$ tale che, se si sceglie $t \in A^*$,

$$|I_{22}| \leq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^{t} |f(t, x^*(t), u^*(t)) - f(s, x^*(s), u^*(s))| \, ds = 0.$$

Per il Teorema 1.3 di differenziazione di Lebesgue applicato alla funzione $s \mapsto f(s, x^*(s), \hat{u})$, esiste un insieme $A_2^{\hat{u}}$ (che dipende dalla scelta di $\hat{u} \in U$) di misura piena in $[a, b]$ tale che, se si sceglie $t \in A_2^{\hat{u}}$,

$$|I_{21}| \leq \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_{t-\varepsilon}^{t} |f(s, x^*(s), \hat{u}) - f(t, x^*(t), \hat{u})| \, ds = 0.$$

In conclusione, abbiamo mostrato che, per ogni $t \in A^* \cap A_1^{\hat{u}} \cap A_2^{\hat{u}}$,

$$\left. \frac{d}{d\varepsilon} x^\varepsilon(t) \right|_{\varepsilon=0} = f(t, x^*(t), \hat{u}) - f(t, x^*(t), u^*(t)).$$

Mutatis mutandis, un argomento del tutto analogo per la prima componente mostra che esistono tre insiemi B^*, $B_1^{\hat{u}}$ e $B_2^{\hat{u}}$ di misura piena in $[a, b]$ tali che, per ogni $t \in B^* \cap B_1^{\hat{u}} \cap B_2^{\hat{u}}$,

$$\left. \frac{d}{d\varepsilon} x_0^\varepsilon(t) \right|_{\varepsilon=0} = l(t, x^*(t), \hat{u}) - l(t, x^*(t), u^*(t)).$$

Possiamo allora scegliere

$$E_{\hat{u}} = A^* \cap A_1^{\hat{u}} \cap A_2^{\hat{u}} \cap B^* \cap B_1^{\hat{u}} \cap B_2^{\hat{u}}.$$

Questo termina la dimostrazione della Proposizione 2.3. ∎

Vogliamo ora sfruttare le informazioni date dalla Proposizione 2.3 per dimostrare la relazione di massimo debole (2).

Step 5: minimalità della traiettoria ottima. La coppia controllo-traiettoria (u^*, x^*) è ottimale per il problema (P-nv), quindi, equivalentemente, la traiettoria (x_0^*, x^*) è ottimale per il problema (P-nv equiv.).

Fissiamo, da qui in poi, $\hat{u} \in U$, e sia $E_{\hat{u}} \subset [a, b]$ l'insieme di misura piena dato dal Proposizione 2.3. Scegliamo un tempo $\hat{t} \in E_{\hat{u}}$ e consideriamo la *needle variation* $(x_0^\varepsilon, x^\varepsilon)$ della traiettoria ottimale (x_0^*, x^*) nel punto $\hat{t}$ con scelta $\hat{u}$, dove $\varepsilon \in [0, \bar{\varepsilon}]$ per un $\bar{\varepsilon} > 0$ opportuno.

Per la minimalità di (x_0^*, x^*) nel problema (P-nv equiv.), dobbiamo avere che

$$x_0^*(b) + g(x^*(b)) \leq x_0^\varepsilon(b) + g(x^\varepsilon(b))$$

per ogni $\varepsilon \in [0, \bar{\varepsilon}]$. Siccome la funzione g è di classe $\mathcal{C}^1$ per l'ipotesi (i) e $(x_0^\varepsilon, x^\varepsilon)\big|_{\varepsilon=0} = (x_0^*, x^*)$ in accordo con la Definizione 2.2, questo implica che[5]

$$0 \leq \left. \frac{d}{d\varepsilon} \Big(x_0^\varepsilon(b) + g(x^\varepsilon(b)) \Big) \right|_{\varepsilon=0} = \begin{pmatrix} 1 & \nabla g(x^*(b)) \end{pmatrix} \left. \frac{d(x_0^\varepsilon(b), x^\varepsilon(b))}{d\varepsilon} \right|_{\varepsilon=0} \tag{2.4}$$

per la regola di derivata di funzione composta. Dobbiamo calcolare il vettore

$$\left. \frac{d(x_0^\varepsilon(b), x^\varepsilon(b))}{d\varepsilon} \right|_{\varepsilon=0},$$

che è il vettore tangente alla curva $\varepsilon \mapsto (x_0^\varepsilon(b), x^\varepsilon(b))$.

[5] Se $\phi \in \mathcal{C}^1([0, \tau]; \mathbb{R})$ ha minimo (locale) in 0, allora $\phi'(0+) \geq 0$ (derivata destra).

Step 6: trasporto del vettore tangente da $\hat{t}$ a b lungo la traiettoria ottima. Utilizzando le proprietà del flusso di una equazione differenziale, vogliamo calcolare il vettore $\frac{d(x_0^\varepsilon(b),x^\varepsilon(b))}{d\varepsilon}\Big|_{\varepsilon=0}$ a partire dal vettore $\frac{d(x_0^\varepsilon(\hat{t}),x^\varepsilon(\hat{t}))}{d\varepsilon}\Big|_{\varepsilon=0}$, che conosciamo grazie alla Proposizione 2.3. Precisamente, vale il seguente risultato.

Proposizione 2.4 (trasporto del vettore tangente)**.** *Assumiamo le ipotesi* (i)*,* (ii)*,* (iii) *e* (iv)*. Siano $\hat{u} \in U$, $E_{\hat{u}} \subset [a,b]$, $\hat{t} \in E_{\hat{u}}$, $(x_0^\varepsilon, x^\varepsilon)$ come li abbiamo fissati nello Step 5. Sia*

$$\begin{pmatrix} v_0 \\ v \end{pmatrix} : [\hat{t}, b] \to \mathbb{R} \times \mathbb{R}^n$$

la soluzione del problema di Cauchy (lineare, ai valori iniziali)

$$\text{(PC-}v\text{)} \qquad \begin{cases} \begin{pmatrix} \dot{v}_0(t) \\ \dot{v}(t) \end{pmatrix} = \left[\dfrac{\partial l}{\partial x}(t, x^*(t), u^*(t)) \quad \dfrac{\partial f}{\partial x}(t, x^*(t), u^*(t)) \right] \begin{pmatrix} v_0(t) \\ v(t) \end{pmatrix}, \quad t \in (\hat{t}, b) \\ \begin{pmatrix} v_0(\hat{t}) \\ v(\hat{t}) \end{pmatrix} = \dfrac{d(x_0^\varepsilon(\hat{t}), x^\varepsilon(\hat{t}))}{d\varepsilon}\Big|_{\varepsilon=0}. \end{cases}$$

Allora

$$\begin{pmatrix} v_0(b) \\ v(b) \end{pmatrix} = \frac{d(x_0^\varepsilon(b), x^\varepsilon(b))}{d\varepsilon}\Big|_{\varepsilon=0}.$$

Dimostrazione della Proposizione 2.4. [6] Dalla Definizione 2.2, abbiamo infatti che

$$\begin{cases} \dot{x}^\varepsilon(t) = f(t, x^\varepsilon(t), u^\varepsilon(t)), \ t \in (a,b) \\ x^\varepsilon(a) = x_a, \end{cases}$$

e quindi, in particolare,

$$x^\varepsilon(t) = x^\varepsilon(\hat{t}) + \int_{\hat{t}}^t f(s, x^\varepsilon(s), u^\varepsilon(s)) \, ds \quad \forall t \in [\hat{t}, b].$$

Siccome, dalla (2.2), $u^\epsilon(t) = u^*(t)$ quando $t \in [\hat{t}, b]$, l'equazione precedente si riscrive semplicemente come

$$x^\varepsilon(t) = x^\varepsilon(\hat{t}) + \int_{\hat{t}}^t f(s, x^\varepsilon(s), u^*(s)) \, ds \quad \forall t \in [\hat{t}, b].$$

Derivando[7] in ε e calcolando poi in $\varepsilon = 0$ si trova che

$$\frac{d}{d\varepsilon}x^\varepsilon(t)\Big|_{\varepsilon=0} = \frac{d}{d\varepsilon}x^\varepsilon(\hat{t})\Big|_{\varepsilon=0} + \int_{\hat{t}}^t \frac{\partial f}{\partial x}(s, x^*(s), u^*(s)) \frac{d}{d\varepsilon}x^\varepsilon(s)\Big|_{\varepsilon=0} ds \quad \forall t \in [\hat{t}, b].$$

Definiamo $v(t) = \frac{d}{d\varepsilon}x^\varepsilon(t)\big|_{\varepsilon=0}$. Derivando in t, l'equazione precedente mostra che v risolve il problema di Cauchy

$$\begin{cases} \dot{v}(t) = \dfrac{\partial f}{\partial x}(t, x^*(t), u^*(t)) v(t), \quad t \in (\hat{t}, b) \\ v(\hat{t}) = \dfrac{d}{d\varepsilon}x^\varepsilon(\hat{t})\Big|_{\varepsilon=0}. \end{cases}$$

[6] Questi conti sono stati lasciati come esercizio durante il corso.

[7] La derivata in ε passa sotto il segno di integrale per il Teorema di derivazione di integrali dipendenti da parametro, perché f è di classe $\mathcal{C}^1$ in x per l'ipotesi (iii)

Mutatis mutandis, un argomento del tutto analogo mostra che $v_0(t) = \frac{d}{d\varepsilon} x_0^\varepsilon(t)\big|_{\varepsilon=0}$ risolve il problema di Cauchy

$$\begin{cases} \dot{v}_0(t) = \dfrac{\partial l}{\partial x}(t, x^*(t), u^*(t))v_0(t), \quad t \in (\hat{t}, b) \\ v_0(\hat{t}) = \dfrac{d}{d\varepsilon} x_0^\varepsilon(\hat{t})\Big|_{\varepsilon=0}. \end{cases}$$

Questo termina la dimostrazione della Proposizione 2.4. ■

Step 7: dualità dei problemi (PC-v) **e** (PC-λ). Osserviamo ora che il problema (PC-λ) è il problema di Cauchy duale del problema (PC-v). Applichiamo allora il Lemma 1.1 ai problemi (PC-v) e (PC-λ): otteniamo che il prodotto delle soluzioni diretta e duale deve essere costante, cioè

$$\begin{pmatrix} -1 & \lambda(t) \end{pmatrix} \begin{pmatrix} v_0(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} -1 & \lambda(b) \end{pmatrix} \begin{pmatrix} v_0(b) \\ v(b) \end{pmatrix} \quad \forall t \in [\hat{t}, b].$$

Sappiamo che, dal problema (PC-λ) e dalla Proposizione 2.4,

$$\lambda(b) = -\nabla g(x^*(b)) \quad \text{e} \quad \begin{pmatrix} v_0(b) \\ v(b) \end{pmatrix} = \frac{d(x_0^\varepsilon(b), x^\varepsilon(b))}{d\varepsilon}\bigg|_{\varepsilon=0},$$

quindi

$$\begin{pmatrix} -1 & \lambda(t) \end{pmatrix} \begin{pmatrix} v_0(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} -1 & -\nabla g(x^*(b)) \end{pmatrix} \frac{d(x_0^\varepsilon(b), x^\varepsilon(b))}{d\varepsilon}\bigg|_{\varepsilon=0}$$

per ogni $t \in [\hat{t}, b]$. In particolare, per $t = \hat{t}$ troviamo che

$$\begin{pmatrix} -1 & \lambda(\hat{t}) \end{pmatrix} \frac{d(x_0^\varepsilon(\hat{t}), x^\varepsilon(\hat{t}))}{d\varepsilon}\bigg|_{\varepsilon=0} = \begin{pmatrix} -1 & -\nabla g(x^*(b)) \end{pmatrix} \frac{d(x_0^\varepsilon(b), x^\varepsilon(b))}{d\varepsilon}\bigg|_{\varepsilon=0}. \tag{2.5}$$

Step 8: relazione di massimo debole. Dalle equazioni (2.5) e (2.4) otteniamo che

$$\begin{pmatrix} -1 & \lambda(\hat{t}) \end{pmatrix} \frac{d(x_0^\varepsilon(\hat{t}), x^\varepsilon(\hat{t}))}{d\varepsilon}\bigg|_{\varepsilon=0} \leq 0,$$

ma questo, per l'equazione (2.3) della Proposizione 2.3, significa che

$$\begin{pmatrix} -1 & \lambda(\hat{t}) \end{pmatrix} \begin{pmatrix} l(\hat{t}, x^*(\hat{t}), u^*(\hat{t})) \\ f(\hat{t}, x^*(\hat{t}), u^*(\hat{t})) \end{pmatrix} \geq \begin{pmatrix} -1 & \lambda(\hat{t}) \end{pmatrix} \begin{pmatrix} l(\hat{t}, x^*(\hat{t}), \hat{u}) \\ f(\hat{t}, x^*(\hat{t}), \hat{u}) \end{pmatrix},$$

ovvero, svolgendo i prodotti,

$$-l(\hat{t}, x^*(\hat{t}), u^*(\hat{t})) + \lambda(\hat{t}) f(\hat{t}, x^*(\hat{t}), u^*(\hat{t})) \geq -l(\hat{t}, x^*(\hat{t}), \hat{u}) + \lambda(\hat{t}) f(\hat{t}, x^*(\hat{t}), \hat{u}).$$

Per la definizione della funzione H data nella (2.1), possiamo riscrivere la disuguaglianza precedente come

$$H(\lambda(\hat{t}), \hat{t}, x^*(\hat{t}), u^*(\hat{t})) \geq H(\lambda(\hat{t}), \hat{t}, x^*(\hat{t}), \hat{u})$$

con $\hat{u} \in U$ fissato a piacere nello Step 5, $E_{\hat{u}} \subset [a, b]$ di misura piena dato dalla Proposizione 2.3, e $\hat{t} \in E_{\hat{u}}$ fissato a piacere nello Step 5. Abbiamo dimostrato che la funzione λ definita nello Step 2 verifica la relazione di massimo debole (2).

Step 9: relazione di massimo forte. Da qui in poi assumiamo anche l'ipotesi ($\diamond$). Per concludere la dimostrazione controlliamo che la funzione λ definita nello Step 2 verifica la relazione di massimo forte ($\star$).

Siccome $U \subset \mathbb{R}^m$, esiste un sottoinsieme $\mathcal{U} \subset U$ numerabile e denso in U. Definiamo allora

$$\mathcal{E} = \bigcap_{u \in \mathcal{U}} E_u,$$

dove E_u è l'insieme di misura piena relativo a $u \in \mathcal{U}$ dato dalla Proposizione 2.3. Osserviamo che $\mathcal{E} \subset [a,b]$ è un insieme misurabile di misura piena, perché è intersezione numerabile di insiemi di misura piena in $[a,b]$.

Dalla relazione di massimo debole (2) segue che, per ogni $u \in \mathcal{U}$, vale

$$H(\lambda(t), t, x^*(t), u^*(t)) \geq H(\lambda(t), t, x^*(t), u)$$

per ogni $t \in \mathcal{E}$. Questo implica che

$$H(\lambda(t), t, x^*(t), u^*(t)) \geq \sup_{u \in \mathcal{U}} H(\lambda(t), t, x^*(t), u) \tag{2.6}$$

per ogni $t \in \mathcal{E}$.

Vogliamo dimostrare che la (2.6) implica la relazione di massimo forte ($\star$), con $\mathcal{E}$ come insieme di misura piena in $[a,b]$. Infatti,[8] per assurdo, supponiamo che esista $\bar{u} \in U$ tale che

$$H(\lambda(\bar{t}), \bar{t}, x^*(\bar{t}), u^*(\bar{t})) < H(\lambda(\bar{t}), \bar{t}, x^*(\bar{t}), \bar{u}) \tag{2.7}$$

per un certo $\bar{t} \in \mathcal{E}$. Siccome $\mathcal{U}$ è denso in U, esiste una successione $(u_k) \subset \mathcal{U}$ tale che $u_k \to \bar{u}$ per $k \to \infty$, ed inoltre, per la (2.6),

$$H(\lambda(\bar{t}), \bar{t}, x^*(\bar{t}), u^*(\bar{t})) \geq H(\lambda(\bar{t}), \bar{t}, x^*(\bar{t}), u_k) \quad \forall\, k \in \mathbb{N}. \tag{2.8}$$

Per l'ipotesi ($\diamond$), la funzione $(x,u) \mapsto (l,f)(t,x,u)$ è continua su $\mathbb{R}^n \times U$, pertanto la funzione $(x,u) \mapsto H(\lambda, t, x, u)$ è continua su $\mathbb{R}^n \times U$. Allora, passando al limite nella (2.8) per $k \to +\infty$, troviamo

$$H(\lambda(\bar{t}), \bar{t}, x^*(\bar{t}), u^*(\bar{t})) \geq H(\lambda(\bar{t}), \bar{t}, x^*(\bar{t}), \bar{u}),$$

che contraddice la (2.7). La dimostrazione è finalmente completa. $\square$

3. Il caso vincolato

In questa sezione studiamo il problema di minimizzazione vincolato

$$\begin{cases} \inf\limits_{(x,u)} \int_a^b l(t,x(t),u(t))\, dt + g(x(b)) \\ \dot{x}(t) = f(t,x(t),u(t)),\ t \in (a,b) \\ x(a) = x_a \\ x(b) \in S \end{cases} \tag{P-v}$$

dove $S \subset \mathbb{R}^n$ è detto *insieme bersaglio* o *target finale*.

3.1. **Notazioni per il caso vincolato.** Fissiamo le notazioni. Diciamo che la coppia controllo-traiettoria (u,x) è *ammissibile* per il problema (P-v) se risolve il problema di Cauchy

$$\begin{cases} \dot{x}(t) = f(t,x(t),u(t)),\ t \in (a,b) \\ x(a) = x_a \end{cases}$$

e la traiettoria x soddisfa condizione di vincolo finale $x(b) \in S$. La coppia controllo-traiettoria (u^*, x^*) è *ottimale* se è ammissibile ed è un minimizzante per il problema (P-nv), ovvero

$$\int_a^b l(t,x^*(t),u^*(t))\, dt + g(x^*(b)) = \inf_{(u,x)} \int_a^b l(t,x(t),u(t))\, dt + g(x(b))$$

dove l'estremo inferiore è fatto su tutte le coppie (u,x) ammissibili. Poniamo infine

$$\begin{aligned} H(\lambda_0, \lambda, t, x, u) &= \begin{pmatrix} \lambda_0 & \lambda \end{pmatrix} \begin{bmatrix} l(t,x,u) & f(t,x,u) \end{bmatrix} = \\ &= \lambda_0 l(t,x,u) + \lambda f(t,x,u) \end{aligned} \tag{3.1}$$

[8]Quanto segue è l'applicazione alla funzione H, continua nella variabile u per la ($\diamond$), del seguente fatto: se $\phi \in C^0(A;\mathbb{R})$ e $\mathcal{A} \subset A$ è denso in A, allora $\sup\limits_{x \in \mathcal{A}} \phi(x) = \sup\limits_{x \in A} \phi(x)$.

la *Hamiltoniana (non massimizzata)* del problema vincolato (P-v).

3.2. **Teorema del Massimo nel caso vincolato.** Vale il seguente risultato.

Teorema 3.1 (Principio del Massimo di Pontryagin, vincolato)**.** *Sia dato il problema di minimizzazione non vincolato* (P-v) *e sia* (u^*, x^*) *una coppia controllo-traiettoria ottimale. Facciamo le seguenti ipotesi.*

(i) La funzione $g\colon \mathbb{R}^n \to \mathbb{R}$ *è di classe* C^1*.*

(ii) Per ogni $(x,u) \in \mathbb{R}^n \times U$ *fissato, la funzione*

$$t \mapsto (l,f)(t,x,u)$$

è misurabile su $[a,b]$*.*

(iii) Per ogni $(t,u) \in [a,b] \times \mathbb{R}^n$ *fissato, la funzione*

$$x \mapsto (l,f)(t,x,u)$$

è di classe C^1 *su* $\mathbb{R}^n$*.*

(iv) Per ogni insieme compatto $Q \subset \mathbb{R}^n$ *fissato, esiste una funzione* $k \in L^1([a,b])$ *tale che, per ogni* $(x,u) \in Q \times U$*, si ha*

$$\max\left\{ |(l,f)(t,x,u)|, \left|\frac{\partial(l,f)}{\partial x}(t,x,u)\right| \right\} \le k(t) \quad \forall t \in [a,b].$$

Allora esistono un numero reale $\lambda_0 \le 0$ *e una funzione* $\lambda \in AC([a,b];\mathbb{R}^n)$ *tali che*

(1) (equazione aggiunta) *per q.o.* $t \in [a,b]$ *si ha*

$$\begin{aligned}\dot\lambda(t) &= -\begin{pmatrix}\lambda_0 & \lambda(t)\end{pmatrix}\begin{bmatrix}\dfrac{\partial l}{\partial x}(t,x^*(t),u^*(t)) & \dfrac{\partial f}{\partial x}(t,x^*(t),u^*(t))\end{bmatrix} = \\ &= -\left(\lambda_0\frac{\partial l}{\partial x}(t,x^*(t),u^*(t)) + \lambda(t)\frac{\partial f}{\partial x}(t,x^*(t),u^*(t))\right);\end{aligned}$$

(2) (relazione di massimo **debole**) *per ogni* $u \in U$ *esiste un insieme misurabile* $E_u \subset [a,b]$ *di misura piena,* $|[a,b] \setminus E_u| = 0$*, tale che*

$$H(\lambda_0,\lambda(t),t,x^*(t),u^*(t)) \ge H(\lambda_0,\lambda(t),t,x^*(t),u)$$

per ogni $t \in E_u$*;*

(3) (non banalità)

$$\begin{pmatrix}\lambda_0 & \lambda(t)\end{pmatrix} \neq \begin{pmatrix}0 & 0\end{pmatrix}$$

per almeno un $t \in [a,b]$*;*

(4) (condizione di trasversalità) *si ha*

$$\lambda(b) - \lambda_0 \nabla g(x^*(b)) \in -C^\perp$$

dove $C \subset \mathbb{R}^n$ *è un cono approssimante (nel senso di Boltyansky) l'insieme bersaglio* S *in* $x^*(b)$*.*

Se, in aggiunta alle ipotesi precedenti, supponiamo che l'insieme dei controlli U *sia un sottoinsieme di* $\mathbb{R}^m$ *e che, per ogni* $t \in [a,b]$ *fissato, si abbia*

$$(x,u) \mapsto (l,f)(t,x,u) \in C(\mathbb{R}^n \times U), \tag{$\diamond\diamond$}$$

allora la (2) può essere migliorata come segue:

(2⋆) (relazione di massimo **forte**) *per q.o.* $t \in [a,b]$ *si ha*

$$H(\lambda_0,\lambda(t),t,x^*(t),u^*(t)) = \max_{u\in U} H(\lambda_0,\lambda(t),t,x^*(t),u). \tag{$\star\star$}$$