Reinforcement Learning and Optimal Control

IFT6760C, Fall 2021

Pierre-Luc Bacon

November 15, 2021

Deterministic Discrete-time Trajectory Optimization

Mayer problem:

$$
\begin{aligned}
&\text{minimize} \;\; c_T(x_T) \\
&\text{subject to} \;\; x_{t+1} = f_t(x_t, u_t), \;\; t = 1, \dots T - 1 \\
&\text{given} \;\; x_1
\end{aligned}
$$

In this formulation, the optimization variables are $x_1, \dots, x_T$ and $u_1, \dots, u_{T-1}$. We solve for both the controls **and the states** that minimize the terminal cost function.

Lagrange problem:

$$\begin{aligned}
&\text{minimize} \quad \sum_{t=1}^{T} c(x_T, u_t) \\
&\text{subject to} \quad x_{t+1} = f_t(x_t, u_t), \quad t = 1, \ldots T - 1 \\
&\text{given} \quad x_1
\end{aligned}$$

Bolza problem:

$$\begin{aligned}
&\text{minimize} \quad c_T(x_T) + \sum_{t=1}^{T-1} c(x_T, u_t) \\
&\text{subject to} \quad x_{t+1} = f_t(x_t, u_t), \quad t = 1, \ldots T - 1 \\
&\text{given} \quad x_1
\end{aligned}$$

Transformation as unconstrained problem

Idea: we can reconstruct the state at any point in time by simulating the deterministic system forward in time. More precisely, let $\phi_t(u)$ be a function such that $\phi_t(u) = x_t$, and where $u = (u_1, \ldots, u_T - 1)$.

$$\text{minimize} \quad c_T(\phi_T(u)) \quad .$$

By the chain rule, we have $D_i J(u) = Dc_T(x_T) D_i \phi_T(u)$ with the recursion:

$$\begin{aligned} D_i \phi_{t+1}(u) &= D_1 f_t(x_t, u_t) D_i \phi_t(u), \forall t = i+1, \ldots, T \\ D_i \phi_{i+1} &= D_2 f_i(x_i, u_i) \quad . \end{aligned}$$

Adjoint/Costate vector

We can choose to accumulate the derivatives from left to right, in which case:

$$D_iJ(u) = \underbrace{\underbrace{\underbrace{Dc_T(x_T)}_{\lambda_T} D_1 f_{T-1}(x_{t-1}, u_{T-1})}_{\lambda_{T-1}} \dots D_1 f_{i+1}(x_{i+1}, u_{i+1})}_{\lambda_{i+1}} D_2 f_i(x_i, u_i)$$

Therefore, $D_iJ(u) = \lambda_i D_2 f_i(x_i, u_i)$. The $\{\lambda_i\}_{i+1}^{T}$ are called **adjoint** or **co-state** vectors.

Adjoint Equation

The accumulation of derivative backwards in time can therefore be described by:

$$
\begin{aligned}
\lambda_T &= Dc_T(x_T) \\
\lambda_t &= \lambda_{t+1} D_1 f_t(x_t, u_t), \quad t = T-1, \ldots, i+1
\end{aligned}
$$

which we call the **adjoint equation**. In order to solve this equation we need to first run the system forward in time:

$$x_{t+1} = f_t(x_t, u_t), \quad t = 1, \ldots, T \quad .$$

We can then finally get:

$$D_i J(u) = \lambda_{i+1} D_2 f_i(x_i, u_i) \quad .$$

Constrained Optimization

Consider the following nonlinear program with equality constraints:

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad h(x) = 0 \quad . \end{aligned}$$

where $x \in \mathbb{R}^n$ and $h : \mathbb{R}^n \to \mathbb{R}^m$.

We know that in the unconstrained case that if $x^\star$ is a local minimum then, $Df(x^\star) = 0$. But what if we have constraints?

Lagrange Multiplier Theorem

In the following, we assume that f and h are continuously differentiable functions.

Theorem Let $x^\star$ be a local minimum of f and feasible solution for which the constraint gradients $Dh_1(x^\star), \ldots, Dh_m(x^\star)$ are linearly independent. There exists a unique *Lagrange multiplier* row vector $\lambda^\star \in \mathbb{R}^{1\times m}$ such that:

$$D_1L(x^\star, \lambda^\star) = 0 \quad \text{and} \quad D_2L(x^\star, \lambda^\star) = 0$$

$$\text{where} \quad L(x, \lambda) \triangleq f(x) + \lambda h(x) \quad .$$

Interpretation

$$D_1 L(x^\star, \lambda^\star) = Df(x^\star) + \lambda^\star Dh(x^\star) = 0$$

Therefore $Df(x^\star) = -\sum_{i=1}^m \lambda_i^\star Dh_i(x^\star)$. This means that $Df(x^\star)$ can be expressed a linear combination of constraint gradients. The objective gradient is in the span of the constraint gradients at $x^\star$.

Intuition for Proof

We can try to "reparameterize" our problem such that the constraints are pushed inside the main objective (similar to what we did in the optimal control case). This time however, we proceed to this transformation using the implicit function theorem.

Consider partition of the variables such that $x = (y, z)$. We will view y as an dependent variable and z as an independent one.

$$\begin{aligned} &\text{minimize} \quad f(y, z) \\ &\text{subject to} \quad h(y, z) = 0 \quad . \end{aligned}$$

Implicit Function

By the implicit function theorem, if we have $h(y^\star, z^\star) = 0$ for some pair $x^\star, y^\star$ and that $[D_1 h(y^\star, z^\star)]^{-1}$ exists then there exists a function ϕ such that $h(\phi(z^\star), z^\star) = 0$ and $D\phi(z^\star) = [D_1 h(\phi(z^\star), z^\star)]^{-1} D_2 h(\phi(z^\star), z^\star)$.

We can then re-write our constrained optimization problem as:

$$\text{minimize} \quad f(\phi(z^\star), z^\star) \quad ,$$

and we know that the constraint will be satisfied (since ϕ maps to $y^\star$)

First-order Condition

Now that are dealing with a constrained problem, we know that

$$DJ(z^\star) = 0 \quad \text{where} \quad J(z^\star) = f(\phi(z^\star), z^\star) \quad .$$

By the chain rule:

$$\begin{aligned}
DJ(z^\star) &= D_1 f(\phi(z^\star), z^\star) D\phi(z^\star) + D_2 f(\phi(z^\star), z^\star) \\
&= D_1 f(\phi(z^\star), z^\star)[D_1 h(\phi(z^\star), z^\star)]^{-1} D_2 h(\phi(z^\star), z^\star) + D_2 f(\phi(z^\star), z^\star) \\
&= \lambda^\star D_2 h(\phi(z^\star), z^\star) + D_2 f(\phi(z^\star), z^\star) \quad .
\end{aligned}$$

Therefore there exists a unique vector $\lambda^\star$ such that the above holds.

Generalized Reduced Gradient Methods

This idea is the basis for the so-called **Generalized Reduced Gradient Method** (GRG) for solving nonlinear programs. Pick a subset of the variables, satisfy the corresponding constraints (by Newton's method for example), take a gradient step using implicit differentiation

From Lagrange Multipliers to Adjoint Equation

Let's tackle our OCP using the Lagrange multiplier theorem.

$$
\begin{aligned}
&\text{minimize} \quad c_T(x_T) + \sum_{t=1}^{T-1} c_t(x_t, u_t) \\
&\text{subject to} \quad x_{t+1} = f_t(x_t, u_t), \;\; t = 1, \dots T-1 \\
&\text{given} \quad x_1
\end{aligned}
$$

We then know that if there exists feasible $x_1, \dots, x_T, u_1, \dots, u_{T-1}$, then it must be that there exists a unique set $\{\lambda_t^\star\}_1^{T-1}$ such that $DL(x^\star, u^\star, \lambda^\star) = 0$ in:

$$
L(x, u, \lambda) \triangleq c_T(x_T) + \sum_{t=1}^{T-1} c_t(x_t, u_t) + \sum_{t=1}^{T-1} \lambda_t(f_t(x_t, u_t) - x_{t+1}) \;\; .
$$

First-Order Optimality

For mathematical convenience, we can write the Lagrangian as:

$$
\begin{aligned}
L(x, u, \lambda) &\triangleq c_T(x_T) + \sum_{t=1}^{T-1} \left(c_t(x_t, u_t) + \lambda_{t+1} \left(f_t(x_t, u_t) - x_{t+1}\right)\right) \\
&= c_T(x_T) + \lambda_T x_T - \lambda_1 x_1 \sum_{t=1}^{T-1} \left(c_t(x_t, u_t) + \lambda_{t+1} f_t(x_t, u_t) - \lambda_t x_t\right)
\end{aligned}
$$

By noting that:

$$
\sum_{t=1}^{T-1} \lambda_{t+1} x_{t+1} = \lambda_T x_T - x_1 \lambda_1 + \sum_{t=1}^{T-1} \lambda_t x_t
$$

The advantage being that x_t appears only once in the summation.

Adjoint Equation

If there exists a feasible local minimum $(x^\star, u^\star)$ then there exists a Lagrange multiplier $\lambda^\star$ such that $DL(x^\star, u^\star, \lambda^\star) = 0$.

$$D_{x_i} L(x^\star, u^\star, \lambda^\star) = \begin{cases} Dc_T(x_T^\star) + \lambda_T^\star & i = T \\ D_1 c_t(x_t^\star, u_t^\star) + \lambda_{t+1} D_1 f_t(x_t^\star, u_t^\star) - \lambda_t^\star & i \in \{i, \ldots, T-1\} \end{cases}$$

In other words:

$$\begin{aligned} \lambda_T^\star &= Dc_T(x_T^\star) \\ \lambda_t^\star &= D_1 c_t(x_t^\star, u_t^\star) + \lambda_{t+1} D_1 f_t(x_t^\star, u^\star) \quad . \end{aligned}$$

Optimality of the controls:

$$D_{u_t}L(x^\star, u^\star, \lambda^\star) = D_2 c_t(x_t^\star, u_t^\star) + \lambda_{t+1} D_2 f_t(x_t^\star, u_t^\star) \quad .$$

Forward dynamics:

$$D_{\lambda_{t+1}}L(x^\star, u^\star, \lambda^\star) = f_t(x_t, u_t) - x_{t+1} = 0 \quad .$$

The three partial derivatives form the basis for the discrete-time Pontryagin maximum principle (PMP)

Algorithms for constrained optimization

(Shown on the board. More in next lecture)