1 Optimization on linear spaces

1.1 Unconstrained optimization

Definition 1. A normed linear space is a vector space $X$ with a real-valued norm $\|x\|$ which satisfies:

1. $\|x\| \geq 0$ for all $x \in X$ and $\|x\| = 0$ if and only if $x = 0$
2. $\|x + y\| \leq \|x\| + \|y\|$
3. $\|\alpha x\| = |\alpha| \cdot \|x\|$

Example 1. The set $C[a, b]$ consisting of all real-valued, continuous functions on the interval $[a, b] \in \mathbb{R}$ with $\|x\| = \max |x(t)|$.

Example 2. The set $D[a, b]$ consisting of all real-valued, continuously differentiable functions on the real interval with $\|x\| = \max |x(t)| + \max |\dot{x}(t)|$.

Example 3. The set $l_p$ consisting of real-valued sequences $\{x_1, x_2, \ldots \}$ with norm

$$\|x\|_p = \left( \sum_{i=1}^{\infty} |x_i|^p \right)^{\frac{1}{p}}$$

Example 4. The set $L_p[a, b]$ consisting of real-valued functions on the interval $[a, b] \in \mathbb{R}$ with norm

$$\|x\|_p = \left( \int_a^b |x(t)|^p \, dt \right)^{\frac{1}{p}}$$

and where we identify functions that disagree only on a set of measure zero.

Example 5. Let $X$ and $Y$ be linear spaces with norms $\| \cdot \|_X$ and $\| \cdot \|_Y$, respectively. Let $B(X, Y)$ be the set of bounded linear operators taking $X$ to $Y$ with norm

$$\|A\| = \sup_{\|x\|_X} \|Ax\|_Y.$$ 

Definition 2. Let $X$ and $Y$ be normed linear spaces and $T : D \subseteq X \rightarrow Y$. The mapping $T$ is Frechet differentiable at $x \in D$ if for every $h \in X$ there exists $\delta T(x; h) \in Y$ which is linear and continuous in $h$ and satisfies

$$\lim_{\|h\| \rightarrow 0} \frac{\|T(x + h) - T(x) - \delta T(x; h)\|}{\|h\|} = 0$$
Proposition 1. If $\delta T$ exists, it is unique.

It will be convenient to introduce some additional notation for Frechet derivatives. For $T : X \to Y$ it follows from the definition that $\delta T(x; h) = A_x h$ for some bounded, linear operator $A_x : X \to Y$. Let $B(X, Y)$ be the normed linear space of bounded linear operators from $X$ to $Y$ and define $T' : X \to B(X, Y)$ as

$$T'(x)h = \delta T(x; h).$$

We call the mapping $T'$ the Frechet derivative of the mapping $T$.

Definition 3. Let $f : \Omega \to \mathbb{R}$ be a real-valued function on $\Omega \subset X$. A point $x_0 \in \Omega$ is a local minimum of $f$ on $\Omega$ if there exists a neighborhood $N \subset X$ containing $x_0$ such that $f(x_0) \leq f(x)$ for all $x \in \Omega \cap N$. The point $x_0$ is called a strict local minimum if $f(x_0) < f(x)$ for all $x \in \Omega \cap N, x \neq x_0$.

A similar definition holds for maxima. A point is called an extremum if it is either a maximum or a minimum.

Theorem 2. Let $f : X \to \mathbb{R}$ be Frechet differentiable on $x$. Then $f$ has an extremum at $x_0$ in $X$ only if $\delta f(x_0; h) = 0$ for all $h \in X$

Proof. If $f$ has an extremum at $x_0$ then $f(x_0 + \alpha h) : \mathbb{R} \to \mathbb{R}$ achieves an extrema at $\alpha = 0$. By ordinary calculus

$$\frac{d}{d\alpha} f(x_0 + \alpha h) \bigg|_{\alpha=0} = \delta f(x_0; h) = 0$$

A point $x_0$ where $\delta f(x_0; h) = 0$ for all $h \in X$ (or, equivalently, $f'(x_0) = 0$) is called a stationary point of $f$. Hence a necessary condition for $f$ to achieve a local extremum at $x_0$ is that $x_0$ be a stationary point of $f$.

Remark 1. In the finite dimensional case, we can rewrite the Frechet derivative using the gradient and so if $x_0$ is an extrema then $\nabla f = 0$. Note that the gradient of a function is only defined relative to a metric and so we are implicitly using additional structure in asking that the gradient be zero at an extrema. This can be avoided by requiring $df = 0$, where $df$ is a one-form and hence doesn’t require the notion of a metric.
Example 6 (Euler-Lagrange equations). Let $X = D[t_1, t_2]$ and consider the problem of finding a function $x(\cdot)$ defined on $[t_1, t_2]$ that minimizes the cost index

$$J(x) = \int_{t_1}^{t_2} L(x, \dot{x}, t) dt.$$  

We further assume that the endpoints of the function $x(\cdot)$ fixed.

Let $h$ represent a vector in $D[t_1, t_2]$ that vanishes at the end points. Define

$$\delta J(x; h) = \frac{d}{d\alpha} \left. \int_{t_1}^{t_2} L(x + \alpha h, \dot{x} + \alpha \dot{h}, t) dt \right|_{\alpha=0}$$  

$$= \int_{t_1}^{t_2} \frac{\partial L}{\partial x}(x, \dot{x}, t) h(t) dt + \int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t) \dot{h}(t) dt$$  

(1)

Claim 3. $\delta J$ is the Frechet derivative of $J : X \rightarrow \mathbb{R}$.

We can now compute the necessary condition given by $\delta J(x; h) = 0$. Integrating equation (1) by parts, we obtain:

$$0 = \int_{t_1}^{t_2} \frac{\partial L}{\partial x}(x, \dot{x}, t) h(t) dt + \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t) h(t) \bigg|_{t_1}^{t_2} - \int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t) \dot{h}(t) dt$$

$$= \int_{t_1}^{t_2} \left[ \frac{\partial L}{\partial x}(x, \dot{x}, t) - \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t) \right] h(t) dt.$$

If this equation is to hold for all admissible $h \in X$ then it most follow that

$$\frac{\partial L}{\partial x}(x, \dot{x}, t) - \frac{\partial L}{\partial \dot{x}}(x, \dot{x}, t) = 0.$$  

These equations are known as the Euler-Lagrange equations.

Theorem 4. Let $f : X \rightarrow \mathbb{R}$ be twice Frechet differentiable on $x$ and let $x_0 \in X$ be a local extrema. Then $x_0$ is a local maximum if

$$f''(x_0)(h, h) < 0 \quad \text{for all } h \in X.$$  

Remark 2. This gives a sufficient condition for an extrema to be a local maxima. For a function $f(x)$ defined on $\mathbb{R}^n$, the matrix of derivatives $f''$ is called the Hessian.
1.2 Optimization with equality constraints

We now consider the problem of minimizing a functional $f : X \to \mathbb{R}$ subject to a finite number of equality constraints, $g_i(x) = 0$, $i = 1, \ldots, m$ where each $g_i : X \to \mathbb{R}$ is a Frechet differentiable function.

**Definition 4.** Let $g_i : X \to \mathbb{R}$, $i = 1, \ldots, m$ be a Frechet differentiable function. A point $x_0 \in X$ satisfying $g_i(x_0) = 0$, $i = 1, \ldots, m$ is a regular point if $g'_1(x_0), \ldots, g'_m(x_0)$ are linearly independent.

**Theorem 5.** If $x_0$ is an extremum of $f : X \to \mathbb{R}$ subject to $g_i(x_0) = 0$, $i = 1, \ldots, m$ and if $x_0$ is a regular point of the $g_i$, then

$$
\delta f(x_0; h) = 0 \quad \text{for all } h \text{ such that } \delta g_i(x_0; h) = 0, \ i = 1, \ldots, m.
$$

**Proof.** [Finite dimensional case] Choose coordinates $x = (\xi, \eta) \in \mathbb{R}^{n-m} \times \mathbb{R}^m$ such that $\left[ \frac{\partial g_i}{\partial \eta} \right]$ is full rank. By the implicit function theorem, $g_i(\xi, \eta)$, $i = 1, \ldots, m$ gives $\eta_i = \phi_i(\xi)$ locally. Infinitesimally, we have

$$
\begin{bmatrix}
\frac{\partial g}{\partial \xi} \\
\frac{\partial g}{\partial \eta}
\end{bmatrix}
\begin{bmatrix} I \\
\frac{\partial \phi}{\partial \xi}
\end{bmatrix} = 0 \quad (2)
$$

(by differentiating $g(\xi, \eta) = 0$). Now, if $x_0 = (\xi_0, \eta_0)$ is an extremum on $g(\xi, \eta) = 0$ then $\tilde{f}(\xi) = f(\xi, \phi(\xi))$ has an extremum at $x_0$. Thus

$$
0 = \frac{\partial \tilde{f}}{\partial \xi} = \frac{\partial f}{\partial \xi} + \frac{\partial f}{\partial \eta} \frac{\partial \phi}{\partial \xi} = \begin{bmatrix} \frac{\partial f}{\partial \xi} & \frac{\partial f}{\partial \eta} \end{bmatrix} \begin{bmatrix} I \\
\frac{\partial \phi}{\partial \xi}
\end{bmatrix} \quad (3)
$$

Equations (2) and (3) imply that

$$
\delta f(x_0; h) = 0 \quad \text{for all } h \text{ such that } \delta g_i(x_0; h) = 0, \ i = 1, \ldots, m.
$$

For the infinite dimensional proof, one needs a version of the implicit function theorem on normed linear spaces plus an approximation result which shows that locally ker $g'$ approximates the constraints (see Luenberger). Note that Luenberger only proves the implicit function theorem for Banach spaces.
Example 7 (Local version in finite dimensions).

Let \((x, y)\) represent a point in the linear space and write the function to be minimized as real-valued function \(f(x, y)\) and the constraints as a vector-valued equation \(g(x, y) = 0\). We assume that \(x\) and \(y\) have been chosen such that \(\frac{\partial g}{\partial x}\) is full rank. Then a necessary condition for \(x_0\) to be an extremum is that
\[
\frac{\partial f}{\partial y} - \left(\frac{\partial g}{\partial y}\right)^T \left(\frac{\partial g}{\partial x}\right)^{-T} \frac{\partial f}{\partial x} = 0.
\]

Remark 3 (Euler-Lagrange equations, revisited). In the derivation of the Euler-Lagrange equations in Example 6, we restricted the end points of the function to be fixed. This gives a constrained optimization problem by imposing the constraints:
\[
g_1(x) = x(t_1) - x_0 = 0 \quad g_2(x) = x(t_2) - x_f = 0
\]
Using Theorem 5, we see that in fact choosing functions \(h\) which vanish at the endpoints is precisely choosing the set of \(h\) such that \(\delta g(x; h) = 0\).

Theorem 6. If \(x_0\) is an extremum of \(f : X \to \mathbb{R}\) subject to \(g_i(x) = 0, i = 1, \ldots, m\) then there exist scalars \(\lambda_i \in \mathbb{R}\) such that
\[
h(x) = f(x) + \sum_{i=1}^{n} \lambda_i g_i(x)
\]
has a stationary point at \(x_0\).

Proof. Homework exercise \(\square\)

Example 8. Let \(x(t) \in \mathbb{R}\) be a curve on the interval \([-1, 1]\) which is zero at its endpoints. We wish to maximize the integral of \(x\), subject to the constraint that the “length” of the curve in the \(t\)-\(x\) plan is \(l\): [PICTURE]
\[
\max \int_{-1}^{1} x(t) \, dt \quad \text{subject to} \quad \int_{-1}^{1} \sqrt{x + 1} \, dt = l.
\]
Using Theorem 6, we seek to maximize, for some fixed \(\lambda\) to be determined
\[
J(x) = \int_{-1}^{1} \left(x + \lambda \sqrt{x^2 + 1}\right) \, dt.
\]
Applying the Euler-Lagrange equations,
\[
\frac{d}{dt} \frac{\partial J}{\partial \dot{x}} - \frac{\partial J}{\partial x} = \lambda \frac{d}{dt} \frac{\dot{x}}{\sqrt{\dot{x}^2 + 1}} - 1 = 0,
\]
and hence
\[
\frac{\dot{x}}{\sqrt{\dot{x}^2 + 1}} = \frac{t}{\lambda} + c.
\]
This gives us a differential equation which the optimal solution must satisfy, in addition to the boundary and length constraints. It can be shown that in this case the answer is given by
\[
(x - x_1)^2 + (t - t_1)^2 = r^2
\]
where the constants \(x_1, t_1,\) and \(r\) are chosen to satisfy the end conditions and length constraint.

**Corollary 6.1.** If \(x_0\) is an extremum of \(f\) subject to constraints \(g_i\) then
\[
f' + \lambda_i g'_i = 0.
\]

This corollary roughly corresponds to the geometric picture that the derivative of the cost function be aligned with the “normal” to the constraint surface. [PICTURE]

**Theorem 7.** If \(x_0\) is an extremum of \(f\) subject to \(g(x) = 0\) where \(g : X \rightarrow Y\) then there exists a \(\lambda_0 \in Y^*\) such that \((x_0, \lambda_0)\) is a stationary point for \(H : X \times Y^* \rightarrow \mathbb{R}\) given by
\[
H(x, \lambda) = f(x) + \langle \lambda, g(x) \rangle.
\]

Note that in this version we do not require that \(Y\) be finite dimensional. This relies on the infinite dimensional version of Theorem 6 (see Luenberger, Section 9.3).

**Proof.** A stationary point of \(H\) satisfies
\[
\delta H = \begin{bmatrix} \delta_x H \\ \delta_\lambda H \end{bmatrix} = \begin{bmatrix} \delta f(x) + \lambda \delta g(x) \\ g(x) \end{bmatrix}
\]

**Remark 4.** Note that in the case where \(X\) and \(Y\) are finite dimensional, the conditions for a stationary point give \(n + m\) equations in \(n + m\) unknowns.

**Example 9.** Rework example from Luenberger.
1.3 Minimum norm problems on Hilbert spaces

**Definition 5.** A pre-Hilbert space is a linear vector space $X$ together with an inner product mapping $X \times X \to \mathbb{R}$ and satisfying the following properties:

1. $\langle x, y \rangle = \langle y, x \rangle$
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
3. $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$
4. $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

**Proposition 8.** The inner product on a pre-Hilbert space satisfies the following properties:

1. $\|x\| = \sqrt{\langle x, x \rangle}$ is a norm
2. $|\langle x, y \rangle| \leq \|x\| \|y\|$, with equality if and only if $x = \lambda y$ or $y = 0$.
3. $\langle x, y \rangle = 0$ for all $y \in X \implies x = 0$.
4. $\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2$ (parallelogram law).
5. $\langle x, y \rangle$ is continuous in $x$ and $y$

**Example 10.** The set $L^2[a, b]$ consisting of real-valued functions on the interval $[a, b] \in \mathbb{R}$ with inner product

$$\langle x, y \rangle = \left( \int_a^b x(t)y(t) \, dt \right)^\frac{1}{2}$$

and where we identify functions that disagree only on a set of measure zero.

**Example 11.** The space of real-valued polynomial functions on $[a, b]$ with inner product $\langle x, y \rangle = \int_a^b x(s)y(s) \, ds$.

**Definition 6.** In a pre-Hilbert space $X$, two vectors $x, y \in X$ are orthogonal (written $x \perp y$) if $\langle x, y \rangle = 0$. A vector $x$ is orthogonal to a set $S$ (written $x \perp S$) if $x \perp s$ for each $s \in S$.

**Proposition 9.** If $x \perp y$ then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$. 
Theorem 10. Let $X$ be a pre-Hilbert space, $M$ a subspace of $X$, and $x \in X$ a given vector. If there exists a vector $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$ then $m_0$ is unique. Furthermore, $m_0 \in M$ is a unique minimizing vector in $M$ if and only if $(x - m_0) \perp M$.

Proof.

Necessity. By contradiction. Let $m \in M$ be a vector which is not orthogonal and assume WLOG that $\|m\| = 1$ and $\langle x - m_0, m \rangle = \delta$. Let $m_1 = m_0 + \delta m$. Then
\[ \|x - m_1\|^2 = \|(x - m_0) - \delta m\|^2 = \|x - m_0\|^2 - 2\langle x - m_0, \delta m \rangle + \delta^2 = \|x - m_0\|^2 - \delta^2 < \|x - m_0\|^2, \]
which is a contradiction.

Sufficiency. Let $m_0$ be such that $(x - m_0) \perp M$ and let $m \in M$. Then
\[ \|x - m\|^2 = \|x - m_0 + m_0 - m\|^2 = \|x - m_0\|^2 + \|m_0 - m\|^2 \]
and $m_0$ is minimizing.

Uniqueness. Suppose $m_0$ and $m'_0$ are both minimizing. Then $(m_0 - m'_0) \perp M \implies m_0 - m'_0 = 0$. \hfill \Box

Definition 7. A pre-Hilbert space $H$ is complete if every Cauchy sequence in $H$ converges to a point in $H$. A Hilbert space is a pre-Hilbert space which is complete.

Theorem 11 (Projection theorem). Let $M$ be a closed subspace of a Hilbert space $H$. Given any $x \in H$, there exists a unique $m_0 \in M$ such that $\|x - m_0\| \leq \|x - m\|$ for all $m \in M$. Furthermore, $m_0 \in M$ is a unique minimizing vector in $M$ if and only if $(x - m_0) \perp M$.

Proof. (Sketch) The only thing that has to be proven is the existence of the minimizer. Roughly, since $M$ is a closed subspace of a Hilbert space, $M$ is a Hilbert space. Now construct a Cauchy sequence $m_i$ such that $\|m_i - x\|$ converges to $\delta = \inf \|x - m\|$. Therefore, $m_i$ approaches $m_0 \in M$ and by continuity of the norm, $\|x - m\| = \delta$. For more details, see Luenbuerger. \hfill \Box
Example 12. Consider the (trivial) problem of finding a point on a line through the origin which is closed to a given point in $\mathbb{R}^2$ [PICTURE]. For this problem, $X = \mathbb{R}^2$,

$$M = \{(m, am) : m \in \mathbb{R}\} \quad x = (x_1, x_2) \in \mathbb{R}$$

and we wish to minimize $\|x - m\|$. Applying the projection theorem

$$\begin{bmatrix} x_1 - m_1 \\ x_2 - m_2 \end{bmatrix} \cdot \begin{bmatrix} m \\ am \end{bmatrix} = (x_1 - m_1)m + a(x_2 - m_2)m.$$ 

Using the fact that $m_2 = am_1$ on $M$, we can solve to obtain

$$m_0 = \frac{1}{1 + a^2} \begin{bmatrix} x_1 + ax_2 \\ a(x_1 + ax_2) \end{bmatrix}$$

Example 13. To see why we need the structure of a Hilbert space for this theorem, let $X = \mathbb{R}^2$, $\|x\| = \max |x_i|$, and $M = \{ (\lambda, 0) : \lambda \in \mathbb{R} \}$. If we seek to find the minimum on $M$ from the point $(2, 1)$ we see that this point is not unique.

Theorem 12. Let $X$ be a normed linear space, $x \in X$ be a given point, and let $d \in \mathbb{R}$ denote the distance from $x$ to a subspace $M$. Then

$$d = \inf_{m \in M} \|x - m\| = \max_{\|x^*\| \leq 1} \langle x, x^* \rangle$$

and if the infemum is achieved for some $m_0 \in M$ then $\langle x^*, x - m_0 \rangle = 0$ (i.e., $X^*$ and $x - m_0$ are aligned).
2 Optimal Control of Discrete Time Systems

2.1 Discrete-time, finite-horizon problems

Consider a discrete time, nonlinear control system of the form

\[ x_{k+1} = f(x_k, u_k, k) \]

\[ x_k \in \mathbb{R}^n \]

\[ u_k \in \mathbb{R}^m \]

\[ k = 0, \ldots, N - 1 \]

with initial condition \( x_0 = x_0^d \) and (desired) final condition \( x_N = x_N^d \). We assume that \( f \) is smooth in each of its arguments. If \( f \) is independent of \( k \), we say that the system is \textit{time invariant}.

We consider a general cost function of the form

\[ J(x, u) = \sum_{k=0}^{N-1} L(x_k, u_k, k) \]

where \( L \) is a smooth function, possibly depending on time \( k \). The \textit{optimal control problem} is to find a control \( u^*_k \) and associated state \( x^*_k \) that steer the system from \( x_0^d \) to \( x_N^d \) and minimize \( J(x, u) \).

**Example 14 (Minimum fuel problems).** To minimize the size of the input, choose

\[ J_{mf}(x, u) = \sum_{k=0}^{N-1} \| u_k \|. \]

Weighted norms and time-varying weights can also be used.

**Example 15 (Minimum energy problems).**

\[ J_{me}(x, u) = \sum_{k=0}^{N-1} x_k^T Q_k x_k + u_k^T R_k u_k \]

where \( Q_k > 0 \) and \( R_k > 0 \). Typically \( Q_k \) and \( R_k \) are constant, leading to a time-invariant cost function.

An important variation of this problem, to be studied in Section 2.3, is to remove the endpoint constraint and instead add the desired endpoint to the cost function. This leads to a cost function of the form

\[ \tilde{J}_{me}(x, u) = x_N^T S x_N + \sum_{k=0}^{N-1} x_k^T Q_k x_k + u_k^T R_k u_k \]
and \( x_N \) free.

**Example 16 (Minimum time problems).** For many problems, we wish to get to the endpoint as quickly as possible, and hence we choose

\[
J_{mt}(x, u) = \sum_{k=0}^{N-1} 1 = N.
\]

This problem is ill-posed unless we also constrain the inputs to be bounded, typically by requiring \(|u_{k,i}| < M_i\) for all \( i,k \). We shall defer solution of this problem until Section \( ?? \).

The discrete-time, finite horizon problem is a finite-dimensional optimization problem on the space

\[
X = \mathbb{R}^{n \times N}_{\text{states}} \times \mathbb{R}^{m \times N}_{\text{inputs}}
\]

with constraints

\[
\begin{align*}
g_0(x, u) &= x_0 - x_0^d = 0 \\
g_{k+1}(x, u) &= x_{k+1} - f(x_k, u_k) = 0 \\
g_N(x, u) &= x_N - x_N^d = 0.
\end{align*}
\]

A necessary condition for \((x^*, u^*) \in X\) to be a solution is that

(a) \( \delta J(x; \eta) = 0 \) for all \( \eta \) such that \( \delta g(x; \eta) = 0 \); or

(b) There exist a \( \lambda^* \in \mathbb{R}^{N+1} \) such that \( J'(x, u, \lambda) = J(x, u) + \sum \lambda^T_{k+1} g_k \)

has a stationary point at \((x^*, u^*, \lambda^*)\).

Taking the differential of \( J' \), we have

\[
dJ' = \sum_{k=0}^{N-1} \left( \frac{\partial L}{\partial x_k} \, dx_k + \frac{\partial L}{\partial u_k} \, du_k \right) + \sum_{k=0}^{N-1} (\lambda^T_{k+1} \frac{\partial g_k}{\partial x_k} \, dx_k + \lambda^T_{k+1} \frac{\partial g_k}{\partial u_k} \, du_k) + g_{k,i} d\lambda_{k+1,i},
\]

which we can rewrite as

\[
\begin{align*}
\sum_{k=0}^{N-1} \left( \frac{\partial L}{\partial x_k} + \lambda^T_{k+1} \frac{\partial g_k}{\partial x_k} \right) dx_k &= 0 \\
\lambda^T_N dx_N &= 0 \\
\sum_{k=0}^{N-1} \left( \frac{\partial L}{\partial u_k} + \lambda^T_{k+1} \frac{\partial g_k}{\partial u_k} \right) du_k &= 0 \\
x_{k+1} - f(x_k, u_k, k) &= 0
\end{align*}
\]
The boundary conditions require that \( dx_0 \) and \( dx_N \) are both identically zero (this is encoded in the first two equations). We are left with a two point boundary value problem (PDE) which we must solve.

To convert the solution to a more useful form, we define the Hamiltonian function as

\[
H(x_k, u_k, \lambda_k, k) := L(x_k, u_k, k) + \lambda_{k+1}^T f(x_k, u_k, k)
\]

and the cost function becomes

\[
J'(x, u, \lambda) = H(x_0, u_0, 0) + \sum_{k=1}^{N-1} (H(x_k, u_k) - \lambda_k^T x_k)
\]

(after a shift of indices). The condition that \( dJ' = 0 \) can now be written as

\[
x_{k+1} = \frac{\partial H}{\partial \lambda_{k+1}} = f(x_k, u_k, k) \tag{4}
\]

\[
\lambda_k = \frac{\partial H}{\partial x_k} = \frac{\partial L}{\partial x_k} + \lambda_{k+1}^T \frac{\partial f}{\partial x_k} \tag{5}
\]

\[
0 = \frac{\partial H}{\partial u_k} = \frac{\partial L}{\partial u_k} + \lambda_{k+1}^T \frac{\partial f}{\partial u_k} \tag{6}
\]

Equations (4) and (5) describe the evolution of the state \( x \) and costate \( \lambda \). Equation (5) is also called the adjoint system. Equation (6) is called the stationarity condition and can be solved (under some regularity assumptions) for \( u_k \).

### 2.2 Linear Quadratic Regulator: fixed final state

Consider now the special case of a linear discrete time system

\[
x_{k+1} = A_k x_k + B_k u_k
\]

(note that we are not assuming that the system is time invariant). We consider a cost function of the form

\[
J(x, u) = \sum_{k=0}^{N-1} x_k^T Q_k x_k + u_k^T R_k u_k.
\]
Applying the general conditions, we obtain
\[
\begin{align*}
x_{k+1} &= A_k x_k + B_k u_k \\
\lambda_k &= Q_k x_k + A_k^T \lambda_{k+1} \\
0 &= R_k u_k + B_k^T \lambda_{k+1}.
\end{align*}
\]
Solving the last equation for \( u_k \),
\[
u_k = R_k^{-1} B_k^T \lambda_{k+1}
\]
we get a solution for the optimal input sequence and optimal state sequence in terms of the costate sequence. Combining equations, the system has the form
\[
\begin{bmatrix}
x_{k+1} \\
\lambda_k
\end{bmatrix} =
\begin{bmatrix}
A & -BR^{-1}B^T \\
Q & A^T
\end{bmatrix}
\begin{bmatrix}
x_k \\
\lambda_{k+1}
\end{bmatrix}
\]
Note that this equation evolves backwards and forwards in time and that it is still a two point boundary value problem: \( x_0 \) and \( x_N \) are given.

In the case when \( A \) is nonsingular (e.g., \( A \) is hyperbolic), we can write
\[
x_k = A^{-1} x_{k+1} + A^{-1} BR^{-1} B^T \lambda_{k+1}
\]
and the system becomes
\[
\begin{bmatrix}
x_k \\
\lambda_k
\end{bmatrix} =
\begin{bmatrix}
A^{-1} & A^{-1}BR^{-1}B^T \\
QA^{-1} & A^T + QA^{-1}BR^{-1}B^T
\end{bmatrix}
\begin{bmatrix}
x_{k+1} \\
\lambda_{k+1}
\end{bmatrix}
\]
which evolves purely backwards in time. Again, \( x_0 \) and \( x_N \) are given, turning this into a two point boundary value problem.

2.3 Linear Quadratic Regulator: free final state