Computational Methods in Optimal Control
Lecture 8. *hp*-Collocation

William W. Hager

July 26, 2018

10,000 Yen Prize Problem (google this)

Let p be a polynomial of degree at most N and let $-1 < \tau_1 < \tau_2 < \ldots < \tau_N < 1$ be the Gauss quadrature points. Suppose that $p(-1) = 0$ and $|p'(\tau_i)| \leq 1$ for all $1 \leq i \leq N$. Show that $|p(\tau_i)| \leq 2$ for all $1 \leq i \leq N$.

This gives $\|\mathbf{D}^{-1}\|_\infty \leq 2$: Let $p \in \mathcal{P}_N$ with $p(-1) = 0$. Let $\mathbf{p}$ be the vector with components $p_i = p(\tau_i)$, and let $\dot{\mathbf{p}}$ be the vector with components $\dot{p}_i = \dot{p}(\tau_i)$. Since $\mathbf{Dp} = \dot{\mathbf{p}}$, we have $\mathbf{p} = \mathbf{D}^{-1}\dot{\mathbf{p}}$. Since

$$\|\mathbf{D}^{-1}\|_\infty = \max\{\|\mathbf{D}^{-1}\mathbf{y}\|_\infty : -\mathbf{1} \leq \mathbf{y} \leq \mathbf{1}\},$$

it follows that $\|\mathbf{D}^{-1}\|_\infty$ is the maximum value $|p(\tau_i)|$ over all polynomials for which $p(-1) = 0$ and $|\dot{p}(\tau_i)| \leq 1$.

NOTE: the polynomial that achieves the maximum value for $|p(\tau_i)|$ is $p(\tau) = 1 + \tau$.

Model Problem for *hp* Pseudospectral Method

$$\begin{aligned} &\text{minimize} && C(\mathbf{x}(1)) \\ &\text{subject to} && \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{u}(t) \in \mathcal{U}, \quad t \in \Omega_0, \\ & && \mathbf{x}(0) = \mathbf{x}_0. \end{aligned}$$

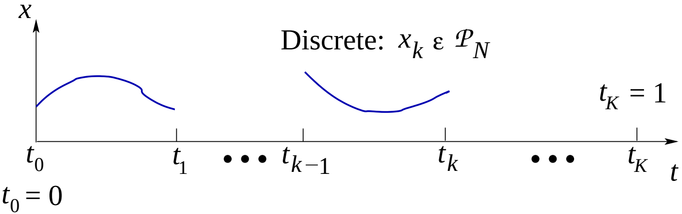
- $\Omega_0 = [0, 1]$, $\mathbf{x}_0$ given, $\mathbf{x}(t) \in \mathbb{R}^n$,
- $\mathcal{U} \subset \mathbb{R}^m$ closed and convex,
- $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $C : \mathbb{R}^n \to \mathbb{R}$

hp Pseudospectral Idea

Partition time domain Ω_0 into K subintervals and use different polynomial to approximate state on each subinterval. Require continuity across subintervals. Change variables so that $\mathbf{x}_k(\tau)$, $\tau \in [-1, 1]$ corresponds to $\mathbf{x}(t)$, $t \in [t_{k-1}, t_k]$.

$$h = 1/2K = \frac{t_k - t_{k-1}}{2} \qquad \tau \in [-1, 1]$$

$$\text{Continuous: } x_k(\tau) = x(t_{k-1/2} + h\tau)$$



Radau Collocation Points

- Radau quadrature points (zeros of $P_{N-1}^{(1,0)}$ plus $\tau_N = 1$):

$$-1 < \tau_1 < \tau_2 < \ldots < \tau_N = +1.$$

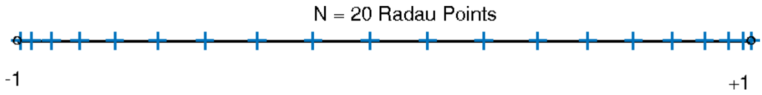
Additional points in analysis: $\tau_0 = -1$

- Quadrature weights:

$$\omega_i = \frac{2(1+\tau_i)}{[(1-\tau_i^2)\dot{P}_{N-1}^{(1,0)}(\tau_i)]^2}, \quad 1 \le i \le N-1, \quad \omega_N = \frac{2}{N^2}.$$

For every $p \in \mathcal{P}_{2N-2}$:

$$\int_{-1}^{1} p(\tau)d\tau = \sum_{i=1}^{N} \omega_i p(\tau_i).$$



Continuous and Discrete Problems on K Mesh Intervals

Continuous:

$$\begin{aligned} &\text{minimize} \quad C(\mathbf{x}_K(1)) \\ &\text{subject to} \quad \dot{\mathbf{x}}_k(\tau) = h\mathbf{f}(\mathbf{x}_k(\tau), \mathbf{u}_k(\tau)), \quad \mathbf{u}_k(\tau) \in \mathcal{U}, \quad \tau \in \Omega, \\ &\qquad\qquad \mathbf{x}_k(-1) = \mathbf{x}_{k-1}(1), \qquad 1 \le k \le K. \end{aligned}$$

Discrete:

$$\begin{aligned} &\text{minimize} \quad C(\mathbf{x}_K(1)) \\ &\text{subject to} \quad \dot{\mathbf{x}}_k(\tau_i) = h\mathbf{f}(\mathbf{x}_k(\tau_i), \mathbf{u}_{ki}), \quad \mathbf{u}_{ki} \in \mathcal{U}, \quad 1 \le i \le N, \\ &\qquad\qquad \mathbf{x}_k(-1) = \mathbf{x}_{k-1}(1), \qquad \mathbf{x}_k \in \mathcal{P}_N^n, \quad 1 \le k \le K. \end{aligned}$$

Easy to Satisfy Continuity with Radau

- Lagrange interpolating polynomials: For $0 \le j \le N$,

$$\Phi_i(\tau) = \prod_{\substack{j=0 \\ j \ne i}}^{N} \frac{\tau - \tau_j}{\tau_i - \tau_j}, \quad \Phi_i(\tau_j) = \begin{cases} 1 \text{ for } j = i \\ 0 \text{ otherwise.} \end{cases}$$

- If $\mathbf{x}_k \in \mathcal{P}_N$, then $\mathbf{x}_k(\tau) = \sum_{j=0}^{N} \mathbf{X}_{kj}(\tau_j)\Phi_j(\tau), \quad \mathbf{X}_{kj} = \mathbf{x}_k(\tau_j)$.
- Continuity $\Leftrightarrow \mathbf{X}_{k-1,N} = \mathbf{X}_{k0}$.
- Differentiation matrix $\mathbf{D} \in \mathbb{R}^{N \times (N+1)}$

$$\dot{\mathbf{x}}(\tau_i) = \sum_{j=0}^{N} \dot{\Phi}_j(\tau_i)\mathbf{x}(\tau_j) = \sum_{j=0}^{N} D_{ij}\mathbf{x}(\tau_j), \quad D_{ij} = \dot{\Phi}_j(\tau_i)$$

Discrete Control Problem

Hence, in terms of the state and control values $\mathbf{X}_{kj}$ and $\mathbf{U}_{kj}$ at the N collocation points on each of the K intervals, the discrete control problem can be formulated as

$$
\begin{array}{lll}
\text{minimize} & C(\mathbf{X}_{KN}) & \\
\text{subject to} & \sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj} = h\mathbf{f}(\mathbf{X}_{ki}, \mathbf{U}_{ki}), \quad \mathbf{U}_{ki} \in \mathcal{U}, & 1 \le i \le N, \\
 & \mathbf{X}_{k0} = \mathbf{X}_{k-1,N}, & 1 \le k \le K,
\end{array}
$$

where $\mathbf{X}_{0N} = \mathbf{X}_0$ is the starting condition.

Lagrangian and KKT Conditions

$$\mathcal{L}(\boldsymbol{\lambda}, \mathbf{X}, \mathbf{U}) = C\left(\mathbf{X}_{KN}\right) +$$
$$\sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \boldsymbol{\lambda}_{ki}, h\mathbf{f}(\mathbf{X}_{ki}, \mathbf{U}_{ki}) - \sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj} \right\rangle +$$
$$\sum_{k=1}^{K} \left\langle \boldsymbol{\lambda}_{k0}, (\mathbf{X}_{k-1,N} - \mathbf{X}_{k0}) \right\rangle .$$

Differentiating with respect to each of the variables:

$$\mathbf{X}_{k0} \quad \Rightarrow \quad \sum_{i=1}^{N} D_{i0}\boldsymbol{\lambda}_{ki} = -\boldsymbol{\lambda}_{k0},$$

$$\mathbf{X}_{kj} \quad \Rightarrow \quad \sum_{i=1}^{N} D_{ij}\boldsymbol{\lambda}_{ki} = h\nabla_x H(\mathbf{X}_{kj}, \mathbf{U}_{kj}, \boldsymbol{\lambda}_{kj}), \quad 1 \leq j < N,$$

$$\mathbf{X}_{kN} \quad \Rightarrow \quad \sum_{i=1}^{N} D_{iN}\boldsymbol{\lambda}_{ki} = h\nabla_x H(\mathbf{X}_{kN}, \mathbf{U}_{kN}, \boldsymbol{\lambda}_{kN}) + \boldsymbol{\lambda}_{k+1,0},$$

$$\boldsymbol{\lambda}_{K+1,0} := \nabla C(\mathbf{X}_{KN}),$$

$$\mathbf{U}_{ki} \quad \Rightarrow \quad -\nabla_u H\left(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \boldsymbol{\lambda}_{ki}\right) \in N_{\mathcal{U}}(\mathbf{U}_{ki}).$$

First-order Optimality in Polynomial Setting

Lemma. The multipliers $\boldsymbol{\lambda}_k \in \mathbb{R}^{Nn}$, $1 \le k \le K$, satisfy the KKT conditions if and only if the polynomial $\boldsymbol{\lambda}_k \in \mathcal{P}^n_{N-1}$ given by $\boldsymbol{\lambda}_k(\tau_i) = \boldsymbol{\lambda}_{ki}/\omega_i$, $1 \le i \le N$, satisfies the following conditions and $\boldsymbol{\lambda}_{k0} = \boldsymbol{\lambda}_k(-1)$.

$$
\begin{aligned}
\dot{\boldsymbol{\lambda}}_k(\tau_i) &= -h\nabla_x H\left(\mathbf{x}_k(\tau_i), \mathbf{u}_{ki}, \boldsymbol{\lambda}_k(\tau_i)\right), \quad 1 \le i < N, \\
\dot{\boldsymbol{\lambda}}_k(1) &= -h\nabla_x H\left(\mathbf{x}_k(1), \mathbf{u}_{kN}, \boldsymbol{\lambda}_k(1)\right) + \left(\boldsymbol{\lambda}_k(1) - \boldsymbol{\lambda}_{k+1}(-1)\right)/\omega_N,
\end{aligned}
$$

$$
\text{where } \boldsymbol{\lambda}_{K+1}(-1) := \nabla C\left(\mathbf{x}_K(1)\right)
$$

$$
N_{\mathcal{U}}(\mathbf{u}_{ki}) \ni -\nabla_u H\left(\mathbf{x}_k(\tau_i), \mathbf{u}_{ki}, \boldsymbol{\lambda}_k(\tau_i)\right), \quad 1 \le i \le N.
$$

NOTE: The discrete $\boldsymbol{\lambda}$ is typically discontinuous at the mesh points.

$\mathbf{D}^\dagger$ for Radau

The differentiation matrix $\mathbf{D}^\dagger$ for the space of polynomials of degree $N-1$ evaluated at τ_i, $1 \le i \le N$, is given by

$$D^\dagger_{NN} = -D_{NN} + \frac{1}{\omega_N} \text{ and } D^\dagger_{ij} = -\frac{\omega_j}{\omega_i} D_{ji} \text{ otherwise.}$$

In other words, if p is a polynomial of degree at most $N-1$ and if $\mathbf{p} \in \mathbb{R}^N$ is the vector with i-th component $p_i = p(\tau_i)$, then

$$(\mathbf{D}^\dagger \mathbf{p})_i = \dot{p}(\tau_i), \quad 1 \le i \le N.$$

Proof

If $p \in \mathcal{P}_N$ and $q \in \mathcal{P}_{N-1}$ with $p(-1) = 0$, then integration by parts gives

$$\int_{-1}^{1} \dot{p}(\tau)q(\tau)d\tau = p(1)q(1) - \int_{-1}^{1} p(\tau)\dot{q}(\tau)d\tau.$$

Since $\dot{p}q$ and $p\dot{q}$ are polynomials of degree at most $2N-2$, Radau quadrature is exact and we have

$$\sum_{j=1}^{N} w_j \dot{p}_j q_j = p_N q_N - \sum_{j=1}^{N} w_j p_j \dot{q}_j,$$

where $p_j = p(\tau_j)$ and $\dot{p}_j = \dot{p}(\tau_j)$. This yields

$$(\mathbf{W}\dot{\mathbf{p}})^{\mathsf{T}}\mathbf{q} = p_N q_N - (\mathbf{W}\mathbf{p})^{\mathsf{T}}\dot{\mathbf{q}}.$$

Substituting $\dot{\mathbf{p}} = \mathbf{D}_{1:N}\mathbf{p}$ and $\dot{\mathbf{q}} = \mathbf{D}^{\dagger}\mathbf{q}$ gives

$$\mathbf{p}^{\mathsf{T}}\mathbf{D}_{1:N}^{\mathsf{T}}\mathbf{W}\mathbf{q} = p_N q_N - \mathbf{p}^{\mathsf{T}}\mathbf{W}\mathbf{D}^{\dagger}\mathbf{q}.$$

Proof continued . . .

Rearrange into the following form:

$$\mathbf{p}^{\mathsf{T}}(\mathbf{D}_{1:N}^{\mathsf{T}}\mathbf{W} + \mathbf{W}\mathbf{D}^{\dagger} - \mathbf{e}_N\mathbf{e}_N^{\mathsf{T}})\mathbf{q} = 0,$$

where $\mathbf{e}_N$ is the last column of $\mathbf{I}$. Since this identity must be satisfied for all choices of $\mathbf{p}$ and $\mathbf{q}$, we deduce that

$$\mathbf{D}^{\dagger} = \mathbf{W}^{-1}\mathbf{e}_N\mathbf{e}_N^{\mathsf{T}} - \mathbf{W}^{-1}\mathbf{D}_{1:N}^{\mathsf{T}}\mathbf{W}.$$

Proof of the Lemma

Define $\mathbf{\Lambda}_{ki} = \boldsymbol{\lambda}_{ki}/\omega_i$ for $1 \le i \le N$, $\mathbf{\Lambda}_{k0} = \boldsymbol{\lambda}_{k0}$, and $D_{ij}^{\ddagger} = -\omega_j D_{ji}/\omega_i$. These substitutions in the KKT conditions yield

$$\begin{aligned} \sum_{j=1}^{N} D_{ij}^{\ddagger}\mathbf{\Lambda}_{kj} &= -h\nabla_x H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}), \quad 1 \le i < N, \\ \sum_{i=1}^{N} D_{Ni}^{\ddagger}\mathbf{\Lambda}_{ki} &= -[h\nabla_x H(\mathbf{X}_{kN}, \mathbf{U}_{kN}, \mathbf{\Lambda}_{kN}) + \mathbf{\Lambda}_{k+1,0}/\omega_N], \\ N_{\mathcal{U}}(\mathbf{U}_{ki}) &\ni -\nabla_u H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}), \quad 1 \le i \le N. \end{aligned} \tag{1}$$

Proof continued

Since the polynomial that is identically equal to $\mathbf{1}$ has derivative $\mathbf{0}$, we have $\mathbf{D1} = \mathbf{0}$, which implies that $\mathbf{D}_0 = -\sum_{j=1}^{N} \mathbf{D}_j$, where $\mathbf{D}_j$ is the j-th column of $\mathbf{D}$. Hence,

$$\begin{aligned}
\boldsymbol{\Lambda}_{k0} &= -\sum_{i=1}^{N} \boldsymbol{\lambda}_{ki} D_{i0} = \sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{\lambda}_{ki} D_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_j \left(\frac{\boldsymbol{\lambda}_{ki}}{\omega_i}\right)(\omega_i D_{ij}/\omega_j) \\
&= -\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i D_{ij}^{\ddagger} \boldsymbol{\Lambda}_{kj} \qquad\qquad (\ddagger) \\
&= \boldsymbol{\Lambda}_{k+1,0} + h\sum_{i=1}^{N} \omega_i \nabla_x H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \boldsymbol{\Lambda}_{ki}),
\end{aligned}$$

where the discrete costate equations are used in the last step.

Let $\boldsymbol{\lambda}_k \in \mathcal{P}^n_{N-1}$ satisfy $\lambda_k \tau_i = \boldsymbol{\Lambda}_{ki}$, $1 \leq i \leq N$.

$$\begin{aligned} \sum_{j=1}^{N} D^{\ddagger}_{ij} \boldsymbol{\Lambda}_{kj} &= \dot{\boldsymbol{\lambda}}_k(\tau_i), \quad 1 \leq i < N, \quad \text{and} \\ \sum_{j=1}^{N} D^{\ddagger}_{Nj} \boldsymbol{\Lambda}_{kj} &= \dot{\boldsymbol{\lambda}}_k(1) - \boldsymbol{\lambda}_k(1)/\omega_N. \end{aligned} \tag{2}$$

This substitution in (‡) yields

$$\boldsymbol{\Lambda}_{k0} = \boldsymbol{\lambda}_k(1) - \sum_{i=1}^{N} \omega_i \dot{\boldsymbol{\lambda}}_k(\tau_i).$$

Since $\dot{\boldsymbol{\lambda}}_k \in \mathcal{P}^n_{N-2}$ and N-point Radau quadrature is exact for these polynomial, we have

$$\sum_{i=1}^{N} \omega_i \dot{\boldsymbol{\lambda}}_k(\tau_i) = \int_{-1}^{1} \dot{\boldsymbol{\lambda}}_k(\tau) d\tau = \boldsymbol{\lambda}_k(1) - \boldsymbol{\lambda}_k(-1).$$

Combine last two equations to obtain

$$\boldsymbol{\Lambda}_{k0} = \boldsymbol{\lambda}_k(-1) \qquad ((3))$$

(1) + (2) + (3) yield the formula for $\dot{\boldsymbol{\lambda}}_k(1)$.

Fundamental Differences: Gauss/hp Radau

- For Gauss, the entries of the matrices $\mathbf{D}$ and $\mathbf{D}^\dagger$ satisfy

$$D_{ij} = -D^\dagger_{N+1-i,N+1-j}, \quad 1 \le i \le N, \quad 1 \le j \le N.$$

  In other words, $\mathbf{D}_{1:N} = -\mathbf{J}\mathbf{D}^\dagger_{1:N}\mathbf{J}$ where $\mathbf{J}$ is the exchange matrix with ones on its counterdiagonal and zeros elsewhere. Hence, $\|\mathbf{D}^{-1}\|_\infty = \|(\mathbf{D}^\dagger)^{-1}\|_\infty$.
- 20,000 Yen Prize Problem: Show that $\|(\mathbf{D}^\ddagger)^{-1}\|\infty \le 2$ and the rows of $[\mathbf{W}^{1/2}\mathbf{D}^\ddagger]^{-1}$ have Euclidean length bounded by $\sqrt{2}$.
- In the *hp*-scheme, only need to have K is large enough, or equivalently h is small enough, that $2hd_1 < 1$ and $2hd_2 < 1$, where

$$\begin{aligned} d_1 &= \sup_{t\in\Omega_0} \|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))\|_\infty \\ d_2 &= \sup_{t\in\Omega_0} \|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))^\mathsf{T}\|_\infty \end{aligned}$$

Discrete and continuous vectors

$$\begin{aligned}
\mathbf{X}^*_{kj} &= \mathbf{x}^*_k(\tau_j), \quad 0 \le j \le N, \quad 1 \le k \le K, \\
\mathbf{U}^*_{kj} &= \mathbf{u}^*_k(\tau_j), \quad 1 \le j \le N, \quad 1 \le k \le K, \\
\mathbf{\Lambda}^*_{kj} &= \boldsymbol{\lambda}^*_k(\tau_j), \quad 0 \le j \le N, \quad 1 \le k \le K. \\
\mathbf{X}^N_{kj} &= \mathbf{x}^N_k(\tau_j), \quad 0 \le j \le N, \quad 1 \le k \le K, \\
\mathbf{U}^N_{kj} &= \mathbf{u}^N_{kj}, \qquad\quad 1 \le j \le N, \quad 1 \le k \le K, \\
\mathbf{\Lambda}^N_{kj} &= \boldsymbol{\lambda}^N_k(\tau_j), \quad 0 \le j \le N, \quad 1 \le k \le K.
\end{aligned}$$

NOTE: $\boldsymbol{\lambda}^*_k(1) = \boldsymbol{\lambda}^*_{k+1}(-1)$ but $\boldsymbol{\lambda}^N_k(1) \ne \boldsymbol{\lambda}^N_{k+1}(-1)$.

Main Theorem

Theorem. Suppose $(\mathbf{x}^*, \mathbf{u}^*)$ is a local minimizer for the continuous problem with $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{PH}^\eta(\Omega_0; \mathbb{R}^n)$ for some $\eta \geq 2$. If some smoothness and local convexity conditions hold, then for N sufficiently large or for h sufficiently small with $N \geq 2$, the discrete problem has a local minimizer $\mathbf{x}_k^N \in \mathcal{P}_N^n$ and $\mathbf{u}_k^N \in \mathbb{R}^{mN}$, and an associated multiplier $\boldsymbol{\lambda}_k^N \in \mathcal{P}_{N-1}^n$, $1 \leq k \leq K$; moreover, there exists a constant c, independent of h, N and η, such that

$$\max\left\{\left\|\mathbf{X}^N - \mathbf{X}^*\right\|_\infty, \left\|\mathbf{U}^N - \mathbf{U}^*\right\|_\infty, \left\|\boldsymbol{\Lambda}^N - \boldsymbol{\Lambda}^*\right\|_\infty\right\}$$
$$\leq h^{p-1}\left(\frac{c}{N}\right)^{p-1} |\mathbf{x}^*|_{\mathcal{PH}^p(\Omega_0)} + h^{q-1}\left(\frac{c}{N}\right)^{q-1.5} |\boldsymbol{\lambda}^*|_{\mathcal{PH}^q(\Omega_0)},$$

where $p = \min(\eta, N+1)$ and $q = \min(\eta, N)$.

Example

$$\text{minimize} \quad \frac{1}{2}\int_0^1 [x^2(t) + u^2(t)]\, dt$$

$$\text{subject to} \quad \dot{x}(t) = u(t), \quad u(t) \le 1, \quad x(t) \le \frac{2\sqrt{e}}{1-e}, \quad t \in [0,1],$$

$$x(0) = \frac{5e+3}{4(1-e)}.$$

Exact solution:

$$0 \le t \le \tfrac{1}{4}: \quad x^*(t) = t - \tfrac{1}{4} + \tfrac{1+e}{1-e}, \qquad u^*(t) = 1,$$

$$\tfrac{1}{4} \le t \le \tfrac{3}{4}: \quad x^*(t) = \tfrac{e^{t-\frac{1}{4}}}{1-e}(1 + e^{\frac{3}{2}-2t}), \quad u^*(t) = \tfrac{e^{t-\frac{1}{4}}}{1-e}(1 - e^{\frac{3}{2}-2t}),$$

$$\tfrac{3}{4} \le t \le 1: \quad x^*(t) = \tfrac{2\sqrt{e}}{1-e}, \qquad u^*(t) = 0.$$

Error in Solution