Fuel Consumption in Optimal Control
J. Redmond and L. Silverberg

Reprinted from
Journal of Guidance, Control, and Dynamics
Volume 15, Number 2, March-April 1992, Pages 424-430
Multiplying Eq. (3) by any $\eta$ in $V$ yields

$$\eta^T y(T) = \int_0^T g(\eta,t) u(t) \, dt$$

(7)

in which

$$g(\eta,t) = (\eta^T e^{-AT} B)^T$$

(8)

By convexity theorem 1, if a control maximizes Eq. (7) over $U$, then it produces a reachable state on the boundary of $R$. Upon maximizing Eq. (7), the controls that produce reachable states on the boundary of $R$ are expressed as a function of $\eta$. The control that produces $z_1$ in convexity theorem 2 can then be expressed as $u_1 = u(\eta)$. From Eq. (3), the control that produces $y_1$ is $u_1/\alpha$, since $\alpha y_1 = z_1$. Indeed, any reachable state in the interior of $R$ can as well be expressed as a function of $\eta$ by its extension to the boundary of $R$.

Now define the cost function $C(u)$ that in turn is used to define the admissible set as follows:

$$U = \{ u : C(u) \leq C_{\text{max}} \}$$

(9)

From Eq. (9), the admissible set bounds the cost by $C_{\text{max}}$. Let us assume that the cost function satisfies the multiplicative property

$$C(\beta u) = |\beta| C(u)$$

(10)

for any $\beta$ and for any control $u$. This property is recognized as the third property associated with the definition of a norm.\textsuperscript{7} We now show that the control that produced $y_1$, given previously by

$$u^* = u_1/\alpha$$

(11)

is the optimal control, that is, minimizes $C(u)$ over all $u$ in $U$ that produce $y_1$. Toward that end, we note from Eq. (11) that $C(u^*) = C(u_1)/\alpha = C_{\text{max}}/\alpha$ since $u_1$ is on the boundary of $U$ from which it follows that $C(u_1) = C_{\text{max}}$. Let us now consider any other control $u_2$ in $U$ that produces $y_1$ and show that $C(u_2) \geq C_{\text{max}}/\alpha$. To this end, we define $u_2 = u_1[C_{\text{max}}/C(u_2)]$. Although it does not produce $y_1$, $u_2$ is admissible since $C(u_2) = C(u_1)[C_{\text{max}}/C(u_2)]$. Thus, the state attained by $u_2$ is contained in the reachable set. From Eq. (3),

$$\int_0^T e^{-AT} Bu_2(t) \, dt = C_{\text{max}}/C(u_2)$$

$$= y_1 C_{\text{max}}/C(u_2)$$

But $y_1 C_{\text{max}}/C(u_2) \leq \alpha y_1$, since $\alpha y_1$ is on the boundary of the reachable set. It follows that $C(u_2) \geq C_{\text{max}}/\alpha$, and so $C(u_2) \geq C(u^*)$.

### III. Some Classical Cost Functions

The following describes three optimal control problems. In each problem, a control that produces a reachable state on the boundary is constructed and then expressed as a function of $\eta$. The construction of the optimal control follows directly.

#### Bang-Bang Control

Consider the case of minimizing the largest control input. This can be regarded as a saturation condition. The cost function is

$$C_i(u) = \max_{1 \leq j \leq m} \sup_{0 \leq t \leq T} |u_j(t)|$$

(12)

in which $u = [u_1, u_2, \ldots, u_m]^T$. From Eq. (9), the set of admissible controls is

$$U_i = \{ u : \max_{1 \leq j \leq m} \sup_{0 \leq t \leq T} |u_j(t)| \leq C_{\text{max}} \}$$

(13)
The control that produces reachable states on the boundary of $R_0$ is given by

$$u_i(t) = \sum_{j=1}^{M_i} c_{ij} \text{sgn}(g_j, (\tau_j, \eta)) \delta(t - \tau_j), \quad 1 = \sum_{j=1}^{M_i} c_{ij}$$  \hspace{1cm} (32)

As stated previously, although any control of the form of Eq. (32) will map the system to the boundary of the reachable set, only certain choices of $c_{ij} (i = 1, 2, \ldots, M_j, j = 1, 2, \ldots, m)$ yield $\eta = \eta_i$. Thus, care must be taken to choose a set of $c_{ij}$ that results in a controller that drives the system from $x_0$ to $x_1$. We obtain by convexity theorem 2,

$$F(\eta) = \sum_{j=1}^{m} \sup_{0 \leq s \leq T} g(s), \quad \alpha = \min F(\eta) = F(\eta_i)$$  \hspace{1cm} (33)

The optimal control is given by Eq. (11) in which $\eta_i$ and $\alpha$ are determined by Eq. (33).

IV. Control of a Damped Harmonic Oscillator

Differences among the solutions given in the previous section are brought out by considering the damped harmonic oscillator. The motion of the oscillator is described by the differential equation

$$m\ddot{x}(t) + c\dot{x}(t) + kx(t) = u(t)$$  \hspace{1cm} (34)

where $m$ denotes mass, $c$ denotes damping, and $k$ denotes stiffness. Equation (34) is recast in the form of Eq. (1) in which

$$\frac{c}{m} = 2\alpha, \quad \frac{k}{m} = \omega^2, \quad x(t) = \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix}$$

leading to the coefficient matrices

$$A = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\alpha \end{bmatrix}, \quad B = \begin{bmatrix} 0 \\ 1/m \end{bmatrix}$$

This system is controllable. The state transition matrix of the damped harmonic oscillator and the inverse at time $T$ are

$$e^{AT} = e^{-\omega T} \begin{bmatrix} \cos \beta T + \frac{\alpha}{\beta} \sin \beta T & \frac{1}{\beta} \sin \beta T \\ -\frac{\alpha + \beta}{\beta} \sin \beta T & \cos \beta T - \frac{\alpha}{\beta} \sin \beta T \end{bmatrix}$$

$$e^{-AT} = e^{\alpha T} \begin{bmatrix} \cos \beta T - \frac{\alpha}{\beta} \sin \beta T & -\frac{1}{\beta} \sin \beta T \\ \frac{\alpha + \beta}{\beta} \sin \beta T & \cos \beta T + \frac{\alpha}{\beta} \sin \beta T \end{bmatrix}$$  \hspace{1cm} (35)

where $\beta = \sqrt{\alpha^2 - \omega^2}$.

It is desired to drive the system from some initial displacement, $x(0) = [x_0, 0]^T$, to the origin, $x(T) = [0, 0]^T$, in some time $T$. From Eq. (3), we compute the desired reachable state

$$y(T) = e^{-AT}x(T) - x(0) = \begin{bmatrix} -x_0 \\ 0 \end{bmatrix}$$  \hspace{1cm} (36)

Since $n = 2$, we write $\eta = [\eta_1, \eta_2]^T$. From Eq. (8), the function $g(\eta, t)$ is reduced to

$$g(\eta, t) = De^{\alpha t} \sin(\beta t + \Phi)$$  \hspace{1cm} (37)

where

$$D = [(\eta_1/\beta + \eta_2/\alpha)^2 + \eta_3^2]^{1/2}/m$$

$$\Phi = \tan^{-1} \left[ \frac{\eta_2}{\eta_1 + \eta_2/\alpha} \right]$$  \hspace{1cm} (38)

For further reference, notice that $g(\eta, t)$ is an exponentially growing oscillatory function of time. Consistent with convexity theorem 2, we define the hyperplane

$$H = \left\{ \eta = [\eta_1, \eta_2]^T: \eta_1 = \frac{-1}{\eta_0} \right\}$$  \hspace{1cm} (39)

Bang-Bang Control

Consider the cost function and corresponding boundary control given in Eqs. (11) and (15), respectively. Substituting $g(\eta, t)$ from Eq. (38) into Eq. (16) yields

$$F(\eta) = D \int_0^T e^{\alpha t} \sin(\beta t + \Phi) dt$$  \hspace{1cm} (40)

for $\alpha \ll \beta$, the minimum value of Eq. (40) over the hyperplane occurs when $\eta_2$ is zero. Thus, from Eq. (40)

$$\bar{\alpha} = \frac{1}{x_0 \beta m} \int_0^T e^{\alpha t} \sin \beta t dt$$  \hspace{1cm} (41)

From Eq. (11), the optimal control is of the form

$$u(t) = \frac{1}{\bar{\alpha}} \text{sgn}(\sin \beta t)$$  \hspace{1cm} (42)

For example, let $m = k = 1$, $x_0 = 1$, and let the decay rate $\alpha = 0.03$. The selection of the final time $T$ depends on the admissible set of controls. Let $T = 6\pi/\beta$, representing the period of time of three oscillations, and assume $C_{\text{max}}$ is large enough that the optimal control is within $U$. Figure 3a shows the control input, revealing the bang-bang nature of the control. This is also characteristic of the minimum time solution for the same problem with bounded control inputs. Indeed, if the final time $T$ is chosen so as to place the bang-bang control on the boundary of the admissible set, that is, $C(u) = C_{\text{max}}$, then the minimum time solution is obtained. The phase diagram shown in Fig. 3b confirms that the system is driven to the origin in three oscillations.
From Eq. (8)

$$g(\eta, t) = -\frac{\eta_1 t}{m} + \frac{\eta_2}{m} = -\frac{t}{m_\tau} + \frac{\eta_2}{m}$$

(53)

**Bang-Bang Control**

Considering the cost function of Eq. (12), we obtain from Eqs. (53) and (16)

$$F(\eta) = \int_0^T \left[ -\frac{t}{m_\tau} + \frac{\eta_2}{m} \right] \, dt$$

(54)

The minimization of Eq. (54) over the hyperplane is accomplished by distinguishing between three cases depending on the sign of the integrand (positive, negative, and sign change). We obtain $\eta_2 = T/2m_\tau$ and

$$\tilde{\alpha} = \frac{T^2}{4m_\tau}$$

(55)

The optimal control is

$$u^*(t) = \frac{4m_\tau}{T^2} \text{sgn} \left( \frac{T}{2} - t \right)$$

(56)

For example, let $m = 1$, $T = 1$, and $x_\tau = 1$. The associated control and the system response are shown in Fig. 6. Note that choosing $T$ to place the control on the boundary of $U_i$ corresponds to the minimum time solution, as was the case with the damped oscillator.

**Continuous Control**

Consider the cost function of Eq. (18) for $p = q = 2$. Repeating the previous steps we obtain

$$F(\eta) = \frac{1}{m} \left[ \frac{T^3}{3x_\tau^2} - \frac{\eta_2 T^2}{x_\tau} + \eta_2 T \right]$$

(57)

The minimum of Eq. (57) occurs when $\eta_2 = T/2x_\tau$, yielding

$$\tilde{\alpha} = \frac{1}{m} \left( \frac{T^3}{12x_\tau^2} \right)^{\frac{1}{2}}$$

(58)

The optimal control becomes

$$u^*(t) = \frac{12x_\tau m}{T^3} \left[ \frac{T}{2} - t \right]$$

(59)

Figure 7 shows the optimal control forces and the system response. The linearity of the control yields a smooth transition of the system from the initial state to the final state.

**Impulse Control**

Consider the cost function given in Eq. (25). From Eq. (30)

$$F(\eta) = \sup_{0 \leq t \leq T} \left| -\frac{t}{x_\tau m} + \frac{\eta_2}{m} \right|$$

(60)

From Eq. (60) and convexity theorem 2

$$\tilde{\alpha} = \min_{\eta_2} \sup_{0 \leq t \leq T} \left| -\frac{t}{x_\tau m} + \frac{\eta_2}{m} \right|$$

(61)

Considering three possible ranges of the argument (positive, negative, sign change), the minimum of Eq. (60) occurs when $\eta_2 = T/2x_\tau$, yielding

$$\tilde{\alpha} = \frac{T}{2x_\tau m}$$

(62)

This minimum occurs both at $\tau_1 = 0$ and $\tau_2 = T$. From Eq. (29) or (32) with $N$ or $M = 2$, it is determined that a uniform distribution of the impulsive forces will drive the system from $x_0$ to $x_1$. Thus, letting $c_1 = c_2 = \frac{1}{2}$ in Eq. (32), the optimal control is

$$u^*(t) = \frac{x_\tau m}{T} \text{sgn} \left( \frac{T}{2} - t \right) [\delta(t) + \delta(t - T)]$$

(63)

This solution consists of one initial impulse to impart a velocity to the system as shown in Fig. 8. Then the system drifts to the desired position at time $T$, at which time a final impulse terminates the motion, leaving the system at rest.