

ECE 6553: Optimal Control Notes

Klaus Z. Okkelberg

Spring 2017

Chapter 1

Parameter Optimization

1.1 What is optimal control?

Optimal Maximize/minimize cost (subject to constraints): $\min_u g(u)$

With constraints,

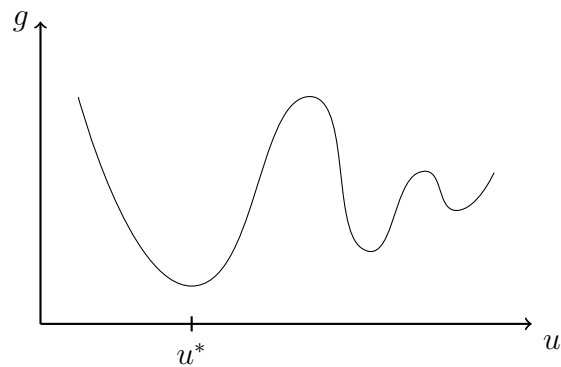
$$\begin{aligned} \min_u \quad & g(u) \\ \text{s.t.} \quad & \begin{cases} h_1(u) = 0 \\ h_2(u) \leq 0 \end{cases} \end{aligned}$$

First-order necessary condition (FONC):

$$\frac{\partial g}{\partial u}(u^*) = 0$$

Optimality can be

- local vs global
- max vs min



Control control design: pick u such that specifications are satisfied:

$$\dot{x} = f(x, u), \quad \dot{x} = Ax + Bu,$$

where $x(t) \in \mathbb{R}^n$ is the state, $u(t) \in \mathbb{R}^m$ is the control, and $f(\cdot)$ is the dynamics.

Actually, x and u are signals:

$$x : [0, T] \rightarrow \mathbb{R}^n, \quad u : [0, T] \rightarrow \mathbb{R}^m$$

Optimal control find the “best” u !

For “best” to mean anything, we need a cost. The big/deep question is

$$\frac{\partial \text{“cost”}}{\partial u} = 0$$

Example

Suppose we have a car with position p . Its acceleration \ddot{p} is controlled by the gas/brake input u ($\ddot{p} = u$). In order to express the dynamics of the system in the form $\dot{x} = f(x, u)$, we introduce state variables:

$$\begin{aligned} x_1 = p \\ x_2 = \dot{p} \end{aligned} \implies \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u \end{cases}$$

The task is to move the car from its initial position to a stop at a distance c away.

Minimum energy problem

$$\begin{aligned} \min_u \quad & \int_0^T u^2(t) dt \\ \text{s.t.} \quad & \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u \end{cases} \\ & x_1(0) = 0, x_2(0) = 0 \\ & x_1(T) = c, x_2(T) = 0 \end{aligned}$$

Minimum time problem

$$\begin{aligned} \min_{u, T} \quad & T = \int_0^T dt \\ \text{s.t.} \quad & \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = u \end{cases} \\ & x_1(0) = 0, x_2(0) = 0 \\ & x_1(T) = c, x_2(T) = 0 \\ & u(t) \in [u_{\min}, u_{\max}] \end{aligned}$$

The general optimal control problem we will solve will look like

$$\begin{aligned} \min_{u,T} \quad & \int_0^T L(x(t), u(t), t) dt + \Psi(x(T)) \\ \text{s.t.} \quad & \dot{x}(t) = f(x(t), u(t), t), \quad t \in [0, T] \\ & x(0) = x_0 \\ & x(T) \in S \\ & u(t) \in \Omega, \quad t \in [0, T] \end{aligned}$$

where $\Psi(\cdot)$ is the terminal cost and S is the terminal manifold. This is a so-called **Bolza Problem**.

What tools do we need to solve this?

1. optimality conditions $\partial \text{cost} / \partial u = 0$
2. some way of representing the optimal signal $u^*(x, t)$
3. some way of actually finding/computing the optimal controllers

1.2 Unconstrained Optimization

Let the decision variable be $u = [u_1, \dots, u_m]^T \in \mathbb{R}^m$. The cost is $g(u) \in C^1$ (C^k means k times continuously differentiable). The problem is

$$\min_u g(u), \quad g : \mathbb{R}^m \rightarrow \mathbb{R}$$

For u^* to be a minimizer, we need

$$\frac{\partial g}{\partial u}(u^*) = 0$$

Definition. u^* is a (local) minimizer to g if $\exists \delta > 0$ s.t.

$$\begin{aligned} g(u^*) &\leq g(u) \quad \forall u \in B_\delta(u^*) \\ B_\delta(u^*) &= \{u \mid \|u - u^*\| \leq \delta\} \end{aligned}$$

Note:

- $\frac{\partial g}{\partial u}(u^*) \delta u \in \mathbb{R}$ and δu is $m \times 1$, so $\frac{\partial g}{\partial u}$ is a $1 \times m$ row vector. For the column vector,

$$\nabla g = \frac{\partial g^T}{\partial u} \in \mathbb{R}^m$$

- $\frac{\partial g}{\partial u} \delta u$ is an inner product

$$\langle \nabla g, \delta u \rangle = \left\langle \frac{\partial g^T}{\partial u}, \delta u \right\rangle$$

- $o(\epsilon)$ encodes higher-order terms

$$\lim_{\epsilon \rightarrow 0} \frac{o(\epsilon)}{\epsilon} = 0 \quad \text{“faster than linear”}$$

This is opposed to big-O notation:

$$\lim_{\epsilon \rightarrow 0} \frac{\mathcal{O}(\epsilon)}{\epsilon} = c$$

- δu has direction and scale so we could write it as

$$\delta u = \epsilon v, \quad \epsilon \in \mathbb{R}, \quad v \in \mathbb{R}^m$$

Theorem. For u^* to be a minimizer, we need

$$\frac{\partial g}{\partial u}(u^*) = 0$$

or, equivalently,

$$\frac{\partial g}{\partial u}(u^*)v = 0 \quad \forall v \in \mathbb{R}^m$$

Proof. Let u^* be a minimizer. Evaluating the cost $g(u)$ in the ball and using Taylor's expansion,

$$g(u^* + \delta u) = g(u^*) + \frac{\partial g}{\partial u}(u^*)\delta u + o(\|\delta u\|) = g(u^*) + \epsilon \frac{\partial g}{\partial u}(u^*)v + o(\epsilon)$$

Assume that $\frac{\partial g}{\partial u} \neq 0$. Then we could pick $v = -\frac{\partial g^T}{\partial u}(u^*)$, i.e.

$$g(u^* + \epsilon v) = g(u^*) - \epsilon \left\| \frac{\partial g^T}{\partial u}(u^*) \right\|^2 + o(\epsilon)$$

Note that the second term is negative per our assumptions. So, for ϵ sufficiently small, we have

$$g\left(u^* - \epsilon \frac{\partial g^T}{\partial u}(u^*)\right) < g(u^*)$$

This contradicts u^* being a minimizer. \times (crossed swords) □

Definition (Positive definite). $M = M^T \succ 0$ if

$$\begin{aligned} z^T M z &> 0 \quad \forall z \neq 0, \quad z \in \mathbb{R}^m \\ \iff M &\text{ has real and positive eigenvalues} \end{aligned}$$

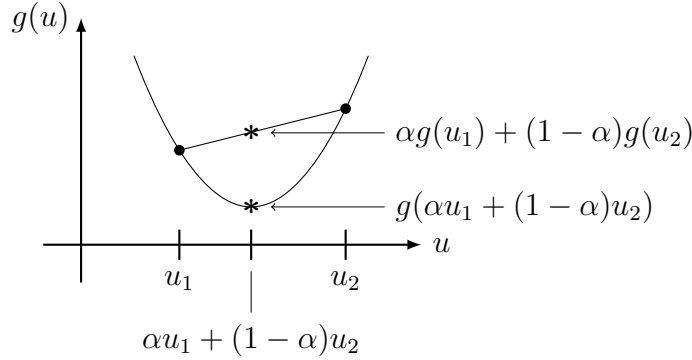
Theorem. If $g \in C^2$, then a **sufficient** condition for u^* to be a (local) minimizer is

$$1. \quad \frac{\partial g}{\partial u}(u^*) = 0$$

2. $\frac{\partial^2 g}{\partial u^2}(u^*) \succ 0$ (the Hessian is positive definite)

Definition. $g : \mathbb{R}^m \rightarrow \mathbb{R}$ is convex if

$$g(\alpha u_1 + (1 - \alpha)u_2) \leq \alpha g(u_1) + (1 - \alpha)g(u_2) \quad \forall \alpha \in [0, 1], \quad u_1, u_2 \in \mathbb{R}^m$$



Theorem. If $\frac{\partial^2 g}{\partial u^2}(u) \succeq 0 \quad \forall u \in \mathbb{R}^m$, then g is convex. (\iff for $g \in C^2$)

Example $\min_u u^T Q u - b^T u$ where $Q = Q^T \succ 0$ (positive definite matrix)

$$\begin{aligned} \frac{\partial g}{\partial u} &= \frac{\partial}{\partial u} (u^T Q u - b^T u) \\ &= u^T Q^T + u^T Q - b^T \\ &= 2u^T Q - b^T \\ \frac{\partial^2 g}{\partial u^2} &= 2Q \end{aligned} \quad \frac{\partial^2 g}{\partial u^2} = \begin{bmatrix} \frac{\partial^2 g}{\partial u_1^2} & \cdots & \frac{\partial^2 g}{\partial u_1 \partial u_m} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial u_m \partial u_1} & \cdots & \frac{\partial^2 g}{\partial u_m^2} \end{bmatrix}$$

From $\frac{\partial g}{\partial u} = 2u^T Q - b^T = 0$,

$$u = \frac{1}{2} Q^{-1} b$$

To see whether this is a minimizer, consider the Hessian. Since $Q \succ 0$, it follows that $\frac{\partial^2 g}{\partial u^2}(u^*) \succ 0$ and $u^* = \frac{1}{2} Q^{-1} b$ is a (local) minimizer. Additionally, since $\frac{\partial^2 g}{\partial u^2} \succ 0$, g is convex and u^* is a global minimizer. In fact, since we have strict convexity ($\succ 0$ rather than $\succeq 0$), it is the unique global minimizer.

In optimal control, *local* is typically all we can ask for. In optimization, we can do better!

But wait, just because we know $\frac{\partial g}{\partial u} = 0$, it doesn't follow that we can actually find $u^* \dots$

1.3 Numerical Methods

Idea: $u_{k+1} = u_k + \text{step}_k$. What should step_k be? For small $\text{step}_k = \gamma_k v_k$,

$$g(u_k + \text{step}_k) = g(u_k) + \frac{\partial g}{\partial u}(u_k) \cdot \text{step}_k + o(\|\text{step}_k\|) = g(u_k) + \gamma_k \frac{\partial g}{\partial u}(u_k) v_k + o(\gamma_k)$$

A perfectly reasonable choice of step direction is

$$v_k = -\frac{\partial g^T}{\partial u}(u_k),$$

known as the *steepest descend* direction. This produces

$$g\left(u_k - \gamma_k \frac{\partial g}{\partial u}(u_k)\right) = g(u_k) - \gamma_k \left\| \frac{\partial g}{\partial u}(u_k) \right\|^2 + o(\gamma_k)$$

Steepest descent

$$u_{k+1} = u_k - \gamma_k \frac{\partial g^T}{\partial u}(u_k)$$

Note:

- What should γ_k be?
- This method “pretends” that $g(u)$ is linear. If we pretend $g(u)$ is quadratic, we get

$$u_{k+1} = u_k - \left(\frac{\partial^2 g}{\partial u^2}(u_k) \right)^{-1} \frac{\partial g^T}{\partial u}(u_k),$$

i.e. Newton’s Method

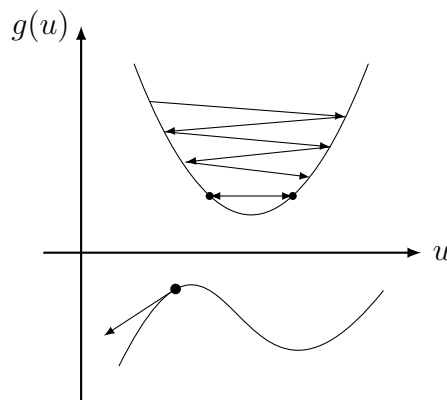
This course: steepest descent

Step-size selection?

- Choice 1: $\gamma_k = \gamma$ “small” $\forall k$; will get close to a minimizer if u_0 is close enough and γ small enough

Problems:

- You may not converge! (but you’ll get close)
- You may go off to infinity (diverge)



- Choice 2: Reduce γ_k as a function of k ; will get close to a minimizer if u_0 is close enough

Problem: slow

Theorem. If u_0 is close enough to u^* and γ_k satisfies

$$\begin{aligned} - \sum_{k=0}^{\infty} \gamma_k &= \infty \\ - \sum_{k=0}^{\infty} \gamma_k^2 &< \infty \end{aligned}$$

e.g. $\gamma_k = c/k$, then $u_k \rightarrow u^*$ as $k \rightarrow \infty$.

- Choice 3: **Armijo step-size:** Take as big a step as possible, but no larger
Pick $\alpha \in (0, 1)$, $\beta \in (0, 1)$. Let i be the smallest non-negative integer such that

$$\begin{aligned} g\left(u_k - \beta^i \frac{\partial g^T}{\partial u}(u_k)\right) - g(u_k) &< -\alpha \beta^i \left\| \frac{\partial g}{\partial u}(u_k) \right\|^2 \\ u_{k+1} &= u_k - \beta^i \frac{\partial g^T}{\partial u}(u_k) \end{aligned}$$

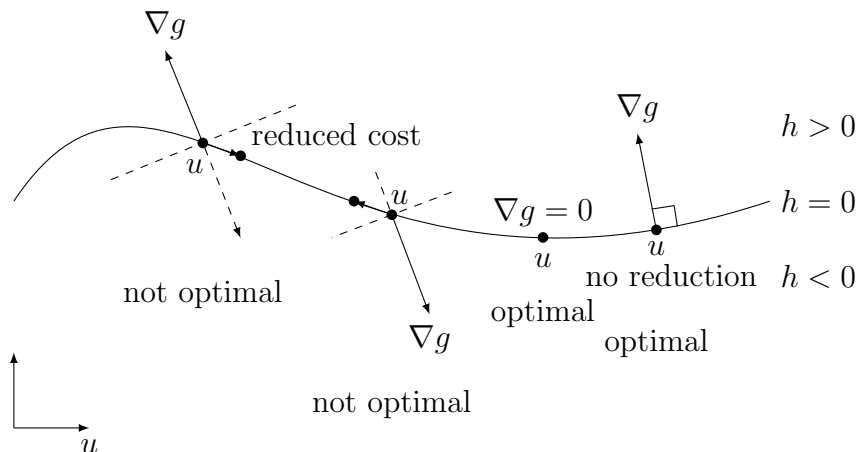
This will get to a minimizer blazingly fast if u_0 is close enough.

1.4 Constrained Optimization

Equality constraints:

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & g(u) \\ \text{s.t.} \quad & h(u) = 0 \end{aligned}$$

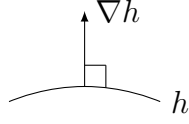
Consider $u \in \mathbb{R}^2$, $h : \mathbb{R}^2 \rightarrow \mathbb{R}$



So u is (locally) optimal if $\nabla g \parallel$ (is parallel to) the normal vector to tangent plane to h .

Fact: (HW# 1)

$$\nabla h \perp Th \quad (\text{tangent plane to } h)$$



We need $\nabla g \parallel \nabla h$ at u^* for optimality, i.e.

$$\frac{\partial g}{\partial u}(u^*) = \alpha \frac{\partial h}{\partial u}(u^*), \quad \text{for some } \alpha \in \mathbb{R}$$

or $(\lambda = -\alpha)$,

$$\frac{\partial g}{\partial u}(u^*) + \lambda \frac{\partial h}{\partial u}(u^*) = 0$$

or

$$\frac{\partial}{\partial u}(g(u^*) + \lambda h(u^*)) = 0, \quad \text{for some } \lambda \in \mathbb{R}$$

More generally,

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & g(u) \\ \text{s.t.} \quad & h(u) = \mathbf{0}, \quad h : \mathbb{R}^m \rightarrow \mathbb{R}^k \end{aligned}$$

Note that $h(u) = [h_1(u), \dots, h_k(u)]^T$.

We need $\frac{\partial g}{\partial u}(u^*)$ to be a linear combination of $\frac{\partial h_i}{\partial u}(u^*)$, $i = 1, \dots, k$, for exactly the same reasons, i.e.

$$\frac{\partial g}{\partial u}(u^*) = \sum_{i=1}^k \alpha_i \frac{\partial h_i}{\partial u}(u^*)$$

or $(\lambda = -[\alpha_1, \dots, \alpha_k]^T)$

$$\frac{\partial g}{\partial u}(u^*) + \lambda^T \frac{\partial h}{\partial u}(u^*) = 0$$

or

$$\frac{\partial}{\partial u}(g(u^*) + \lambda^T h(u^*)) = 0, \quad \text{for some } \lambda \in \mathbb{R}^k$$

Theorem. *If u^* is a minimizer to*

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & g(u) \\ \text{s.t.} \quad & h(u) = \mathbf{0}, \quad h : \mathbb{R}^m \rightarrow \mathbb{R}^k \end{aligned}$$

then $\exists \lambda \in \mathbb{R}^k$ s.t.

$$\begin{cases} \frac{\partial L}{\partial u}(u^*, \lambda) = 0 \\ \frac{\partial L}{\partial \lambda}(u^*, \lambda) = 0 \end{cases}$$

where the Lagrangian L is given by

$$L(u, \lambda) = g(u) + \lambda^T h(u)$$

Note:

- λ are the Lagrange multipliers
- $\frac{\partial L}{\partial \lambda} = 0$ is fancy speak for $h(u^*) = 0$

Example

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & \frac{1}{2} \|u\|^2 \\ \text{s.t.} \quad & Au = b \end{aligned}$$

where A is $k \times m$, $k \leq m$. Assume $(AA^T)^{-1}$ exists (constraints are linearly independent, none of the constraints are “duplicates”, all the constraints are essential).

$$\begin{aligned} L &= \frac{1}{2} u^T u + \lambda^T (Au - b) \\ \frac{\partial L}{\partial u} &= u^T + \lambda^T A = 0 \\ u^* &= -A^T \lambda \end{aligned}$$

Using the equality constraint,

$$\begin{aligned} Au^* &= b \\ -AA^T \lambda &= b \\ \lambda &= -(AA^T)^{-1} b \\ u^* &= A^T (AA^T)^{-1} b \end{aligned}$$

Example

$$\begin{aligned} \min \quad & u_1 u_2 + u_2 u_3 + u_1 u_3 \\ \text{s.t.} \quad & u_1 + u_2 + u_3 = 3 \end{aligned}$$

$$L = u_1 u_2 + u_2 u_3 + u_1 u_3 + \lambda(u_1 + u_2 + u_3 - 3)$$

$$\begin{cases} \frac{\partial L}{\partial u_1} = u_2 + u_3 + \lambda = 0 \\ \frac{\partial L}{\partial u_2} = u_1 + u_3 + \lambda = 0 \\ \frac{\partial L}{\partial u_3} = u_2 + u_1 + \lambda = 0 \\ \frac{\partial L}{\partial \lambda} = u_1 + u_2 + u_3 = 3 \end{cases} \implies \begin{cases} u_1^* = 1 \\ u_2^* = 1 \\ u_3^* = 1 \\ \lambda = -2 \end{cases} \quad \text{optimal solution}$$

Note: This was actually the worst we can do—maximize! Even weirder: no local minimizer!

1.4.1 Equality Constraints

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & g(u) \\ \text{s.t.} \quad & h(u) = \mathbf{0}, \quad h : \mathbb{R}^m \rightarrow \mathbb{R}^k \end{aligned}$$

Theorem. If u^* is a minimizer/maximizer then $\exists \lambda \in \mathbb{R}^k$ s.t.

$$\begin{aligned} \frac{\partial L}{\partial u}(u^*, \lambda) &= 0 \\ \frac{\partial L}{\partial \lambda}(u^*, \lambda) &= 0 \quad (\iff h(u^*) = 0) \end{aligned}$$

where $L(u, \lambda) = g(u) + \lambda^T h(u)$.

Example [Entropy Maximization]

Given $S = \{x_1, \dots, x_n\}$ and a distribution over S such that it takes the value x_j with probability p_j . The entropy is

$$E(p) = \sum_{j=1}^n (-p_j \ln p_j).$$

The mean is

$$m = \sum_{j=1}^n p_j x_j.$$

Problem: Given m , find p such that E is maximized.

$$\begin{aligned} \min_p \quad & - \sum_{j=1}^n p_j \ln p_j \\ \text{s.t.} \quad & \sum_{j=1}^n p_j x_j = m \\ & \sum_{j=1}^n p_j = 1 \\ & p_j \geq 0, \quad j = 1, \dots, n \quad (\text{ignore this...}) \end{aligned}$$

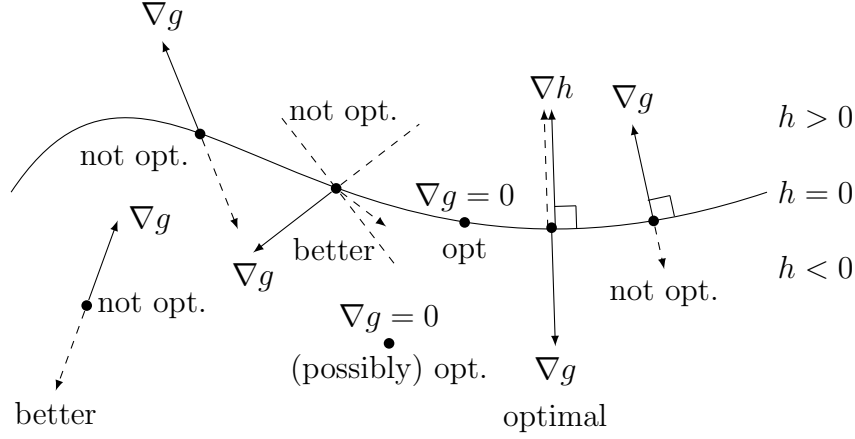
$$\begin{aligned} L &= - \sum p_j \ln p_j + \lambda_1 \left[\sum p_j x_j - m \right] + \lambda_2 \left[\sum p_j - 1 \right] \\ \frac{\partial L}{\partial p_j} &= - \ln p_j - 1 + \lambda_1 x_j + \lambda_2 = 0 \\ p_j &= e^{\lambda_2 - 1 + \lambda_1 x_j}, \quad j = 1, \dots, n \quad (p_j \geq 0 \text{ so we're ok with ignoring that}) \end{aligned}$$

$$\begin{aligned} \sum e^{\lambda_2 - 1 + \lambda_1 x_j} x_j &= m & n + 2 \text{ equations and} \\ \sum e^{\lambda_2 - 1 + \lambda_1 x_j} &= 1 & n + 2 \text{ unknowns...} \end{aligned}$$

No analytical solution, but numerically “solvable”

1.4.2 Inequality Constraints

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & g(u) \\ \text{s.t.} \quad & h(u) \leq \mathbf{0}, \quad h : \mathbb{R}^m \rightarrow \mathbb{R}^k \end{aligned}$$



We need:

- if $h(u^*) < 0$ then $\frac{\partial g}{\partial u}(u^*) = 0$
- if $h(u^*) = 0$ then we need either

$$\frac{\partial g}{\partial u}(u^*) = 0$$

or

$$\frac{\partial g}{\partial u}(u^*) = -\lambda \frac{\partial h}{\partial u}(u^*) \quad \text{for } \lambda > 0$$

Or, even better,

$$\frac{\partial}{\partial u} (g(u^*) + \lambda h(u^*)) = 0 \quad \text{for } \lambda \geq 0,$$

where $\lambda h(u^*) = 0$. ($h < 0 \rightarrow \lambda = 0$, $h = 0 \rightarrow \lambda \geq 0$)

In general, if $u \in \mathbb{R}^m$ and $h : \mathbb{R}^m \rightarrow \mathbb{R}^k$, we have that u^* , if optimal, has to satisfy

$$\begin{aligned} \frac{\partial}{\partial u} L(u^*, \lambda) &= 0 \\ h(u^*) &\leq \mathbf{0} \\ \lambda^T h(u^*) &= 0 \\ \lambda &\geq \mathbf{0} \end{aligned}$$

where the Lagrangian is $L(u, \lambda) = g(u) + \lambda^T h(u)$. Note that if we're maximizing, the same holds except we need $\lambda \leq 0$.

Example

$$\begin{aligned} \min \quad & 2u_1^2 + 2u_1u_2 + u_2^2 - 10u_1 - 10u_2 \\ \text{s.t.} \quad & \begin{cases} u_1^2 + u_2^2 \leq 5 \\ 3u_1 + u_2 \leq 6 \end{cases} \end{aligned}$$

$$L = 2u_1^2 + 2u_1u_2 + u_2^2 - 10u_1 - 10u_2 + \lambda_1(u_1^2 + u_2^2 - 5) + \lambda_2(3u_1 + u_2 - 6)$$

FONC:

- i) $\partial L / \partial u_1 = 4u_1 + 2u_2 - 10 + 2\lambda_1u_1 + 3\lambda_2$
- ii) $\partial L / \partial u_2 = 2u_1 + 2u_2 - 10 + 2\lambda_1u_2 + \lambda_2$
- iii) $u_1^2 + u_2^2 \leq 5$
- iv) $3u_1 + u_2 \leq 6$
- v) $\lambda_1(u_1^2 + u_2^2 - 5) = 0$
- vi) $\lambda_2(3u_1 + u_2 - 6) = 0$
- vii) $\lambda_1 \geq 0$
- viii) $\lambda_2 \geq 0$

To solve, assume different constraints are active/inactive:

1. Both constraints are inactive ($u_1^2 + u_2^2 < 5$, $3u_1 + u_2 < 6$) $\implies \lambda_1 = \lambda_2 = 0$

$$\begin{cases} 4u_1 + 2u_2 - 10 = 0 \\ 2u_1 + 2u_2 - 10 = 0 \end{cases} \implies \begin{cases} u_1 = 0 \\ u_2 = 5 \end{cases}$$

Note: iii) $0^2 + 5^2 \not\leq 5$

Not feasible

2. Assume constraint 1 is active and constraint 2 is inactive ($u_1^2 + u_2^2 = 5$, $\lambda_2 = 0$)

$$\begin{cases} 4u_1 + 2u_2 - 10 + 2\lambda_1u_1 = 0 \\ 2u_1 + 2u_2 - 10 + 2\lambda_1u_2 = 0 \\ u_1^2 + u_2^2 = 5 \end{cases} \implies \begin{cases} u_1 = 1 \\ u_2 = 2 \\ \lambda_1 = 1 \end{cases}$$

$$\checkmark \lambda_1 \geq 0$$

$$\checkmark 3 \cdot 1 + 2 \leq 6$$

This is a local minimizer

3. Assume constraint 2 is active and constraint 1 is inactive
4. Assume both constraints are active

Kuhn-Tucker Conditions (KKT conditions, Karush-Kuhn-Tucker)

Problem:

$$\begin{aligned} \min_{u \in \mathbb{R}^m} \quad & g(u) \\ \text{s.t.} \quad & \begin{cases} h_1(u) = 0, & h_1 : \mathbb{R}^m \rightarrow \mathbb{R}^p \\ h_2(u) \leq 0, & h_2 : \mathbb{R}^m \rightarrow \mathbb{R}^k \end{cases} \end{aligned} \quad (1.1)$$

Theorem. Let u^* be feasible ($h_1 = 0$, $h_2 \leq 0$). If u^* is a minimizer to (1.1) then there exists vectors $\lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^k$ with $\mu \geq \mathbf{0}$ such that

$$\begin{cases} \frac{\partial g}{\partial u}(u^*) + \lambda^T \frac{\partial h_1}{\partial u}(u^*) + \mu^T \frac{\partial h_2}{\partial u}(u^*) = 0 \\ \mu^T h_2(u^*) = 0 \end{cases}$$

Looking ahead: $\min \text{cost}(u(\cdot))$ s.t. $\dot{x} = f(x, u)$ (dynamics), where u is a function. Note the equality constraint

Question: How do we go from $u \in \mathbb{R}^m$ to $u \in \mathcal{U}$ (function space)?

Note: Function space is a set of functions of a given kind from a set X to a set Y

1. linear function
2. square-integrable functions: $L_2[0, T] : \int_0^T \|u(t)\|^2 dt < \infty$
3. $C^\infty(\mathbb{R})$

What would ∂ “cost”/ ∂u mean?

1.5 Directional Derivatives

Recall: To minimize $g(u)$, let u^* be a candidate minimizer and pitch a perturbation on u^*

$$\begin{aligned} g(u^* + \epsilon v) &= g(u^*) + \epsilon \frac{\partial g}{\partial u}(u^*)v + o(\epsilon) \\ \text{FONC: } \frac{\partial g}{\partial u}(u^*) &= 0 \end{aligned}$$

Note: $\frac{\partial g}{\partial u}(u^*)v$ tells us how much $g(u)$ increases/decreases in the direction of v .

Definition. The directional (Gateaux) derivative is given by

$$\delta g(u; v) = \lim_{\epsilon \rightarrow 0} \frac{g(u + \epsilon v) - g(u)}{\epsilon}$$

Example

$$g(u) = \frac{1}{2}u_1^2 - u_1 + 2u_2, \quad g : \mathbb{R}^2 \rightarrow \mathbb{R}$$

Let's consider $e_1 = [1 \ 0]^T$, $e_2 = [0 \ 1]^T$. What is $\delta g(u; e_i)$, $i = 1, 2$?

$$\begin{aligned} \delta g(u; v) &= \lim_{\epsilon \rightarrow 0} \frac{g(u + \epsilon v) - g(u)}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{g(u) + \epsilon \frac{\partial g}{\partial u}(u)v + o(\epsilon) - g(u)}{\epsilon} \\ &= \frac{\partial g}{\partial u}(u)v \end{aligned}$$

$$\begin{aligned} \frac{\partial g}{\partial u}(u) &= [u_1 - 1 \ 2] \\ \delta g(u; e_1) &= [u_1 - 1 \ 2]e_1 = u_1 - 1 \\ \delta g(u; e_2) &= [u_1 - 1 \ 2]e_2 = 2 \end{aligned}$$

But the beauty of directional derivatives is that they generalize beyond vectors, $u \in \mathbb{R}^m$, to function spaces (\mathcal{U}) or other “objects” like matrices.

Example $M \in \mathbb{R}^{n \times n}$, $F(M) = M^2$

What is $\frac{\partial F}{\partial M}$? (ponder at home...)

We can easily compute $\delta F(M; N)$!

$$\begin{aligned} F(M + \epsilon N) &= (M + \epsilon N)(M + \epsilon N) = M^2 + \epsilon MN + \epsilon NM + \epsilon^2 N^2 \\ \delta F(M; N) &= \lim_{\epsilon \rightarrow 0} \frac{F(M + \epsilon N) - F(M)}{\epsilon} \\ &= \lim_{\epsilon \rightarrow 0} \frac{\epsilon MN + \epsilon NM + \epsilon^2 N^2}{\epsilon} = MN + NM \end{aligned}$$

Infinite Dimensional Optimization Let $u \in \mathcal{U}$ (function space) and let $J(u)$ be the cost:

$$\min_{u \in \mathcal{U}} J(u)$$

Theorem. If $u^* \in \mathcal{U}$ is a (local) minimizer then

$$\delta J(u^*; v) = 0, \quad \forall v \in \mathcal{U}$$

Example Find minimizer u^* to

$$J(u) = \int_0^T L(u(t)) dt$$

$$\begin{aligned}
J(u + \epsilon v) - J(u) &= \int_0^T L(u(t) + \epsilon v(t)) \, dt - \int_0^T L(u(t)) \, dt, \quad u, v \in \mathcal{U} \\
&= \int_0^T \left[L(u(t)) + \epsilon \frac{\partial L}{\partial u}(u(t))v(t) + o(\epsilon) - L(u(t)) \right] \, dt \\
\delta J(u^*; v) &= \lim_{\epsilon \rightarrow 0} \frac{J(u + \epsilon v) - J(u)}{\epsilon} \\
&= \lim_{\epsilon \rightarrow 0} \frac{\int_0^T \epsilon \frac{\partial L}{\partial u}(u(t))v(t) \, dt + o(\epsilon)}{\epsilon} \\
&= \int_0^T \frac{\partial L}{\partial u}(u(t))v(t) \, dt
\end{aligned}$$

u^* optimizer:

$$\begin{aligned}
\delta J(u^*; v) &= \int_0^T \frac{\partial L}{\partial u}(u(t))v(t) \, dt = 0 \quad \forall v \in \mathcal{U} \\
&\Downarrow \\
\frac{\partial L}{\partial u}(u(t)) &= 0 \quad \forall t \in [0, T]
\end{aligned}$$

But, we want *optimal control*! We want our cost to look like

$$\begin{aligned}
&\int_0^T L(x(t), u(t)) \, dt \\
&\dot{x} = f(x, u)
\end{aligned}$$

1.6 Calculus of Variations

What happens to $x(t)$ when $u(t)$ changes to $u(t) + \epsilon v(t)$?

Instead of $\hat{x} = f(\hat{x}, u + \epsilon v)$, $\hat{x}(0) = x_0$, we can have

$$\tilde{x} = x + \epsilon \eta,$$

where

$$\begin{aligned}
\dot{x} &= f(x, u), & x(0) &= x_0 \\
\dot{\eta} &= \frac{\partial f}{\partial x}(x, u)\eta + \frac{\partial f}{\partial u}(x, u)v, & \eta(0) &= 0
\end{aligned}$$

Theorem. *If f is continuously differentiable in x and u then*

$$\hat{x}(t) = \tilde{x}(t) + o(\epsilon)$$

Proof. i) Initial conditions:

$$\begin{aligned}
\hat{x}(0) &= x_0 \\
\tilde{x}(0) &= x(0) + \epsilon \eta(0) = x_0
\end{aligned}$$

ii) Dynamics:

$$\begin{aligned}
\dot{\hat{x}} &= f(\hat{x}, u + \epsilon v) \\
\dot{\hat{x}} &= \dot{x} + \epsilon \dot{\eta} = f(x, u) + \epsilon \frac{\partial f}{\partial x}(x, u) \eta + \epsilon \frac{\partial f}{\partial x}(x, u) v \\
&= f(x + \epsilon \eta, u + \epsilon v) + o(\epsilon) \\
&= f(\tilde{x}, u + \epsilon v) + o(\epsilon)
\end{aligned}$$

We can see that $\hat{x}(t)$ and $\tilde{x}(t)$ have the same form.

Note: Taylor expansion with two elements is

$$\begin{aligned}
g(z_1 + \epsilon w_1, z_2 + \epsilon w_2) &= g(z_1, z_2 + \epsilon w_2) + \epsilon \frac{\partial g}{\partial z_1}(z_1, z_2 + \epsilon w_2) w_1 + o(\epsilon) \\
&= g(z_1, z_2) + \epsilon \frac{\partial g}{\partial z_2}(z_1, z_2) w_2 + \epsilon \frac{\partial g}{\partial z_1}(z_1, z_2) w_1 \\
&\quad + \epsilon \frac{\partial^2 g}{\partial z_2 \partial z_1}(z_1, z_2) w_1 w_2 + o(\epsilon) \\
&= g(z_1, z_2) + \epsilon \frac{\partial g}{\partial z_2}(z_1, z_2) w_2 + \epsilon \frac{\partial g}{\partial z_1}(z_1, z_2) w_1 + o(\epsilon)
\end{aligned}$$

□