Optimal Control
Lecture 3

Solmaz S. Kia
Mechanical and Aerospace Engineering Dept.
University of California Irvine
solmaz@uci.edu

Note: These slides only cover some parts of the lecture. For details and other discussions consult your class notes.

Reading suggestion: Chapters 1 and 2 of Ref [2] (see syllabus for references)

Parameter static optimization: when time is not a parameter in the problem

- Unconstrained optimization
- **Constrained optimization**

Some notation convention

- Let $F(x,u)$ be a real differentiable function taking values in $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.
- Let $f(x,u)$ be a real differentiable function taking values in $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$.

Then

$$F_x = \frac{\partial F}{\partial x} = \begin{bmatrix} \frac{\partial F}{\partial x_1} \\ \frac{\partial F}{\partial x_2} \\ \cdots \\ \frac{\partial F}{\partial x_n} \end{bmatrix}, \quad F_u = \frac{\partial F}{\partial u} = \begin{bmatrix} \frac{\partial F}{\partial u_1} \\ \frac{\partial F}{\partial u_2} \\ \cdots \\ \frac{\partial F}{\partial u_m} \end{bmatrix},$$

$$f_x = \frac{\partial f}{\partial x} = \begin{bmatrix} \frac{\partial f^1}{\partial x} & \frac{\partial f^2}{\partial x} & \cdots & \frac{\partial f^p}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial f^1}{\partial x_1} & \frac{\partial f^2}{\partial x_1} & \cdots & \frac{\partial f^p}{\partial x_1} \\ \frac{\partial f^1}{\partial x_2} & \frac{\partial f^2}{\partial x_2} & \cdots & \frac{\partial f^p}{\partial x_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f^1}{\partial x_n} & \frac{\partial f^2}{\partial x_n} & \cdots & \frac{\partial f^p}{\partial x_n} \end{bmatrix},$$

$$f_u = \frac{\partial f}{\partial u} = \begin{bmatrix} \frac{\partial f^1}{\partial u} & \frac{\partial f^2}{\partial u} & \cdots & \frac{\partial f^p}{\partial u} \end{bmatrix} = \begin{bmatrix} \frac{\partial f^1}{\partial u_1} & \frac{\partial f^2}{\partial u_1} & \cdots & \frac{\partial f^p}{\partial u_1} \\ \frac{\partial f^1}{\partial u_2} & \frac{\partial f^2}{\partial u_2} & \cdots & \frac{\partial f^p}{\partial u_2} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial f^1}{\partial u_m} & \frac{\partial f^2}{\partial u_m} & \cdots & \frac{\partial f^p}{\partial u_m} \end{bmatrix},$$

Constrained optimization

$$x^\star = \underset{x\in\mathbb{R}^n}{\text{argmin}}\ f(x) \quad \text{s.t.}$$
$$h_i(x) = 0, \quad i \in \{1, \cdots, m\}$$
$$g_i(x) \leqslant 0, \quad i \in \{1, \cdots, r\}$$

or

$$x^\star = \underset{x\in\mathbb{R}^n}{\text{argmin}}\ f(x) \quad \text{s.t.}$$
$$h(x) = 0,$$
$$g(x) \leqslant 0,$$

f, h, g: continuously differentiable function of x
e.g., $f, h, g \in C^1$ continuously differentiable
e.g., $f, h, g \in C^2$ both f and its first derivative are continuously differentiable

First Order Necessary Condition for Optimality: $x^\star$ is a local minimizer then

$$\nabla f(x^\star)^\top \Delta x \geqslant 0, \quad \text{for} \ \ \Delta x \in V(x^\star)$$

- Set of first order feasible variations at x

$$V(x) = \{d \in \mathbb{R}^n \mid \nabla h_i(x)^\top d = 0,\ \nabla g_j(x)^\top d \leqslant 0, \quad j \in A(x^\star)\}$$

- Active inequality constraints at x

$$A(x) = \{j \in \{1, \cdots, r\} \mid g_j(x) = 0\}$$

A feasible vector x is said to be regular of the equality constraint gradients $\nabla h_i(x)$, $i = 1, \cdots, m$, and the active inequality constraint gradients $\nabla g_j(x)$, $j \in A(x)$, are linearly independent.

Necessary Conditions for Optimality: equality and inequality conditions

$$
\begin{aligned}
x^\star = & \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \\
& h_i(x) = 0, \quad i \in \{1, \cdots, m\} \\
& g_j(x) \leqslant 0, \quad j \in \{1, \cdots, r\}
\end{aligned}
$$

or

$$
\begin{aligned}
x^\star = & \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \\
& h(x) = 0, \\
& g(x) \leqslant 0,
\end{aligned}
$$

- A simple approach relies on the theory for equality constraints:
  - Inactive constraints at $x^\star$ do not matter, they can be ignored in the statement of optimality conditions
  - Active inequality constraints can be treated to a large extent as equality constraints

$x^\star$ is also a local minimum of

$$
\begin{aligned}
x^\star = & \operatorname*{argmin}_{x\in\mathbb{R}^n} f(x) \quad \text{s.t.} \\
& h_i(x) = 0, \quad i \in \{1, \cdots, m\} \\
& g_j(x) = 0, \quad \forall j \in A(x^\star)
\end{aligned}
$$

If $x^\star$ is regular for this equivalent optimization problem, then there exists Lagrange multipliers $\lambda_1^\star, \cdots, \lambda_{,}^\star$, and $\mu_j^\star$, $j \in A(x^\star)$:

$$
\nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla h_i(x^\star) + \sum_{j\in A(x^\star)} \mu_j^\star \nabla g_j(x^\star) = 0.
$$

But we need to require that $\mu_j^\star \geqslant 0$ for $j \in A(x^\star)$.

This approach is limited by regularity condition!

Necessary Conditions for Optimality: equality and inequality conditions

Lagrangian function $L : \mathbb{R}^{n+m} \mapsto \mathbb{R}$: $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{r} \mu_j g_j(x)$

Proposition (Karush-Huhn-Tucker Necessary conditions)

Let $x^\star$ be a local minimum of

$$x^\star = \underset{x \in \mathbb{R}^n}{\text{argmin}}\ f(x) \quad \text{s.t.}$$
$$h_1(x) = 0, \cdots, h_m(x) = 0$$
$$g_1(x) \leqslant 0, \cdots, g_r(x) \leqslant 0$$

where f, h_i and g_j are continuously differentiable functions from $\mathbb{R}^n$ to $\mathbb{R}$. Assume the $x^\star$ is regular. Then there exists unique Lagrange multiplier vectors $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_m^\star)$ and $\mu^\star = (\mu_1^\star, \cdots, \mu_r^\star)$, s.t.

$$\nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0$$
$$\mu_j^\star \geqslant 0, \quad j = 1, \cdots, r$$
$$\mu_j^\star = 0, \quad \forall j \notin \underbrace{A(x^\star)}_{\text{active constraint set}} \quad .$$

If in addition f g and h are twice continuously differentiable we have

$$y^\top \nabla_{xx} L(x^\star, \lambda^\star, \mu^\star)\, y \geqslant 0,$$

for all
$y \in V(x^\star) = \{y \in \mathbb{R}^n | \nabla h_i(x^\star)^\top y = 0, \quad \forall i = 1, \cdots, m, \quad \nabla g_j(x^\star)^\top y = 0, \quad j \in A(x^\star)\}.$

Sufficiency Conditions for Optimality

Lagrangian function $L : \mathbb{R}^{n+m} \mapsto \mathbb{R}$: $L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) + \sum_{i=1}^{r} \mu_j g_j(x)$

Second Order Sufficiency Conditions

Assume that f, h_i and g_j are twice continuously differentiable f, and let $x^\star \in \mathbb{R}^n$, $\lambda^\star = (\lambda_1^\star, \cdots, \lambda_m^\star)$ and $\mu^\star = (\mu_1^\star, \cdots, \mu_r^\star)$ satisfy

$$\begin{aligned}
&\nabla_x L(x^\star, \lambda^\star, \mu^\star) = 0, \quad h(x^\star) = 0_m, \\
&\mu_j^\star \geqslant 0, \quad j = 1, \cdots, r, \\
&\mu_j^\star = 0, \quad \forall j \notin A(x^\star), \\
&y^\top \nabla_{xx} L(x^\star, \lambda^\star, \mu^\star) y > 0,
\end{aligned}$$

for all $y \in \mathbb{R}^n$ such that $\nabla h_i(x^\star)^\top y = 0, \ \forall i = 1, \cdots, m, \quad \nabla g_j(x^\star)^\top y = 0, \ j \in A(x^\star)$.
Assume also that

$$\mu_j^\star > 0, \qquad \forall j \in A(x^\star).$$

Then $x^\star$ is a strict local minimum of

$$\begin{aligned}
&\min_{x \in \mathbb{R}^n} \ f(x) \quad \text{s.t.} \\
&h_1(x) = 0, \cdots, h_m(x) = 0 \\
&g_1(x) \leqslant 0, \cdots, g_r(x) \leqslant 0
\end{aligned}$$

Solution approach

One approach for using necessary conditions to solve inequality constrained problems is to consider separately all the possible combinations of constraints being active or inactive.

Constrained optimization: numerical example

$$\text{minimize } f(x) = x_1 + x_2 \text{ subject to}$$
$$g(x) = (x_1 - 1)^2 + x_2^2 - 1 \leqslant 0$$

- H1: Constraint is active. To validate H1, we should have $\mu \geqslant 0$.

$$L(x, \mu) = x_1 + x_2 + \mu(x_1 - 1)^2 + x_2^2 \leqslant 1$$

FONC:

$$\left.\begin{array}{l} \nabla_{x_1} L(x, \mu) = 1 + 2\mu(x_1 - 1) = 0 \\ \nabla_{x_2} L(x, \mu) = 1 + 2\mu(x_2) = 0 \\ \nabla_{\mu} L(x, \mu) = (x_1 - 1)^2 + x_2^2 - 1 = 0 \end{array}\right\} \Rightarrow$$

$$\begin{cases} x_1 = 1, x_2 = 1, \mu = -\frac{1}{2} & \text{since } \mu < 0 \text{ this solution is not acceptable} \\ x_1^\star = 1, x_2^\star = -1, \mu^\star = \frac{1}{2} & \text{since } \mu^\star > 0 \text{ this solution is a candidate for local minimizer} \end{cases}$$

SONC:

$$y \nabla_{xx} L(x^\star, \mu^\star) y \geqslant 0 \text{ for } y \in V(x^\star) = \{y \in \mathbb{R}^2 | \nabla g(x^\star)^\top y = 0\} = \{y \in \mathbb{R}^2 | \begin{bmatrix} 0 & -2 \end{bmatrix} y = 0\}$$

Since $\nabla_{xx} L(x^\star, \mu^\star) = \begin{bmatrix} 2\mu^\star & 0 \\ 0 & 2\mu^\star \end{bmatrix} > 0$ ($\mu^\star = \frac{1}{2}$), then SONC condition is definitely satisfied. Also since the condition holds for strict > 0, then the second order sufficiency condition is satisfied and $x_1^\star = 1, x_2^\star = -1$ is a local minimizer.

- H2: Constraint is not active. To validate H2, we should check that the identified stationary points $x^\star$ satisfy $g(x^\star) < 0$.

$$\left.\begin{array}{l} \nabla_{x_1} f(x) = 1 = 0 \\ \nabla_{x_2} f(x) = 1 = 0 \end{array}\right\} \Rightarrow \text{there is no solution in this case}$$

Constrained optimization: numerical example

$$\text{minimize } f(x) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 \text{ subject to}$$
$$g_1(x) = x_1^2 + x_2^2 - 5 \leqslant 0$$
$$g_2(x) = 3x_1 + x_2 - 6 \leqslant 0$$

$$\nabla_x f(x) = \begin{bmatrix} 4x_1 + 2x_2 - 10 \\ 2x_1 + 2x_2 - 10 \end{bmatrix}, \quad \nabla_x g_1(x) = \begin{bmatrix} 2x_1 \\ 3 \end{bmatrix}, \quad \nabla_x g_2(x) = \begin{bmatrix} 2x_2 \\ 1 \end{bmatrix}$$

- H1: both constraints are inactive: $g_1 < 0$, $g_2 < 0$ and $\mu_1 = \mu_2 = 0$.
  FONC:

$$\left.\begin{array}{l} \nabla_{x_1} f(x) = 4x_1 + 2x_2 - 10 = 0 \\ \nabla_{x_2} f(x) = 2x_1 + 2x_2 - 10 = 0 \end{array}\right\} \Rightarrow \; x_1 = 0, x_2 = 5$$

  $g_1(x_1 = 0, x_2 = 5) = 20 > 0$ and $g_2(x_1 = 0, x_2 = -1 < 0$. Since H1 is not correct, this case is not possible.

- H2: both constraints are active: $g_1 = 0$, $g_2 = 0$ and $\mu_1, \mu_2 \geqslant 0$.

$$L(x, \mu) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 + \mu_1(x_1^2 + x_2^2 - 5) + \mu_2(3x_1 + x_2 - 6)$$

  FONC:

$$\left.\begin{array}{l} \nabla_{x_1} L(x, \mu) = 4x_1 + 2x_2 - 10 + 2\mu_1 x_1 + 3\mu_2 = 0 \\ \nabla_{x_2} L(x, \mu) = 2x_1 + 2x_2 - 10 + 2\mu_2 x_2 + \mu_2 = 0 \\ \nabla_{\mu_1} L(x, \mu) = x_1^2 + x_2^2 - 5 = 0 \\ \nabla_{\mu_1} L(x, \mu) = 3x_1 + x_2 - 6 = 0 \end{array}\right\} \Rightarrow$$

$$\left\{\begin{array}{ll} x = \begin{bmatrix} 2.1742 \\ -0.5225 \end{bmatrix}, \mu = \begin{bmatrix} -2.37 \\ 4.22 \end{bmatrix} & \text{since } \mu_1 < 0 \text{ this solution is not acceptable.} \\ x = \begin{bmatrix} 1.4258 \\ 1.7228 \end{bmatrix}, \mu = \begin{bmatrix} 1.37 \\ -1.02 \end{bmatrix} & \text{since } \mu_2 < 0 \text{ this solution is not acceptable.} \end{array}\right.$$

Constrained optimization: numerical example

- H3: g_1 is inactive ($g_1 < 0,\ \mu_1 = 0$), and g_2 is active ($\mu_2 \geqslant 0$).

$$L(x,\mu) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 + \mu_2(3x_1 + x_2 - 6)$$

FONC:

$$\left.\begin{array}{l} \nabla_{x_1} L(x,\mu) = 4x_1 + 2x_2 - 10 + 3\mu_2 = 0 \\ \nabla_{x_2} L(x,\mu) = 2x_1 + 2x_2 - 10 + \mu_2 = 0 \\ \nabla_{\mu_1} L(x,\mu) = 3x_1 + x_2 - 6 = 0 \end{array}\right\} \Rightarrow x = \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix},\ \mu_2 = -0.4.$$

since $\mu_2 < 0$ this solution is not acceptable.

- H4: g_2 is inactive ($g_2 < 0,\ \mu_2 = 0$), and g_1 is inactive ($\mu_1 \geqslant 0$).

$$L(x,\mu) = 2x_1^2 + 2x_1x_2 + x_2^2 - 10x_1 - 10x_2 + \mu_1(x_1^2 + x_2^2 - 5)$$

FONC:

$$\left.\begin{array}{l} \nabla_{x_1} L(x,\mu) = 4x_1 + 2x_2 - 10 + 2\mu_1 x_1 = 0 \\ \nabla_{x_2} L(x,\mu) = 2x_1 + 2x_2 - 10 + 2\mu_1 x_2 = 0 \\ \nabla_{\mu_1} L(x,\mu) = x_1^2 + x_2^2 - 5 = 0 \end{array}\right\} \Rightarrow x^\star = \begin{bmatrix} 1 \\ 2 \end{bmatrix},\ \mu_1^\star = 1.$$

since $\mu_1 \geqslant 0$ this solution is qualified as KKT solution.

Now we need to validate H4: $g_2(x_1 = 1, x_2 = 2) = -1 < 0$, therefore H4 is correct.
SONC:

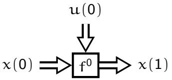
$$y \nabla_{xx} L(x^\star, \mu^\star) y \geqslant 0 \text{ for } y \in V(x^\star) = \{y \in \mathbb{R}^2 | \nabla g_1(x^\star)^\top y = 0\} = \{y \in \mathbb{R}^2 | \begin{bmatrix} 2 & 4 \end{bmatrix} y = 0\}$$

Since $\nabla_{xx} L(x^\star, \mu^\star) = \begin{bmatrix} 4 + 2\mu_1^\star & 2 \\ 2 & 2 + 2\mu_1^\star \end{bmatrix} > 0$ ($\mu^\star = 1$), then SONC condition is definitely satisfied. Also since the condition holds for strict > 0, then the second order sufficiency condition is satisfied and $x_1^\star = 1, x_2^\star = 2$ is a local minimizer.

Optimal control and its connection to constrained optimization

Optimal Control Example

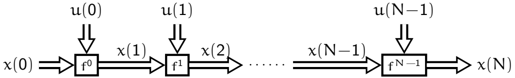
Single stage system



$$u^\star(0) = \text{argmin}\, \underbrace{\phi(x(1)) + L^0(x(0), u(0))}_{J(u(0))}$$

$$\text{s.t.} \quad x(1) = f^0(x(0), u(0)),$$
$$x(0) = x_0 \in \mathbb{R}^n.$$

Multi stage system



$$u^\star = \text{argmin}\, \underbrace{\phi(x(N)) + \sum_{k=0}^{N-1} L^i(x(i), u(i))}_{J(u(0),\cdots,u(N-1))} \quad \text{s.t.}$$

$$x(N) = f^{N-1}(x(N-1), u(N-1)),$$
$$\vdots$$
$$x(1) = f^0(x(0), u(0)),$$
$$x(0) = x_0 \in \mathbb{R}^n.$$

The problems above are in the general form of:

$$u^\star = \underset{u \in \mathbb{R}^m}{\text{argmin}}\ F(x, u), \quad \text{s.t.,}$$
$$f(x, u) = 0$$

where $F : \mathbb{R}^{m+n} \to \mathbb{R}$ and $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ are differentiable.

Constrained optimization

$$u^\star = \underset{u \in \mathbb{R}^m}{\text{argmin}}\ F(x, u), \quad s.t.,$$
$$f(x, u) = 0$$

where $F : \mathbb{R}^{m+n} \to \mathbb{R}$ and $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ are differentiable.

Trivial solution: solve via direct substitution, i.e.,

1. find x in terms of u from $f(x, u) = 0$,
2. substitute in $F(x, u)$ to eliminate x and obtain an unconstrained optimization problem in terms of u.

Works best for simple linear $f's$ (assumption is that not both of f and F are linear)

Constrained optimization

$$u^\star = \underset{u \in \mathbb{R}^m}{\text{argmin}}\ F(x, u), \quad \text{s.t.},$$

$$f(x, u) = 0$$

where $F : \mathbb{R}^{m+n} \to \mathbb{R}$ and $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ are differentiable.

Feasible set: $\mathcal{S}_{\text{feas}} = \{(x, u) \in \mathbb{R}^n \times \mathbb{R}^m \,|\, f(x, u) = 0\}$.

To obtain a criterion for optimality (in minimization sense):

- start with $(x, u) \in \mathcal{S}_{\text{feas}}$ as candidate for local minimum,
- we investigate how dF changes for points in $\mathcal{S}_{\text{feas}}$ in close neighborhood of (x, u)

First-order analysis:

$$f(x + dx, u + du) \approx f(x, u) + f_u^\top du + f_x^\top dx,$$

$$F(x + dx, u + du) \approx F(x, u) + \underbrace{F_u^\top du + F_x^\top dx}_{dF},$$

First-order necessary condition for a point to be a minimizer:

$$\begin{cases} f_u^\top du + f_x^\top dx = 0, \\ F_x^\top dx + F_u^\top du \geqslant 0. \end{cases} \Rightarrow \begin{cases} \text{let } du \text{ vary freely, but } dx = -(f_x)^{-\top}(f_u)^\top du, \\ 1^{st} \text{ order Neces. condition} : F_u - f_u f_x^{-1} F_x = 0. \end{cases}$$

Critical point: $dF = 0$ for neighboring points $(x + dx, u + du) \in \mathcal{S}_{\text{feas}}$

$$\begin{cases} f_u^\top du + f_x^\top dx = 0, \\ F_x^\top dx + F_u^\top du = 0. \end{cases} \Rightarrow \begin{cases} \text{Neces. and sufficient condition for a point to be critical point} : \\ \quad F_u - f_u f_x^{-1} F_x = 0. \end{cases}$$

Constrained optimization: method of Lagrange multipliers

$$u^\star = \underset{u \in \mathbb{R}^m}{\text{argmin}}\ F(x,u), \quad s.t.,$$
$$f(x,u) = 0$$

where $F : \mathbb{R}^{m+n} \to \mathbb{R}$ and $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ are differentiable.

- to determine neighboring points, dx and du are not independent
- Lagrange multiplier $\lambda = [\lambda_1, \cdots, \lambda_n]^\top \in \mathbb{R}^n$ captures this dependency

$$H(x,u,\lambda) = F(x,u) + \lambda^\top f(x,u),$$
$$f(x,u) = 0.$$

First-order analysis

Necessary and sufficient cond. for **critical point**

$$\begin{cases} dx = -(f_x)^{-\top}(f_u)^\top du, \\ \lambda = -(f_x)^{-1} F_x. \end{cases}$$

$$\begin{cases} \frac{\partial H}{\partial x} = F_x + f_x\lambda = 0, \\ \frac{\partial H}{\partial u} = F_u + f_u\lambda =, \\ \frac{\partial H}{\partial \lambda} = 0 \Rightarrow f(x,u) = 0. \end{cases}$$

Example

$$\min \; F(x,u) = x_1^2 + x_2^2 + x_3^2 + 2u_1 u_2 + u_2^2 + u_1^2$$

$$\begin{cases} f^1(x,u) = x_1 - x_2 + u_1 = 0, \\ f^1(x,u) = x_2 + 2\,x_3 - u_2 + 1 = 0, \\ f^2(x,u) = x_2 + x_3 + x_1 = 0. \end{cases}$$

$$f_x = \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 0 & 2 & 1 \end{bmatrix}, \quad f_u = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}, \; F_x = \begin{bmatrix} 2\,x_1 \\ 2\,x_2 \\ 2\,x_3 \end{bmatrix}, \quad F_u = \begin{bmatrix} 2\,u_1 + 2\,u_2 \\ 2\,u_2 + 2\,u_1 \end{bmatrix}$$

$$H = F(x,u) + \lambda^\top f(x,u) = F(x,u) + \lambda_1 f^1(x,u) + \lambda_2 f^2(x,u) + \lambda_3 f^3(x,u)$$

$------$

$$H_u = \begin{bmatrix} \frac{\partial F(x,u)}{\partial u_1} + \lambda_1 \frac{\partial f^1(x,u)}{\partial u_1} + \lambda_2 \frac{\partial f^2(x,u)}{\partial u_1} + \lambda_3 \frac{\partial f^3(x,u)}{\partial u_1} \\ \frac{\partial F(x,u)}{\partial u_2} + \lambda_1 \frac{\partial f^1(x,u)}{\partial u_2} + \lambda_2 \frac{\partial f^2(x,u)}{\partial u_2} + \lambda_3 \frac{\partial f^3(x,u)}{\partial u_2} \end{bmatrix}$$

$$= F_u + f_u \lambda = \begin{bmatrix} 2\,u_1 + 2\,u_2 \\ 2\,u_2 + 2\,u_1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix} \lambda = \begin{bmatrix} 2\,u_1 + 2\,u_2 + \lambda_1 \\ 2\,u_2 + 2\,u_1 - \lambda_2 \end{bmatrix}$$

$------$

$$H_x = F_x + f_x \lambda = \begin{bmatrix} 2\,x_1 \\ 2\,x_2 \\ 2\,x_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 \\ -1 & 1 & 1 \\ 0 & 2 & 1 \end{bmatrix} \lambda = \begin{bmatrix} 2\,x_1 + \lambda_1 + \lambda_3 \\ 2\,x_2 - \lambda_1 + \lambda_2 + \lambda_3 \\ 2\,x_3 + 2\lambda_2 + \lambda_3 \end{bmatrix}$$

Example

$$H_u = 0 \Rightarrow \begin{cases} 2\,u_1 + 2\,u_2 + \lambda_1 = 0 \\ 2\,u_2 + 2\,u_1 - \lambda_2 = 0 \end{cases}$$

$$H_x = 0 \Rightarrow \begin{cases} 2\,x_1 + \lambda_1 + \lambda_3 = 0 \\ 2\,x_2 - \lambda_1 + \lambda_2 + \lambda_3 = 0 \\ 2\,x_3 + 2\lambda_2 + \lambda_3 = 0 \end{cases}$$

$$f(x, u) = 0 \Rightarrow \begin{cases} x_1 - x_2 + u_1 = 0, \\ x_2 + 2\,x_3 - u_2 + 1 = 0, \\ x_2 + x_3 + x_1 = 0. \end{cases}$$

7 equations, 7 unknowns: can be solved to find critical point(s)

$$\begin{aligned} \lambda_1^\star &= 2, \\ \lambda_2^\star &= -2, \\ \lambda_3^\star &= 2, \\ x_1^\star &= -2, \\ x_2^\star &= 1, \\ x_3^\star &= 1, \\ u_1^\star &= 3, \\ u_2^\star &= -4. \end{aligned}$$

Constrained optimization: second order necessary and sufficient conditions for optimality

$$u^\star = \operatorname*{argmin}_{u \in \mathbb{R}^m} F(x,u), \quad \text{s.t.}, \qquad f(x,u) = 0$$

$$u^\star = \operatorname*{argmin}_{u \in \mathbb{R}^m} H(x,u) = F(x,u) + \lambda^\top f(x,u), \quad \text{s.t.}, \qquad f(x,u) = 0$$

Second-order analysis around critical point

Necessary and sufficient cond. for $(x^\star, u^\star)$ to be a **critical point**

$$\begin{cases} \frac{\partial H}{\partial x} = F_x + f_x \lambda = 0, \quad \frac{\partial H}{\partial u} = F_u + f_u \lambda = 0, \ @(x^\star, u^\star) \\ \frac{\partial H}{\partial \lambda} = 0 \Rightarrow f(x^\star, u^\star) = 0. \end{cases}$$

$(x^\star + dx, u^\star + du) \in \mathcal{S}_{\textbf{feas}}$: $dx = -(f_x)^{-\top} f_u^\top du$.

$$H(x^\star + dx, u^\star + du) =$$

$$H(x^\star, u^\star) + H_u^\top du + H_x^\top dx + \frac{1}{2} dx^\top H_{xx} dx + dx^\top H_{xu} du + \frac{1}{2} du^\top H_{uu} du + O(3) =$$

$$H(x^\star, u^\star) + \frac{1}{2} \begin{bmatrix} dx^\top & du^\top \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} dx \\ du \end{bmatrix} + O(3) =$$

$$H(x^\star, u^\star) + \frac{1}{2} du^\top \begin{bmatrix} -f_u (f_x)^{-1} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -(f_x)^{-\top} f_u^\top \\ I \end{bmatrix} du + O(3)$$

Constrained optimization: second order necessary and sufficient conditions for optimality

- 2^{nd} order necessary cond.

$$du^\top \begin{bmatrix} -f_u (f_x)^{-1} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -(f_x)^{-\top} f_u^\top \\ I \end{bmatrix} du \geqslant 0$$

$$\Leftrightarrow$$

$$\begin{bmatrix} -f_u (f_x)^{-1} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -(f_x)^{-\top} f_u^\top \\ I \end{bmatrix} \geqslant 0, \ @(x^\star, u^\star) \ \text{positive semi-definite matrix}$$

- 2^{nd} order necessary cond.

$$du^\top \begin{bmatrix} -f_u (f_x)^{-1} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -(f_x)^{-\top} f_u^\top \\ I \end{bmatrix} du > 0, \quad du \neq 0$$

$$\Leftrightarrow$$

$$\begin{bmatrix} -f_u (f_x)^{-1} & I \end{bmatrix} \begin{bmatrix} H_{xx} & H_{xu} \\ H_{ux} & H_{uu} \end{bmatrix} \begin{bmatrix} -(f_x)^{-\top} f_u^\top \\ I \end{bmatrix} > 0, \ @(x^\star, u^\star) \ \text{positive definite matrix}$$

Constrained optimization: an iterative solver

$$u^\star = \underset{u \in \mathbb{R}^m}{\operatorname{argmin}}\ F(x,u), \quad \text{s.t.,} \qquad f(x,u) = 0$$

$$u^\star = \underset{u \in \mathbb{R}^m}{\operatorname{argmin}}\ H(x,u) = F(x,u) + \lambda^\top f(x,u), \quad \text{s.t.,} \qquad f(x,u) = 0$$

$$\begin{cases} dx = -(f_x)^{-\top}(f_u)^\top du, \\ \lambda^\top = -(f_x)^{-1}F_x, \\ \begin{cases} dF = F_x^\top dx + F_u^\top du \\ \quad = (-f_u\,(f_x)^{-1}F_x + F_u)^\top du \\ \quad = (F_u + f_u\lambda)^\top du \\ \quad = H_u^\top du \end{cases} \end{cases}$$

Necessary and sufficient cond. for **critical point**

$$\begin{cases} \frac{\partial H}{\partial x} = F_x + f_x\lambda = 0, \\ \frac{\partial H}{\partial u} = F_u + f_u\lambda = 0, \\ \frac{\partial H}{\partial \lambda} = 0 \Rightarrow f(x,u) = 0. \end{cases}$$

1. Select initial $u(k)$, $k = 0$
2. Determine $x(k)$ from $f(x(k), u(k)) = 0$
3. Determine $\lambda = -(f_x)^{-1} F_x$
4. Determine $\frac{\partial H}{\partial u} = F_u + f_u\lambda$
5. Determine $u(k+1) = u(k) - \alpha H_u$
6. Determine predicted change in value of F

$$\Delta F = \Delta H = H_u^\top \Delta u = -\alpha H_u^\top H_u.$$

If $\Delta F = \Delta H$ is sufficiently small, stop.
Otherwise go to step 2.

Constrained optimization: fmincon

Problem:

$$\text{minimize } F(x,u) = x^2 + u^2, \ .s.t.,$$
$$x + u + 2 = 0$$

Code for fmincon:

- a function to list equality constraints

```
function [c,ceq] = EqualFun(x)
ceq = x(1) + x(2) +2;
c = [];
```

- the main code

```
nonlcon = @EqualFun;
A = [];
b = [];
Aeq = [];
beq = [];
lb = [];
ub = [];
x0 = [0,0];
[x,fval,exitflag,output] = fmincon(fun,x0,A,b,Aeq,beq,lb,ub,nonlcon)
```

Optimal control and its connection to constrained optimization

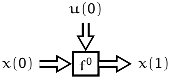
$$u^\star = \underset{u \in \mathbb{R}^m}{\text{argmin}}\ F(x, u), \quad \text{s.t.},$$

$$f(x, u) = 0$$

where $F : \mathbb{R}^{m+n} \to \mathbb{R}$ and $f : \mathbb{R}^{m+n} \to \mathbb{R}^n$ are differentiable.

Optimal Control Example

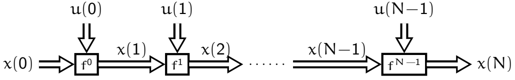
Single stage system



$$u^\star(0) = \text{argmin}\ \underbrace{\phi(x(1)) + L^0(x(0), u(0))}_{J(u(0))}$$

$$\text{s.t.} \quad x(1) = f^0(x(0), u(0)),$$
$$x(0) = x_0 \in \mathbb{R}^n.$$

Multi stage system



$$u^\star = \text{argmin}\ \underbrace{\phi(x(N)) + \sum_{k=0}^{N-1} L^i(x(i), u(i))}_{J(u(0), \cdots, u(N-1))} \quad \text{s.t.}$$

$$x(N) = f^{N-1}(x(N-1), u(N-1)),$$
$$\vdots$$
$$x(1) = f^0(x(0), u(0)),$$
$$x(0) = x_0 \in \mathbb{R}^n.$$

First order optimality condition for single stage optimal control

$$u(0)^\star = \underset{u(0)\in\mathbb{R}^m}{\text{argmin}} \; J(x(1), u(0)) = \phi(x(1)) + L^0(x(0), u(0)), \quad \text{s.t.},$$
$$x(1) = f^0(x(0), u(0)), \quad x(1) \in \mathbb{R}^n, \; u(0) \in \mathbb{R}^m,$$
$$x(0) = x_0 \in \mathbb{R}^n, \quad \text{(given initial condition)}.$$

- $\bar{J} = J + \lambda(1)^\top(f^0(x(0), u(0)) - x(1)) = \phi(x(1)) + L^0(x(0), u(0)) + \lambda(1)^\top(f^0(x(0), u(0)) - x(1))$
- Let $H^0(x(0), u(0)), \lambda(1)) = L^0(x(0), u(0)) + \lambda(1)^\top(f^0(x(0), u(0)))$.
- Then, we can rewrite $\bar{J}$ as $\bar{J} = (\phi(x(1)) - \lambda(1)^\top x(1)) + H^0(x(0), u(0), \lambda(1))$.

First order analysis:

$$\bar{J}(x(1) + dx(1), u(0) + du(0)) = \bar{J}(x(1), u(0)) + \underbrace{\left(\frac{\partial \phi(x(1))}{\partial x(1)} - \lambda(1)\right)^\top dx(1) + \left(\frac{\partial H^0}{\partial x(0)}\right)^\top dx(0) + \left(\frac{\partial H^0}{\partial u(0)}\right)^\top du(0)}_{d\bar{J}}$$

Here $dx(0) = 0$ because the initial condition is given (no need for variation). Think of $du(0)$ as free variable and $dx(1)$ the dependent variable, which is defined from the constraint equation (constraint equation relates $dx(1)$ to $du(0)$). Next, pick $\lambda(1)$ such that

$$\frac{\partial \phi(x(1))}{\partial x(1)} - \lambda(1) = 0,$$

which gives us $\bar{J}(x(1) + dx(1), u(0) + du(0)) = \bar{J}(x(1), u(0)) + \underbrace{\left(\frac{\partial H^0}{\partial u(0)}\right)^\top du(0)}_{d\bar{J}}$.

First order optimality condition for single stage optimal control (cont'd)

- For $(x(1), u(0))$ to be a minimum point we need $d\bar{J} = (\frac{\partial H^0}{\partial u(0)})^\top du(0) \geqslant 0$. Because we are free to vary $du(0)$ in all directions, then the necessary condition for $(x(1), u(0))$ to be a minimum point is $\frac{\partial H^0}{\partial u(0)} = 0$.

- To summarize:
  First order necessary condition for $(x(1), u(0))$ to be a minimum point:

$$\begin{cases} \lambda(1) = \frac{\partial \phi(x(1))}{\partial x(1)}, & n \text{ eq} \\ \frac{\partial H^0}{\partial u(0)} = 0, & m \text{ eq} \\ x(1) = f^0(x(0), u(0)), & n \text{ eq}^1 \end{cases}$$

  Here, we have $2n + m$ equation for $2n + m$ unknowns (the unknowns in the set of equations above are $\lambda(1) \in \mathbb{R}^n$, $x(1) \in \mathbb{R}^n$ and $u(0) \in \mathbb{R}^m$).

[1] which can also be written as $x(1) = \frac{\partial H^0}{\partial \lambda(1)}$

The following material will be discussed in the next lecture.

First order optimality condition for multi-stage optimal control

$$(u^\star(0), \cdots, u^\star(N-1)) = \underset{(u(0)\in\mathbb{R}^m, \cdots, u(N-1)\in\mathbb{R}^m}{\operatorname{argmin}} \quad J = \phi(x(N)) + \sum\nolimits_{i=0}^{N-1} L^i(x(i), u(i)), \quad \text{s.t.},$$

$$x(1) = f^0(x(0), u(0)), \quad x(1) \in \mathbb{R}^n, \ u(0) \in \mathbb{R}^m,$$

$$\vdots$$

$$x(N) = f^{N-1}(x(N-1), u(N-1)), \quad x(N) \in \mathbb{R}^n, \ u(N-1) \in \mathbb{R}^m,$$

$$x(0) = x_0 \in \mathbb{R}^n, \quad \text{(given initial condition)}.$$

- $\bar{J} = J + \lambda_1^\top(f^0(x_0, u_0) - x_1) + \cdots + \lambda_N^\top(f^{N-1}(x_{N-1}, u_{N-1}) - x_N) = \phi(x_N) + \sum_{i=0}^{N-1}(L^i(x_i, u_i) + \lambda_{i+1}^\top(f^i(x_i, u_i) - x_{i+1}))$
- Let $H^i(x_i, u_i), \lambda_{i+1}) = L^i(x_i, u_i) + \lambda_{i+1}^\top(f^i(x_i, u_i))$.
- Then, we can rewrite $\bar{J}$ as $\bar{J} = (\phi(x_N) - \lambda_N^\top x_N) + \sum_{i=1}^{N-1}(H^i(x_i, u_i) - \lambda_i^\top x_i) + H^0$.

First order analysis: Here $dx(0) = 0$ because the initial condition is given (no need for variation).

$$d\bar{J} = \left(\frac{\partial \phi(x_N)}{\partial x_N} - \lambda_N\right)^\top dx_N + \sum_{i=1}^{N-1}\left(\left(\frac{\partial H^i}{\partial x_i}\right)^\top dx_i + \left(\frac{\partial H^i}{\partial u_i}\right)^\top du_i - \lambda_i^\top dx_i\right) + \left(\frac{\partial H^0}{\partial x_0}\right)^\top dx_0 + \left(\frac{\partial H^0}{\partial u_0}\right)^\top du_0$$

$$= \left(\frac{\partial \phi(x_N)}{\partial x_N} - \lambda_N\right)^\top dx_N + \sum_{i=1}^{N-1}\Big(\left(\frac{\partial H^i}{\partial x_i} - \lambda_i\right)^\top dx_i + \left(\frac{\partial H^i}{\partial u_i}\right)^\top du_i\Big) + \left(\frac{\partial H^0}{\partial u_0}\right)^\top du_0$$

First order optimality condition for multi-stage optimal control (cont'd)

- Think of $du(i)$, $i = 0, \cdots, N-1$ as free variable and $dx(i+1)$ the dependent variable, which is defined from the constraint equation (constraint equation relates $dx(i+1)$ to $du(i)$).
- Next, pick λ_N such that
$$\frac{\partial \phi(x_N)}{\partial x_N} - \lambda_N = 0,$$
- also pick λ_i, such that
$$\frac{\partial H^i}{\partial x_i} - \lambda_i = 0, \quad i = 1, \cdots, N-1.$$
- Then, we have
$$d\bar{J} = \sum_{i=1}^{N-1} \Big((\frac{\partial H^i}{\partial u_i})^\top du_i \Big) + (\frac{\partial H^0}{\partial u_0})^\top du_0$$

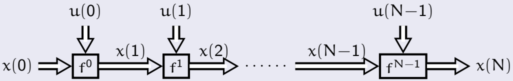
  For $(u(0), \cdots, u(N-1), x(1), \cdots, x(N))$ to be a minimum point we need $d\bar{J} = \sum_{i=1}^{N-1} \Big((\frac{\partial H^i}{\partial u_i})^\top du_i \Big) + (\frac{\partial H^0}{\partial u_0})^\top du_0 \geqslant 0$. Because we are free to vary du_i, $i = 0, \cdots, N-1$ in all directions, then the necessary condition for $(u(0), \cdots, u(N-1), x(1), \cdots, x(N))$ to be a minimum point is $\frac{\partial H^i}{\partial u_i} = 0$, $i = 0, \cdots, N-1$.
- Putting all the conditions we stated and derived, we obtain:
  First order necessary condition for $(x(1), u(0))$ to be a minimum point:
$$\begin{cases} \lambda_N = \frac{\partial \phi(x_N)}{\partial x_N}, & n \text{ eq} \\ \frac{\partial H^i}{\partial u_i} = 0, \quad i = 0, \cdots, N-1 & N\,m \text{ eq} \\ x_{i+1} = f^i(x_i, u_i), \quad i = 0, \cdots, N-1 & N\,n \text{ eq}^2 \end{cases}$$

  Here, we have $2Nn + Nm$ equation for $2Nn + m$ unknowns (the unknowns in the set of equations above are $\lambda_i \in \mathbb{R}^n$, $x_i \in \mathbb{R}^n$ and $u_{i-1} \in \mathbb{R}^m$, $i = 1, \cdots, N$).

[2] which can also be written as $x_{i+1} = \frac{\partial H^i}{\partial \lambda_{i+1}}$

Optimal control of multi-stage systems over finite horizon

$$u^\star = \operatorname{argmin} \underbrace{\phi(x(N)) + \sum_{k=0}^{N-1} L^k(x(k), u(k))}_{J(u(0),\cdots,u(N-1))} \quad \text{s.t.}$$



$$H^k = L^k(x(k), u(k)) + \lambda(k+1)^\top f^k(x(k), u(k)), \quad k = 0, 1, \cdots, N-1$$

Free final state

$$\lambda(N) = \frac{\partial \phi(x(N))}{\partial x(N)},$$

$$\lambda(k) = \frac{\partial H^k}{\partial x(k)}, \ k = 1, \cdots, N-1,$$

$$0 = \frac{\partial H^k}{\partial u(k)}, \ k = 0, \cdots, N-1,$$

$$x(k+1) = \frac{\partial H^k}{\partial \lambda(k+1)}, \ k = 1, \cdots, N-1,$$

$$\begin{cases} x(0) = x_0, & \text{given initial condition,} \\ 0 = \frac{\partial H^0}{\partial x(0)}, & \text{free initial condition.} \end{cases}$$

Constrained final state, i.e., $\psi(x(n)) = 0, \ \ \psi : \mathbb{R}^n \to \mathbb{R}^p, \ p \leqslant N$

$$\psi(x(n)) = 0,$$

$$\lambda(N) = \frac{\partial(\phi(x(N)) + \nu^\top \psi(x(n)))}{\partial x(N)},$$

$$\lambda(k) = \frac{\partial H^k}{\partial x(k)}, \ k = 1, \cdots, N-1,$$

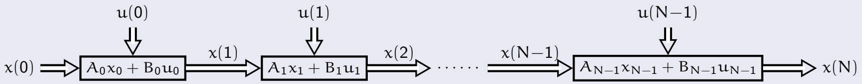
$$0 = \frac{\partial H^k}{\partial u(k)}, \quad k = 0, \cdots, N-1,$$

$$x(k+1) = \frac{\partial H^k}{\partial \lambda(k+1)}, \ k = 1, \cdots, N-1,$$

$$\begin{cases} x(0) = x_0, & \text{given initial condition,} \\ 0 = \frac{\partial H^0}{\partial x(0)}, & \text{free initial condition.} \end{cases}$$

Optimal control of multi-stage systems over finite horizon: regulator problem

$$u^\star = \operatorname{argmin} \frac{1}{2} x_N^\top S_N x_N + \frac{1}{2} \sum_{k=0}^{N-1} x_k^\top Q_k x_k + u_k^\top R_k u_k \quad \text{s.t.}$$



$$H^k = \frac{1}{2} x_k^\top Q_k x_k + \frac{1}{2} u_k^\top R_k u_k + \lambda_{k+1}^\top (A_k x_k + B_k u_k), \quad k = 0, 1, \cdots, N-1$$

Free final state: Linear systems with given initial condition

$$\lambda(N) = \frac{\partial \phi(x(N))}{\partial x(N)} \implies \lambda_N = S_N x_N,$$

$$\lambda(k) = \frac{\partial H^k}{\partial x(k)}, \quad k = 1, \cdots, N-1 \implies \lambda_k = Q_k x_k + A_k^\top \lambda_{k+1}, \quad k = 1, \cdots, N,$$

$$0 = \frac{\partial H^k}{\partial u(k)}, \quad k = 0, \cdots, N-1 \implies 0 = R_k u_k + B_k^\top \lambda_{k+1}, \quad k = 0, \cdots, N-1,$$

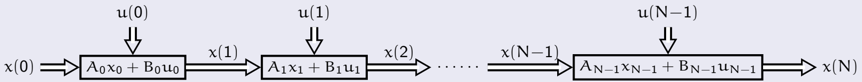
$$x(k+1) = \frac{\partial H^k}{\partial \lambda(k+1)}$$

$$= f^k(x(k), u(k)), \; k = 0, \cdots, N-1 \implies x_{k+1} = A_k x_k + B_k u_k, \quad k = 1, \cdots, N-1,$$

$$x(0) = x_0 \implies x(0) = x_0.$$

Optimal control of multi-stage systems over finite horizon: regulator problem

$$u^\star = \operatorname{argmin} \frac{1}{2} x_N^\top S_N x_N + \frac{1}{2}\sum_{k=0}^{N-1} x_k^\top Q_k x_k + u_k^\top R_k u_k \quad \text{s.t.}$$



$$\lambda_N = S_N x_N,$$

$$\lambda_k = Q_k x_k + A_k^\top \lambda_{k+1}, \quad k = 1, \cdots, N-1,$$

$$0 = R_k u_k + B_k^\top \lambda_{k+1}, \Rightarrow u^\star = -R_k^{-1} B_k^\top \lambda_{k+1}, \quad k = 0, \cdots, N-1$$

$$x_{k+1} = A_k x_k + B_k u_k, \Rightarrow x_{k+1} = A_k x_k - B_k R_k^{-1} B_k^\top \lambda_{k+1} \quad k = 1, \cdots, N-1,$$

$$x(0) = x_0.$$

$$- - - - - - - - - - - - - - - - -$$

$$\begin{bmatrix} x(k+1) \\ \lambda(k) \end{bmatrix} = \begin{bmatrix} A_k & -B_k R_k^{-1} B_k^\top \\ Q_k & A_k^\top \end{bmatrix} \begin{bmatrix} x(k) \\ \lambda(k+1) \end{bmatrix}, \quad x(0) = x_0,\ \lambda_N = S_N x_N.$$

If A_k is invertible: $x_k = A_k^{-1} x_{k+1} + A_k^{-1} B_k R_k^{-1} B_k^\top \lambda_{k+1}$. Then, we can write

$$\begin{bmatrix} x(k) \\ \lambda(k) \end{bmatrix} = \begin{bmatrix} A_k^{-1} & A_k^{-1} B_k R_k^{-1} B_k^\top \\ Q_k A_k^{-1} & A_k^\top + Q_k A_k^{-1} B_k R_k^{-1} B_k^\top \end{bmatrix} \begin{bmatrix} x(k+1) \\ \lambda(k+1) \end{bmatrix}, \quad x(0) = x_0,\ \lambda_N = S_N x_N.$$

If we had x_N and λ_N, we could solve the equation above backward in time, but unfortunately we have x_0 and λ_N.