Ein kurzer Beweis des Pontryaginschen Maximumprinzips ohne Zustandsbeschränkungen und rechte Randbedingung

Als Vorbemerkung betrachten wir das abstrakte Optmierungsproblem

$$\begin{aligned}&\min f(y,u)\\&\text{unter } c(y,u)=0 \quad u\in\Omega\end{aligned}$$

wobei c_y als invertierbar vorausgesetzt wird, wodurch die Gleichungsnebenbedingung zu jedem $u \in \Omega$ implizit ein eindeutiges $y = y(u)$ definiert. Wir bezeichnen mit $(.,.)$ das Skalarprodukt in passenden Funktionenräumen.

Lemma Es gilt für die Lösung $(\hat{y},\hat{u})$ des obigen Optimierungsproblems

$$f(\hat{y},\hat{u}) + \big(\lambda, c(\hat{y},\hat{u})\big) \le f(\hat{y},u) + \big(\lambda, c(\hat{y},u)\big)\ ,\forall u\in\Omega, \|u-\hat{u}\|<\varepsilon$$

für ein $\varepsilon > 0$ und wenn λ das adj. Problem löst

$$C_y(\hat{y},\hat{u})^{ad}\lambda = -\nabla_y f(\hat{y},\hat{u})$$

Beweis: Betrachte $y = y(u)$ via $c(y,u) = 0$

Da u optimal, gilt offenbar

$$f\big(y(u),u\big) - f\big(y(\hat{u}),\hat{u}\big) \ge 0\ ,\forall u\in\Omega$$

Für $\Delta u := u - \hat{u}$ hinreichend klein führen wir eine Taylorentwicklung durch

$$f\big(y(u),u\big) - f\big(y(\hat{u}),\hat{u}\big) = \underbrace{\frac{\partial f}{\partial y}\big(y(\hat{y}),\hat{u}\big)\frac{\partial y}{\partial u}(\hat{u})\Delta u}_{=\left(-\frac{\partial c^{ad}}{\partial u}\left(\frac{\partial c^{ad}}{\partial y}\right)^{-1}\nabla_y f,\Delta u\right)} + \frac{\partial f}{\partial u}\big(y(\hat{u}),\hat{u}\big)\Delta u + \mathcal{O}(\|\Delta u\|^2)$$

$$= \left(\lambda, \frac{\partial c}{\partial u}\big(y(\hat{u}),\hat{u}\big)\Delta u\right) + \frac{\partial f}{\partial u}\big(y(\hat{u}),\hat{u}\big)\Delta u + \mathcal{O}(\|\Delta u\|^2)$$

$$= f(\hat{y},u) + \big(\lambda, c(\hat{y},u)\big) - \Big(f(\hat{y},\hat{u}) + \big(\lambda, c(\hat{y},\hat{u})\big) + \mathcal{O}(\|\Delta u\|^2) \ge 0 \text{ für } \|\Delta u\| \to 0 \qquad \square$$

Damit formulieren wir nun das Maximumprinzip für das Optimalsteuerungsproblem

$$\begin{aligned}&\min \int_{t_0}^{t_f} L\big(t,y(t),u(t)\big)dt\\&\text{unter } \dot{y}(t) = f\big(t,y(t),u(t)\big)\\&\qquad y(t_0) = y_0\end{aligned}$$

wobei

$$u\in\Omega := \{u:[t_0,t_f]\to R^{n_u}\mid u \text{ stückweise stetig}, u(t)\in\Omega(t),\ \forall t\}$$

Als Funktionenraum für y wählen wir ebenfalls:

$$y \in Y := \{y : [t_0, t_f] \to R^n \mid y \text{ stückweise differenzierbar}\}$$

Als Skalarprodukt verwenden wir das $L_2([t_0, t_f])$-Skalarprodukt und ebenso die $L_2([t_0, t_f])$-Norm.

Wir bestimmen die Lagrange-Fkt.:

$$\mathcal{L}(y, u, \lambda) = \int_{t_0}^{t_f} L\big(t, y(t), u(t)\big) + \lambda(t)^\top \left(\dot{y}(t) - f\big(t, y(t), u(t)\big)\right) dt + \lambda_0^\top \left(y(t) - y_0\right)$$

Wir verwenden partielle Integration, um die zeitliche Ableitung von y loszuwerden:

$$\int_{t_0}^{t} \lambda(t)^\top \Big(\dot{y} - f\big(t, y(t), u(t)\big)\Big) dt \stackrel{\text{part.Int.}}{=} -\int_{t_0}^{t_f} \dot{\lambda}(t)^\top y(t) dt + \big[\lambda(t)^\top y(t)\big]_{t_0}^{t_f} - \int_{t_0}^{t_f} \lambda(t)^\top f(t, y(t), u(t)) dt$$

Durch Einsetzen erhalten wir den Ausdruck

$$\begin{aligned} \mathcal{L}(y, u, \lambda) \quad = \quad & \int_{t_0}^{t_f} L\big(t, y(t), u(t)\big) dt - \int_{t_0}^{t_f} \dot{\lambda}(t)^\top y(t) dt + \big[\lambda(t)^\top y(t)\big]_{t_0}^{t_f} - \int_{t_0}^{t_f} \lambda(t)^\top f(t, y(t), u(t)) dt \\ & + \lambda_0^\top \big(y(t) - y_0\big) \end{aligned}$$

Die adjungierte Gleichung $\quad \mathcal{L}_y(y, u, \lambda) = 0 \quad$ liefert nun

$$\dot{\lambda}(t) = -f_y^\top \lambda + \nabla_y L \quad , \lambda(t_f) = 0$$

also $\quad \dot{\lambda}(t) = -\nabla_y H\big(t, y(t), u(t), \lambda(t)\big)$

$$\text{mit } H\big(t, y(t), u(t), \lambda(t)\big) := -L\big(t, y(t), u(t)\big) + \lambda(t)^\top f\big(t, y(t), u(t)\big)$$

außerdem wissen wir nach dem Lemma

$$\mathcal{L}(\hat{y}, \hat{u}, \lambda) \le \mathcal{L}(\hat{y}, u, \lambda) \quad \forall u \in \Omega, \|u - \hat{u}\| < \varepsilon$$

Da nur die Terme, die mit $H(.)$ in $\mathcal{L}$ zusammenhängen, von u abhängen, ist dies äquivalent zu

$$\int_{t_0}^{t_f} -H(\hat{y}, \hat{u}, \lambda) dt \le \int_{t_0}^{t_f} -H(\hat{y}, u, \lambda) dt \quad \forall u \in \Omega, \|u - \hat{u}\| < \varepsilon$$

und somit zu

$$\int_{t_0}^{t_f} H(\hat{y}, \hat{u}, \lambda)dt \geq \int_{t_0}^{t_f} H(\hat{y}, u, \lambda)dt \quad \forall u \in \Omega, \|u - \hat{u}\| < \varepsilon \tag{1}$$

Globalisierung:

Die Ungleichung gilt jedoch auch in ganz Ω. Denn, falls es ein $\bar{u} \in \Omega$ gäbe mit

$$\int_{t_0}^{t_f} H(\hat{y}, \hat{u}, \lambda)dt < \int_{t_0}^{t_f} H(\hat{y}, \bar{u}, \lambda)dt$$

so gäbe es eine Menge $M \subset [t_0, t_f]$ vom Maß $\neq 0$ so dass

$$H(\hat{y}(t), \hat{u}(t), \lambda(t)) < H(\hat{y}(t), \bar{u}(t), \lambda(t)) \quad \forall t \in M$$

Wir können nun folgendermaßen eine neue Steuerung definieren

$$\tilde{u}(t) := \begin{cases} \hat{u}(t), \ t \in [t_0, t_f] \backslash M \\ \bar{u}(t), \ t \in M \end{cases}$$

für diese gilt

$$\int_{t_0}^{t_f} H(\hat{y}, \hat{u}, \lambda)dt < \int_{t_0}^{t_f} H(\hat{y}, \tilde{u}, \lambda)dt$$

Nun können wir M klein genug definieren, dass $\|\tilde{u} - \hat{u}\| < \varepsilon$ und erhalten dadurch einen Widerspruch zur Aussage (1)

Punktweise Ungleichung:

Nun können mit fast dem gleichen Argument die Ungleichung auch noch punktweise formulieren: Fast überall in $[t_0, t_f]$ muss gelten

$$\hat{u}(t) = \arg \max_{v \in \Omega(t)} \ H\big(v, \hat{y}(t), \lambda(t)\big),$$

denn sonst gäbe es ein $\bar{u}$ und eine Menge M mit Maß $\neq 0$

so dass $\quad \displaystyle\int_M H(\hat{y}, \hat{u}, \lambda)dt < \int_M H(\hat{y}, \bar{u}, \lambda)dt$

wir definieren wieder

$$\tilde{u}(t) := \begin{cases} \hat{u}(t), \ t \in [t_0, t_f] \backslash M \\ \bar{u}(t), \ t \in M \end{cases}$$

und erhalten eine bessere Steuerung $\tilde{u}$ als $\hat{u}$ $\lightning$

Wir erhalten somit als notwendige Bedingungen den Satz

Satz Die notwendige Bedingungen für die Lösung des Optimalsteuerungsproblems lassen sich mit Hilfe der Hamiltonfunktion

$$H(t, y(t), u(t), \lambda(t)) := -L(t, y(t), u(t)) + \lambda(t)^\top f(t, y(t), u(t))$$

in folgender Form ausdrücken:

(1) adjungierte Differentialgleichung

$$\dot{\lambda}(t) = -\nabla_y H(t, y(t), u(t), \lambda(t))$$

(2) Transversalitätsbedingung (hier als Endbedingung)

$$\lambda(t_f) = 0$$

(3) Maximumprinzip: es gilt fast überall in $[t_0, t_f]$

$$u(t) = \arg \max_{v \in \Omega(t)} H(t, y(t), v, \lambda(t))$$

Beweis: s.o. □