We will begin with a rather general treatment using nonlinear state space models. However, in future we will specialise the treatment to constrained linear systems since much more can be said about this case.
We will describe the system via a nonlinear state space model of the form
\[ x_{k+1} = f(x_k, u_k), \quad k \geq i \geq 0, \quad x_i = \bar{x}, \] (1)
where \( f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \) is some given nonlinear function, \( x \in \mathbb{R}^n \) is the system state and \( u \in \mathbb{R}^m \) is the control input.
We can think of the system of equations (1) as providing equality constraints that must be satisfied for all time instants $k$ in an interval of interest. It is also common to assume that there exist additional constraints on the system state and input, which can be stated as set constraints of the form

$$
\begin{align*}
    u_k & \in U \quad \text{for } k = i, i + 1, \ldots, i + N - 1, \\
    x_k & \in X \quad \text{for } k = i, i + 1, \ldots, i + N, \\
    x_{i+N} & \in X_f \subset X,
\end{align*}
$$

where $U \subset \mathbb{R}^m$, $X \subset \mathbb{R}^n$, and $X_f \subset \mathbb{R}^n$ are some sets, and $N$ is the optimisation horizon.
The fixed horizon optimal control problem of interest here is the following:

\[ \mathcal{P}_N(\bar{x}) : \quad V^\text{opt}_N(\bar{x}) \triangleq \min V_N(\{x_k\}, \{u_k\}), \]

subject to:
\[
\begin{align*}
    x_{k+1} &= f(x_k, u_k) \quad \text{for } k = i, i + 1, \ldots, i + N - 1, \\
    x_i &= \bar{x}, \\
    u_k &\in \mathbb{U} \quad \text{for } k = i, i + 1, \ldots, i + N - 1, \\
    x_k &\in \mathbb{X} \quad \text{for } k = i, i + 1, \ldots, i + N, \\
    x_{i+N} &\in \mathbb{X}_f \subset \mathbb{X},
\end{align*}
\]

where \( \{x_k\} \triangleq \{x_i, \ldots, x_{i+N}\}, \{u_k\} \triangleq \{u_i, \ldots, u_{i+N-1}\} \) are the state and control sequences.
$V_N({x_k}, {u_k})$ is the objective function given by

$$V_N({x_k}, {u_k}) \triangleq F(x_{i+N}) + \sum_{k=i}^{i+N-1} L(x_k, u_k).$$  \hspace{1cm} (4)

$F$ and $L$ are some functions whose properties will be defined later.
The state and control sequences that attain the minimum are the **optimal sequences**, or *minimisers*. The value of the objective function at the minimisers is $V_{N}^{\text{OPT}}(\bar{x})$. The function $V_{N}^{\text{OPT}}(\cdot)$ is called the *value function*, and is a function of the initial state only.
The remainder of this lecture will be concerned with necessary and sufficient conditions for the sequences \( \{x_i, \ldots, x_{i+N}\} \) and \( \{u_i, \ldots, u_{i+N-1}\} \) to be the minimisers of the above optimisation problem.
Our approach to constrained control (and later constrained estimation) evolves from classical optimal control theory. Our goal in this lecture is to summarise some of the key features of the latter theory.
We assume that the objective function and state equations do not depend explicitly on time. We can then take the initial time as $i = 0$ without loss of generality.
We thus consider the following optimal control problem:

\[ \mathcal{P}_N(\bar{x}) : \text{minimise } V_N(\{x_k\}, \{u_k\}), \]

subject to:

\[ x_{k+1} = f(x_k, u_k) \quad \text{for } k = 0, \ldots, N - 1, \]
\[ x_0 = \bar{x}, \]

where

\[ V_N(\{x_k\}, \{u_k\}) \triangleq F(x_N) + \sum_{k=0}^{N-1} L(x_k, u_k), \]

and \( \{x_k\} \triangleq \{x_0, \ldots, x_N\}, \{u_k\} \triangleq \{u_0, \ldots, u_{N-1}\}. \) We assume that \( f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \) \( F : \mathbb{R}^n \to \mathbb{R} \) and \( L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R} \) are differentiable functions of their variables.
We will derive necessary optimality conditions for the sequences $\{x_0^*, \ldots, x_N^*\}$ and $\{u_0^*, \ldots, u_{N-1}^*\}$ to be minimisers of the optimisation problem using the KKT necessary conditions. Note that problem (5)–(8) has equality constraints (given by the state equations (6)–(7)) and no inequality constraints. In order to apply the KKT optimality conditions, we need to verify the constraint qualification that the gradients of the equality constraints are linearly independent when evaluated at the minimisers.
Let us define a new variable

\[ \mathbf{x} \triangleq \begin{bmatrix} x_0^T & \cdots & x_N^T & u_0^T & \cdots & u_{N-1}^T \end{bmatrix}^T \in \mathbb{R}^{(N+1)n+Nm}, \]  

(9)

which comprises all the variables with respect to which the optimisation is performed.
We can then write the state equations (6)–(7) as $(N + 1)n$ equality constraints on $\mathbf{x}$ as follows:

$$h(\mathbf{x}) \triangleq \begin{bmatrix} \bar{x} - x_0 \\ f(x_0, u_0) - x_1 \\ \vdots \\ f(x_{N-1}, u_{N-1}) - x_N \end{bmatrix} = 0.$$  

(10)
Now let

\[ x_k \triangleq \begin{bmatrix} x^1_k & \cdots & x^n_k \end{bmatrix}^T \text{ for } k = 0, \ldots, N, \]  

(11)

\[ u_k \triangleq \begin{bmatrix} u^1_k & \cdots & u^m_k \end{bmatrix}^T \text{ for } k = 0, \ldots, N - 1, \]  

(12)

\[ f(x_k, u_k) \triangleq \begin{bmatrix} f_1(x_k, u_k) & \cdots & f_n(x_k, u_k) \end{bmatrix}^T, \]  

(13)
Define

\[
\frac{\partial f}{\partial x_k} \triangleq \begin{bmatrix}
\frac{\partial f_1}{\partial x^1_k} & \cdots & \frac{\partial f_1}{\partial x^n_k} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial x^1_k} & \cdots & \frac{\partial f_n}{\partial x^n_k}
\end{bmatrix}, \quad \frac{\partial f}{\partial u_k} \triangleq \begin{bmatrix}
\frac{\partial f_1}{\partial u^1_k} & \cdots & \frac{\partial f_1}{\partial u^m_k} \\
\vdots & \ddots & \vdots \\
\frac{\partial f_n}{\partial u^1_k} & \cdots & \frac{\partial f_n}{\partial u^m_k}
\end{bmatrix},
\] (14)

for \( k = 0, \ldots, N - 1 \).
We can compute the \((N + 1)n \times [(N + 1)n + Nm]\) Jacobian matrix of the vector-valued function \(h(\mathbf{x})\) as

\[
\frac{\partial h}{\partial \mathbf{x}} = \begin{bmatrix}
-l_n & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
\frac{\partial f}{\partial x_0} & -l_n & 0 & \cdots & 0 & 0 & \frac{\partial f}{\partial u_0} & 0 & \cdots & 0 \\
0 & \frac{\partial f}{\partial x_1} & -l_n & \cdots & 0 & 0 & 0 & \frac{\partial f}{\partial u_1} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \frac{\partial f}{\partial x_{N-1}} & -l_n & 0 & 0 & \cdots & \frac{\partial f}{\partial u_{N-1}} \\
\end{bmatrix},
\]

where 0 denotes zero matrices of appropriate dimensions, and \(l_n\) denotes the \(n \times n\) identity matrix.
Clearly, $\partial h/\partial x$ has full row rank for all $x \in \mathbb{R}^{(N+1)n+Nm}$, and hence the gradients of the equality constraints are linearly independent in $\mathbb{R}^{(N+1)n+Nm}$. Thus, the constraint qualification required by the KKT optimality conditions holds for all $x \in \mathbb{R}^{(N+1)n+Nm}$. 
Next, we introduce Lagrange multipliers $\lambda_{-1} \in \mathbb{R}^n$ for the initial state equation, and $\{\lambda_k\} \triangleq \{\lambda_0, \ldots, \lambda_{N-1}\}$, $\lambda_k \in \mathbb{R}^n$ (usually referred to as adjoint variables) for the state equations, and form the (real valued) Lagrangian function

$$\mathcal{L}(x, \lambda) \triangleq F(x_N) + \sum_{k=0}^{N-1} L(x_k, u_k) + \lambda_{-1}^T (\bar{x} - x_0) + \sum_{k=0}^{N-1} \lambda_k^T [f(x_k, u_k) - x_{k+1}],$$

(16)

where $\lambda \triangleq [\lambda_{-1}^T \lambda_0^T \ldots \lambda_{N-1}^T]^T$. 

Centre for Complex Dynamic Systems and Control
Let
\[ \mathbf{x}^* \triangleq \begin{bmatrix} (x_0^*)^T & \cdots & (x_N^*)^T & (u_0^*)^T & \cdots & (u_{N-1}^*)^T \end{bmatrix}^T \]
be the minimising vector corresponding to the sequences \( \{x_0^*, \ldots, x_N^*\} \) and \( \{u_0^*, \ldots, u_{N-1}^*\} \). Observe that the dual feasibility condition in the KKT conditions is equivalent to the statement that there exists \( \lambda^* \triangleq \begin{bmatrix} (\lambda_{-1}^*)^T & (\lambda_0^*)^T & \cdots & (\lambda_{N-1}^*)^T \end{bmatrix}^T \) such that the partial derivative \( \partial L / \partial \mathbf{x} \) of the Lagrangian function vanishes at \( (\mathbf{x}^*, \lambda^*) \).
Hence, the following must hold:

\[
\frac{\partial L(x^*, \lambda^*)}{\partial x_k} = 0 \quad \text{for } k = 0, \ldots, N, \\
\frac{\partial L(x^*, \lambda^*)}{\partial u_k} = 0 \quad \text{for } k = 0, \ldots, N - 1.
\]  

(17)
\[ \frac{\partial L}{\partial x_k} \quad \text{and} \quad \frac{\partial L}{\partial u_k} \] denote the row vectors of partial derivatives

\[ \frac{\partial L}{\partial x_k} \triangleq \begin{bmatrix} \frac{\partial L}{\partial x^1_k} & \cdots & \frac{\partial L}{\partial x^n_k} \end{bmatrix}, \]

\[ \frac{\partial L}{\partial u_k} \triangleq \begin{bmatrix} \frac{\partial L}{\partial u^1_k} & \cdots & \frac{\partial L}{\partial u^m_k} \end{bmatrix}, \]
Before performing the differentiations above, we introduce the Hamiltonian $\mathcal{H} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \rightarrow \mathbb{R}$ defined as

$$\mathcal{H}(x_k, u_k, \lambda_k) \triangleq L(x_k, u_k) + \lambda_k^T f(x_k, u_k) \quad \text{for} \ k = 0, \ldots, N - 1, \quad (18)$$

where $L(\cdot, \cdot)$ is the per-stage weighting in the objective function.
Note that

\[
\frac{\partial H}{\partial x_k} = \frac{\partial L}{\partial x_k} + \lambda_k^T \frac{\partial f}{\partial x_k},
\]

\[
\frac{\partial H}{\partial u_k} = \frac{\partial L}{\partial u_k} + \lambda_k^T \frac{\partial f}{\partial u_k},
\]

where

\[
\frac{\partial L}{\partial x_k} = \begin{bmatrix} \frac{\partial L}{\partial x_1} & \cdots & \frac{\partial L}{\partial x_n} \end{bmatrix},
\]

\[
\frac{\partial L}{\partial u_k} = \begin{bmatrix} \frac{\partial L}{\partial u_1} & \cdots & \frac{\partial L}{\partial u_m} \end{bmatrix}
\]
Then, using the KKT conditions

\[
\frac{\partial L(x^*, \lambda^*)}{\partial x_k} = \frac{\partial H(x_k^*, u_k^*, \lambda_k^*)}{\partial x_k} - (\lambda_{k-1}^*)^T = 0 \quad \text{for } k = 0, \ldots, N - 1,
\]

(19)

\[
\frac{\partial L(x^*, \lambda^*)}{\partial x_N} = \frac{\partial F(x_N^*)}{\partial x_N} - (\lambda_{N-1}^*)^T = 0,
\]

(20)

\[
\frac{\partial L(x^*, \lambda^*)}{\partial u_k} = \frac{\partial H(x_k^*, u_k^*, \lambda_k^*)}{\partial u_k} = 0 \quad \text{for } k = 0, \ldots, N - 1.
\]

(21)
(i) State equations:

\[ x_{k+1}^* = f(x_k^*, u_k^*) \quad \text{for } k = 0, \ldots, N - 1, \]
\[ x_0^* = \bar{x}. \]

(ii) Adjoint equations:

\[ (\lambda_{k-1}^*)^T = \frac{\partial H(x_k^*, u_k^*, \lambda_k^*)}{\partial x_k} \quad \text{for } k = 0, \ldots, N - 1. \]
(iii) Boundary condition:

\[ (\lambda^*_{N-1})^T = \frac{\partial F(x^*_N)}{\partial x_N}. \]

(iv) Hamiltonian condition:

\[
\frac{\partial H(x^*_k, u^*_k, \lambda^*_k)}{\partial u_k} = 0 \quad \text{for } k = 0, \ldots, N - 1. \tag{22}
\]
Notice, in particular, that this means that the minimising control $u^*_k$ is a stationary point of the restricted Hamiltonian defined as

$$\mathcal{H}(x^*_k, u_k, \lambda^*_k) \triangleq L(x^*_k, u_k) + (\lambda^*_k)^T f(x^*_k, u_k) \quad \text{for } k = 0, \ldots, N - 1.$$
In the “constrained” case, necessary conditions for optimality are provided by the *discrete minimum principle*. Although well established for constrained *continuous time* systems, the validity of the minimum principle for constrained *discrete time* systems was a subject of great interest and controversy in the 1960s.
We consider the following optimisation problem.

\[ \mathcal{P}_N(\bar{x}) : \quad \text{minimise } V_N(\{x_k\}, \{u_k\}), \quad (23) \]

subject to:
\[ x_{k+1} = f(x_k, u_k) \quad \text{for } k = 0, \ldots, N - 1, \quad (24) \]
\[ x_0 = \bar{x}, \quad (25) \]
\[ u_k \in U \subset \mathbb{R}^m \quad \text{for } k = 0, \ldots, N - 1, \quad (26) \]
\[ h_N(x_N) = 0, \quad (27) \]

where
\[ V_N(\{x_k\}, \{u_k\}) \triangleq F(x_N) + \sum_{k=0}^{N-1} L(x_k, u_k). \quad (28) \]
Assumptions A

(i) The function $F(x)$ is twice-continuously differentiable.

(ii) For every $u \in \mathbb{U}$, the functions $f(x, u)$ and $L(x, u)$ are twice-continuously differentiable with respect to $x$.

(iii) The terminal constraint function $h_N(x)$ is twice-continuously differentiable and satisfies the “constraint qualification” that the Jacobian matrix $\frac{\partial h_N(x)}{\partial x}$ has full row rank for all $x \in \mathbb{R}^n$. 
(iv) The functions $f(x, u)$ and $L(x, u)$, and all their first and second partial derivatives with respect to $x$, are uniformly bounded on $A \times \mathbb{U}$ for any bounded set $A \subset \mathbb{R}^n$.

(v) The matrix $\partial f(\cdot, \cdot)/\partial x$ is nonsingular on $\mathbb{R}^n \times \mathbb{U}$.

(iv) The set $\left\{ \begin{bmatrix} f(x, u) \\ L(x, u) \end{bmatrix} : u \in \mathbb{U} \right\}$ is convex for all $x \in \mathbb{R}^n$. 
Similarly to the unconstrained case, we define the Hamiltonian as

\[ H(x_k, u_k, \lambda_k, \eta) \triangleq \eta L(x_k, u_k) + \lambda_k^T f(x_k, u_k) \quad \text{for} \ k = 0, \ldots, N - 1, \tag{29} \]

where \( \eta \) is a real number and \( \lambda_k, k = 0, \ldots, N - 1, \) are some vectors in \( \mathbb{R}^n. \)
Subject to Assumptions A, if the sequences \( \{x_0^*, \ldots, x_N^*\} \), \( \{u_0^*, \ldots, u_{N-1}^*\} \) are minimisers of problem \( P_N(\bar{x}) \) defined in (23)–(28), then there exist a sequence of vectors \( \{\lambda_{-1}^*, \ldots, \lambda_{N-1}^*\} \), \( \lambda_k^* \in \mathbb{R}^n \) and a real number \( \eta^* \) such that the following conditions hold:

(i) Adjoint equations:

\[
(\lambda_{k-1}^*)^\top = \frac{\partial H(x_k^*, u_k^*, \lambda_k^*, \eta^*)}{\partial x_k} \quad \text{for } k = 0, \ldots, N - 1. \tag{30}
\]
(ii) Boundary conditions: There exists a real number $\beta \geq 0$ and a vector $\gamma \in \mathbb{R}^\ell$, such that

$$\lambda_{N-1}^* = \left[ \frac{\partial h_N(x_N^*)}{\partial x} \right]^T \gamma + \left[ \frac{\partial F(x_N^*)}{\partial x} \right]^T \beta,$$

$$\eta^* = \beta \geq 0,$$

where $\eta^*$ and $\lambda_{N-1}^*$ are not simultaneously zero. Moreover, if $\eta^* = 0$ in (32), then the vectors $\{\lambda_{-1}^*, \ldots, \lambda_{N-1}^*\}$ satisfying (30)-(31) are all nonzero.
(iii) Minimisation of the Hamiltonian:

$$H(x_k^*, u_k^*, \lambda_k^*, \eta^*) \leq H(x_k^*, u, \lambda_k^*, \eta^*),$$  \hspace{1cm} (33)

for all $k = 0, \ldots, N - 1$ and all $u \in \mathbb{U}$. 

We refer you to the book and the associated references especially those by Halkin. Note that this is an ideal problem to test your understanding of constrained estimation KKT etc.
5. Connections Between the Minimum Principle and the Fritz John and Karush–Kuhn–Tucker Optimality Conditions

We consider the following optimisation problem:

\[ \mathcal{P}_N(\bar{x}) : \quad \text{minimise } V_N(\{x_k\}, \{u_k\}), \quad (34) \]

subject to:

\[ x_{k+1} = f(x_k, u_k) \quad \text{for } k = 0, \ldots, N - 1, \quad (35) \]
\[ x_0 = \bar{x}, \quad (36) \]
\[ g_k(u_k) \leq 0 \quad \text{for } k = 0, \ldots, N - 1, \quad (37) \]
\[ g_N(x_N) \leq 0, \quad (38) \]
\[ h_N(x_N) = 0, \quad (39) \]
\[ V_N(\{x_k\}, \{u_k\}) \triangleq F(x_N) + \sum_{k=0}^{N-1} L(x_k, u_k). \]  

(40)

As before, \( \{x_k\} \triangleq \{x_0, \ldots, x_N\}, x_k \in \mathbb{R}^n \), and \( \{u_k\} \triangleq \{u_0, \ldots, u_{N-1}\}, u_k \in \mathbb{R}^m \), are the state and control sequences.
The functions \( g_k : \mathbb{R}^m \to \mathbb{R}^r \), \( k = 0, \ldots, N - 1 \), represent \( r \) (elementwise) inequality constraints on the input \( u_k \). The functions \( g_N : \mathbb{R}^n \to \mathbb{R}^p \) and \( h_N : \mathbb{R}^n \to \mathbb{R}^\ell \), represent, respectively, inequality and equality constraints on the terminal state. We will assume that all functions are differentiable functions of their variables and that \( f \) and \( h_N \) are continuously differentiable at the optimal solution.
We will derive necessary optimality conditions for the sequences \( \{x_0^*, \ldots, x_N^*\} \) and \( \{u_0^*, \ldots, u_{N-1}^*\} \) to be minimisers of the optimisation problem using the FJ necessary optimality conditions. We observe that, in contrast with the “unconstrained” case where the linear independence constraint qualification required by the KKT conditions holds for all feasible points, here a constraint qualification would need to be imposed if we were to use the KKT conditions as necessary conditions for optimality. On the other hand, the FJ conditions are always a necessary condition for optimality under the differentiability assumption, without requiring any constraint qualification.
Recalling the vector definition

\[ \mathbf{x} \triangleq \begin{bmatrix} x_0^T & \cdots & x_N^T & u_0^T & \cdots & u_{N-1}^T \end{bmatrix}^T \in \mathbb{R}^{(N+1)n+Nm}, \]

we can express problem in the form

\[
\begin{aligned}
& \text{minimise } \phi(\mathbf{x}), \\
& \text{subject to:} \\
& g(\mathbf{x}) \leq 0, \\
& h(\mathbf{x}) = 0, \\
\end{aligned}
\]

(41)
\[ \phi(x) \triangleq F(x_N) + \sum_{k=0}^{N-1} L(x_k, u_k), \quad (42) \]

\[ h(x) \triangleq \begin{bmatrix} \bar{x} - x_0 \\ f(x_0, u_0) - x_1 \\ \vdots \\ f(x_{N-1}, u_{N-1}) - x_N \\ h_N(x_N) \end{bmatrix}, \quad (43) \]

\[ g(x) \triangleq \begin{bmatrix} g_0(u_0) \\ g_1(u_1) \\ \vdots \\ g_{N-1}(u_{N-1}) \\ g_N(x_N) \end{bmatrix}. \quad (44) \]
We will use the Hamiltonian

\[ H(x_k, u_k, \lambda_k, \eta) \triangleq \eta L(x_k, u_k) + \lambda_k^T f(x_k, u_k) \quad \text{for } k = 0, \ldots, N - 1, \]

(45)
The Fritz John conditions immediately yield the following necessary conditions. There exist a scalar $\eta^*$ and vectors $\{\lambda^*_{-1}, \ldots, \lambda^*_N\}$, $\gamma^*$, $\{\nu^*_0, \ldots, \nu^*_N\}$, not all zero, such that the following conditions hold:

(i) Adjoint equations:

\[
(\lambda^*_k) = \frac{\partial H(x_k^*, u_k^*, \lambda_k^*, \eta^*)}{\partial x_k} \quad \text{for } k = 0, \ldots, N - 1. \tag{46}
\]
(ii) Boundary conditions:

\[
\lambda^*_{N-1} = \left[ \frac{\partial h_N(x_N^*)}{\partial x_N} \right]^T \gamma^* + \left[ \frac{\partial F(x_N^*)}{\partial x_N} \right]^T \eta^* + \left[ \frac{\partial g_N(x_N^*)}{\partial x_N} \right]^T \nu_N^*, \quad (47)
\]

\[
(\nu_N^*)^T g_N(x_N^*) = 0, \quad (48)
\]

\[
\eta^* \geq 0, \quad \nu_N^* \geq 0. \quad (49)
\]
(iii) Hamiltonian conditions:

\[
\begin{align*}
\left[ \frac{\partial H(x_k^*, u_k^*, \lambda_k^*, \eta^*)}{\partial u_k} \right]^\top + \left[ \frac{\partial g_k(u_k^*)}{\partial u_k} \right]^\top & \nu_k^* = 0, \\
(\nu_k^*)^\top g_k(u_k^*) & = 0,
\end{align*}
\]

for \( k = 0, \ldots, N - 1. \)

for \( k = 0, \ldots, N - 1. \)
We next study necessary conditions for the above Hamiltonian conditions to hold. Our goal is to directly apply the standard KKT conditions.
We need additional mild convexity assumptions. In particular, suppose that $\mathcal{H}(x_k^*, u_k, \lambda_k^*, \eta^*)$ is pseudoconvex at $u_k^*$, and the constraint function $g_k(u_k)$ in (53) is quasiconvex at $u_k^*$. We can then apply the KKT sufficient optimality conditions to conclude that the Hamiltonian conditions imply:
\[ \mathcal{H}(x_k^*, u_k^*, \lambda_k^*, \eta^*) \leq \mathcal{H}(x_k^*, u_k, \lambda_k^*, \eta^*) \text{ for all } u_k \text{ such that } g_k(u_k) \leq 0. \]  

(53)

for \( k = 0, \ldots, N - 1 \), where \( \mathcal{H}(x_k^*, u_k, \lambda_k^*, \eta^*) \) is the restricted Hamiltonian

\[ \mathcal{H}(x_k^*, u_k, \lambda_k^*, \eta^*) = \eta^* L(x_k^*, u_k) + (\lambda_k^*)^T f(x_k^*, u_k). \]  

(54)
Finally, we observe that, if the functions $h(x)$ and $g(x)$, which define the constraints for the (vector form of the) original problem, satisfy a constraint qualification, then we can apply the KKT necessary optimality conditions to the original problem, that is, we can set $\eta = 1$ in the FJ conditions and in the Hamiltonian.
The necessary conditions for optimality developed above are not, in general, sufficient. However, it is possible to obtain sufficient conditions by using the following *principle of optimality*: “Any part of an optimal trajectory must itself be optimal.”
We consider the optimisation problem as before. To illustrate the idea, we assume that, in addition to the equality constraints provided by the state equations, only the input is constrained to belong to a set $U \subset \mathbb{R}^m$, (no equality constraint on the terminal state). Clearly, the partial value function at time $N$ is

\[ V_0^{\text{opt}}(x_N) = F(x_N), \quad (55) \]
The principle of optimality states that we must satisfy the following sufficient condition at each time $k$:

$$V_{N-k}^{\text{OPT}}(x_k) = \min_{u_k \in U} \left\{ V_{N-(k+1)}^{\text{OPT}}(f(x_k, u_k)) + L(x_k, u_k) \right\}. \quad (56)$$
For some simple problems, such as for linear quadratic unconstrained optimal control, $V_{N-k}^{\text{OPT}}(x_k)$ has a finite parameterisation. However, in general, $V_{N-k}^{\text{OPT}}(x_k)$ will be a complicated function of the states. In the latter case, we really have no option but to store $V_{N-k}^{\text{OPT}}(x_k)$ for all possible values of $x_k$. 
We begin with $V^\text{OPT}_0(x_N)$. Then, we can evaluate the optimiser $u^*_N$ and hence $V^\text{OPT}_1(x_{N-1})$ for all possible values of $x_{N-1}$. We continue backwards until we finally reach $V^\text{OPT}_N(x_0)$. Having reached this point, we realise that we actually do know $x_0$, and we can thus find $u^*_0$. Then, running forward in time, we can calculate $x^*_1$ and, provided we have recorded $u^*_1$ for all possible $x_1$, we can then proceed to $x^*_2$ and so on.
The above procedure can be applied, in principle, to very general problems assuming a finite number of admissible states. This can be achieved, for example, by quantisation of the admissible region in the state space. For problems with large horizons and high state space dimension, however, the number of computations and the storage required by the dynamic programming technique may be impractical, or even prohibitive. Bellman called this problem the “curse of dimensionality.” For some special problems, however, the technique can be successfully applied, even to obtain closed form expressions for the optimal constrained control law.