INTRODUCTION

Systems Optimization - Definition

- The general definition of optimization is to obtain the "best" solution to a problem by choosing certain design variables which maximize or minimize a given (or chosen) criterion for "best".

- The purpose is to formulate mathematical models of the system and maximize, or minimize, some "objective" function subject to some equality and/or inequality constraints on the pertinent variables.

- Optimization is fundamental to design but the two processes are <u>not</u> equivalent. Design seeks a solution to a certain synthesis problem and the solution is generally not unique.
  - Optimization assumes that some form of solution exists, possibly via the design process, and that this solution is specified using a finite set of design variables.
  - The values of the design variables may be restricted to some range and this range is called the *feasible region.*

- We cast the optimization problem in mathematical form.
  - these techniques are usually algorithmic in nature and must be implemented on a computer.

- The difference is that practical systems may not always have well defined mathematical formulations which take into account all the features of that system and therefore we usually refer to a mathematical formulation as a *model* for the system.
- Modelling error is the difference between how the model performs and how the actual physical system performs given certain values for the design variables and the parameters which produce the best approximation of the system using that model.

- Example: model the voltage across a resistor as being proportional to the current through it and find that the constant of proportionality (*i.e.* resistance) is given by 10 [Ω].

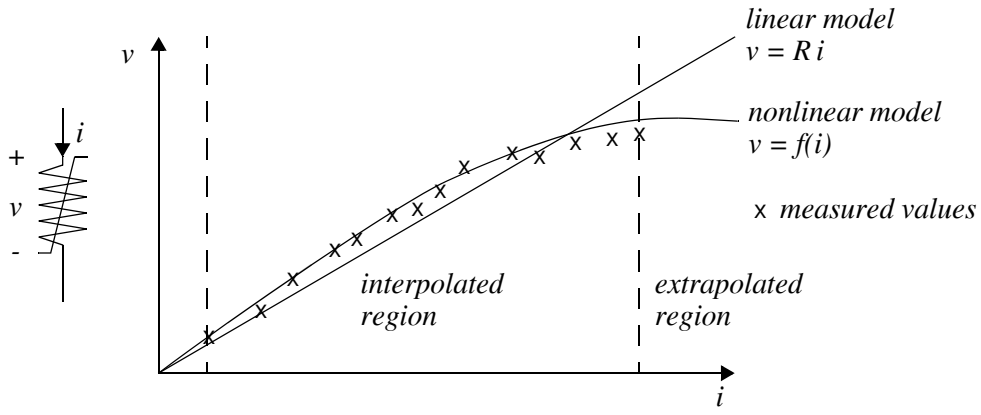

Figure 1. linear and nonlinear models for resistance

Mathematical Notation

- The *objective function,* $f(x)$, will generally be a function of n scalar variables, that is $x \in \subsetneq^n$, and we denote the individual elements of x using vector notation $x = [x_1, x_2, \ldots, x_n]^T$. The vector x is called the *design vector* and the *optimization problem* is written as follows:

$$\text{Minimize } f(x)$$

$$\text{subject to: } g_j(x) \leq 0 \quad j = 1(1)J$$

$$h_k(x) = 0 \quad k = 1(1)K$$

  - $f(x)$ is the objective function (or sometimes called the *cost function*);
  - x is the *design vector* of n variables;
  - $g_j(x) \leq 0$ are the J *equality constraints*;
  - $h_k(x) = 0$ are the K *inequality constraints*.

- The functions g and h define the constraints of the optimization problem. The constraints, in turn, define the *feasibility region* of the optimization problem which is the region of allowed x values where the optimum (*i.e.* maximum or minimum) value of the objective function is to be found.

- It is at once clear that $K < n$, otherwise there may be only one, or possibly no feasible vector x.

- It is only necessary to speak of minimization since if the problem requires maximization then we can convert the problem to a minimization problem by multiplying the objective function by *-1*. That is

$$\text{minimize} \quad -f(x) \equiv \text{maximize} -f(x)$$

Notice that although the minimum of $-f(x)$ and the maximum of $f(x)$ are located at x^*, the values of each function are not the same, that is $f(x^*) \neq -f(x^*)$.

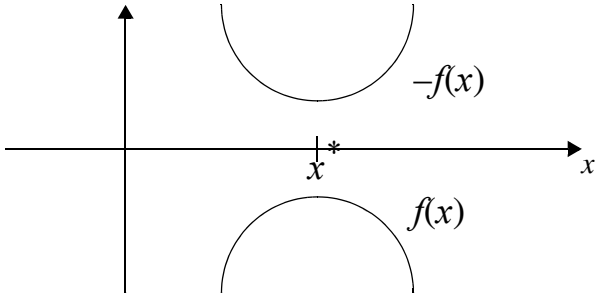

Figure 2. Both the minimum of $-f(x)$ and the maximum of $f(x)$ are located at x^*.

Classification of Optimization Problems

- The function $f(x)$ may be any valid mathematical function or indeed it may even be an algorithm or physical measurement of a process which depends on n variables.

- The method of solution will depend on the exact nature of the objective function as well as on the exact nature of the equality and/or inequality constraints, $g(x)$ and $h(x)$.

    1. single variable optimization - $n = 1$
    2. unconstrained optimization - no constraints on x
    3. linear program - $f(x)$, $g(x)$, and $h(x)$ are all linear functions
    4. quadratic program - $f(x)$ is a quadratic function, $g(x)$ and $h(x)$ are linear functions
    5. linearly constrained nonlinear programs
    6. nonlinear programs - $f(x)$, $g(x)$, and $h(x)$ are nonlinear functions

- Different techniques are used for the different problems; each technique tries to exploit the structure inherent in the problem it was meant for. Clearly, the last classification is the most difficult to solve.

Examples

Minimization of power in resistive circuits

- Consider the current divider circuit of Figure 3. where i_0, R_1 and R_2 are given and we want to determine the branch currents i_1 and i_2.

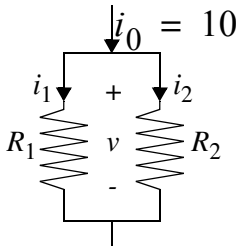

Figure 3. Current divider circuit.

- The laws of physics are such that the power lost in the resistors will be the minimum possible given that the sum of the two resistor currents will have to be equal to 10. Thus, if we are given that the power lost in a resistor is equal to Ri^2 we can state the problem as:

$$\text{minimize } f(i) = R_1 i_1^2 + R_2 i_2^2 \text{ (Power lost in the circuit)}$$

$$\text{subject to: } i_1 + i_2 = i_0 = 10 \text{ (KCL)}$$

  - and we recognize this as a quadratic program.

- One method of finding its solution is to use the equality constraint to eliminate one of the variables:

$$f(i_1) = R_1 i_1^2 + R_2 (10 - i_1)^2 = (R_1 + R_2) i_1^2 - 20 i_1 R_2 + 100 R_2$$

and find the minimum by setting

$$\frac{d}{di_1} f(i_1) = 2(R_1 + R_2)i_1 - 20R_2 = 0$$

which gives

$$i_1 = \frac{20R_2}{2(R_1 + R_2)} = \frac{10R_2}{(R_1 + R_2)}$$

or, for general i_0,

$$i_1 = \frac{i_0 R_2}{R_1 + R_2}$$

and substituting back into the equality constraint, we have

$$i_2 = (i_0 - i_1) = i_0\left(1 - \frac{R_2}{R_1 + R_2}\right) = \frac{i_0 R_1}{R_1 + R_2}.$$

Thus we have derived the *current division law* by an "indirect" procedure of minimizing the power dissipated in the circuit.

Production-inventory problem

Each year, denoted by (i), a company produces x_i items with a cost of production given by $f_i(x_i)$. The demand for the items is d_i and the cost of holding inventory is $c_i(y_i)$ where y_i is the amount of inventory. There is a maximum production capacity of X items for any year and the inventory cannot exceed Y items. The initial inventory is 0 and we wish to formulate the nonlinear programming problem for minimizing total cost over n years.

The total cost is given by

$$C(x, y) = \sum_{i=1}^{n} [f_i(x_i) + C_i(y_i)]$$

and the inventory in any year i by

$$y_i = y_{i-1} + x_i - d_i$$

where $y_0 = 0$ and $i = 1\ (1)\ n$. Therefore the nonlinear programming problem is written as

$$\underset{x_i\, y_i}{\text{minimize}} \sum_{i=1}^{n} [f_i(x_i) + C_i(y_i)]$$

$$\text{subject to: } y_i = y_{i-1} + x_i - d_i,\ y_0 = 0$$

$$x_i \le X \qquad y_i \le Y$$

The last two constraints are, possibly nonlinear, inequality constraints.

If we eliminate x_i using the equality constraints

$$x_i = y_i - y_{i-1} + d_i,\ i = 1(1)n$$

then the problem becomes

$$\underset{y_i}{\text{minimize}} \sum_{i=1}^{n} [f_i(y_i - y_{i-1} + d_i) + C_i(y_i)]$$

$$\text{subject to:} \quad (y_i - y_{i-1} + d_i) \le X$$

$$y_0 = 0$$

$$y_i \le Y$$

- Notice that if we want to minimize the total cost over *say* 10 *years*, then in the second formulation we would have a 10 variable minimization problem whereas in the first formulation we would have a 20 variable minimization problem to solve.

- This reduction in the number of design variables can be an important factor in reducing the computer time required to find the solution.

- Also note that in this problem, as we choose a longer term n, over which to minimize the cost, we inherently have more "uncertainty" which will exist in the demand d_i, inventory costs $C_i(y_i)$, and the costs of production $f_i(x_i)$.

- If the inventory and production costs are linear functions, say $f_i(x_i) = a_i x_i$ and $C_i(y_i) = b_i y_i + c_i$, then the problem becomes a linear program.

Analysis and data reduction

- Consider an experiment wherein we vary an independent variable x, and measure a resulting outcome y.

- We want

$$y = F(x, \theta_1, \theta_2)$$

where θ_1 and θ_2 are undetermined parameters.

- Now, say we run the experiment and vary the independent variable x and measure y n times. How should we determine the values of the parameters θ_1 and θ_2 given this data?

- One criterion for the fit is the least squares criterion, given as:

$$\text{minimize } L(\theta_1, \theta_2) = \sum_{i=1}^{n} [y_1 - f(x_1, \theta_1, \theta_2)]^2$$

Thus, we try to minimize this "goodness of fit" criterion, $L(\theta_1, \theta_2)$, by choosing appropriate θ_1, θ_2 parameters. The difference between the experimental result and the model's prediction for specific values of the independent variable and the parameters:

$$y_i - f(x_i, \theta_1, \theta_2)$$

is called the residual and we are minimizing the sum of the squares of the residuals.

Single Variable Optimization

Definition a function $f(x)$ is said to be *unimodal*, with a *minimum* located at x^*, on the interval $S = \{x | a \le x \le b\}$ if and only if for all $x_1, x_2 \in S$. there exists an x^* such that if $x^* < x_1 < x_2$ then $f(x^*) < f(x_1) < f(x_2)$ or if $x^* > x_1 > x_2$ then $f(x^*) < f(x_1) < f(x_2)$.❐

Definition a function $f(x)$ is said to be *unimodal*, with *maximum* located at x^*, on the interval $S = \{x | a \le x \le b\}$ if and only if for all $x_1, x_2 \in S$, there exists an x^* such that if then $x^* < x_1 < x_2$ then $f(x^*) > f(x_1) > f(x_2)$ or if $x^* > x_1 > x_2$ then $f(x^*) > f(x_1) > f(x_2)$.❐

- Note that in both these definitions, $f(x)$ may be continuous or discontinuous and these are more general than definitions using the derivative of the function. The definitions above are sometimes said to be for *strict* unimodality. If the strict inequalities are relaxed to include equality, then the definitions refere to non-strict unimodality or just unimodality.

Definition the function $f(x)$ has a *global minimum (maximum)* at x^{**}, on the interval $S = \{x | a \le x \le b\}$, if and only if for all $x \in S$, $f(x^{**}) \le (\ge) f(x)$, with $x^{**} \in S$.❒

Definition the function $f(x)$ has a *local minimum (maximum)* at x^*, on the interval $S = \{x | a \le x \le b\}$ if and only if there exists an $\varepsilon > 0$, such that for all x satisfying $|x - x^*| < \varepsilon$, $f(x^*) \le (\ge MP) f(x)$.❒

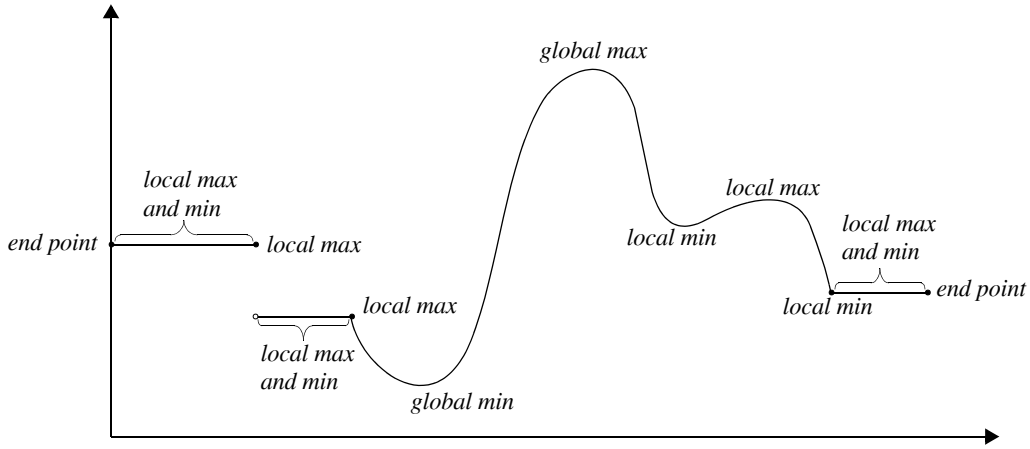

Figure 4. Classification of minimums and maximums for a general function

- In general, it is more difficult to find global extremas than local ones. Thus, most of the methods which have been developed and the ones which we will be dealing with are for finding local extremas.

Definition the point x^* is called a *stationary point* of the function $f(x)$ if $f'(x^*) = 0$.❐

Here we are using the notation $\left.\frac{df}{dx}\right|_{x = x^*} = f'(x^*)$.

Definition an *inflection* or *saddle point* is a stationary point which is not a local maximum or minimum.❐

Theorem (necessary condition for extremum)
if x^* is a local minimum (maximum) of $f(x)$ on $S = \{x \mid a \le x \le b\}$ then if $f(x)$ is differentiable at $x = x^*$ then it is a stationary point, that is $f'(x^*) = 0$. If $f(x)$ is twice differentiable, then $f''(x^*) \ge 0$ $(f''(x^*) \le 0)$.❐

Notice the non-strict inequality in the above theorem as compared to the following theorem giving a sufficient condition for a local extremum.

Theorem (sufficient condition for extremum)
given a stationary point x^* such that the first nonzero higher order derivative is of order n, then if

(i) n is odd $\Rightarrow x^*$ is an inflection point;

(ii) n is even, then x^* is a local optimum such that

a) $\left.\frac{d^n f}{dx^n}\right|_{x^*} > 0 \Rightarrow x^*$ is a local minimum

b) $\left.\frac{d^n f}{dx^n}\right|_{x^*} < 0 \Rightarrow x^*$ is a local maximum❒

Proof: expand $f(x)$ in a Taylor's series about x^*

$$f(x^* + \varepsilon) - f(x^*) = \left.\varepsilon\frac{df}{dx}\right|_{x^*} + \left.\frac{\varepsilon^2}{2!}\frac{d^2 f}{dx^2}\right|_{x^*} + \ldots + \left.\frac{\varepsilon^n}{n!}\frac{d^n f}{dx^n}\right|_{x^*} + O_{n+1}(\varepsilon)$$

At a stationary point $\left.\frac{d}{dx}\right|_{x^*} = 0$ and if the first non-zero higher order derivative is the n^{th} one then we have

$$f(x^* + \varepsilon) - f(x^*) = \left.\frac{\varepsilon^n}{n!}\frac{d^n f}{dx^n}\right|_{x^*} + O_{n+1}(\varepsilon)$$

Since we can make $|\varepsilon|$ arbitrarily small, the first term on the right dominates, and therefore, if n is even this implies that

x^* is a local minimum if $\left.\frac{d^n}{dx^n}\right|_{x^*} > 0$ and that

x* is a local maximum if $\left.\frac{d^n f}{dx^n}\right|_{x^*} < 0$

since $\varepsilon^n > 0$. On the other hand, if n is odd, then this implies that x^* is an inflection point since we can choose ε and thus ε^n as positive or negative.■

Example - constrained continuous function

Consider the function $f(x) = \sin x$ with feasible region given as $S = \{x | 0 \le x \le 2\pi\}$. The stationary points are given by $f'(x^*) = \cos(x^*) = 0 \Rightarrow x^* = \{\pi/2, 3\pi/2\}$. The values of the second derivative at these stationary points are $f''(x^*) = -\sin(x^*) = \{-1, 1\}$ and therefore, by the above theorem we can say that $x^* = \pi/2$ is a local maximum and $x^* = 3\pi/2$ is a local minimum. By evaluating the function at the two endpoints $x = \{0, 2\pi\}$, and by the fact that it is a continuous function, we see that these local extrema are in fact global extrema. Thus, we see that for constrained continuous functions we only need to check all local extrema and the endpoints to determine the global extrema.■