3M1 LECTURE NOTES — OPTIMIZATION TECHNIQUES

Books

- A. Antoniou and W-S. Lu "Practical Optimization: Algorithms and Engineering Applications", Springer ER 227
- P.E. Gill, W. Murray and M.H. Wright "Practical Optimization", Academic Press ER 115
- D.G. Luenberger and Y. Ye "Linear and Non-linear Programming", Springer ER 239

Introduction

> *"'What's new?' is an interesting and broadening eternal question, but one which, if pursued exclusively, results only in an endless parade of trivia and fashion, the silt of tomorrow. I would like, instead, to be concerned with the question 'What is best?', a question which cuts deeply rather than broadly, a question whose answers tend to move the silt downstream."*
>
> Robert M. Pirsig "Zen and the Art of Motorcycle Maintenance" (1974)

Mathematical optimization is the formal title given to the branch of computational science that seeks to answer the question 'What is best?' for problems in which the quality of any answer can be expressed as a numerical value. Such problems arise in all areas of mathematics, the physical, chemical and biological sciences, engineering, architecture, economics, and management, and the range of techniques available to solve them is nearly as wide.

The primary objective of this course is to provide a broad overview of standard optimization techniques and their application to practical problems.

Aims

When using optimization techniques, one must:

- understand clearly where optimization fits into the problem;
- be able to formulate a criterion for optimization;
- know how to simplify a problem to the point at which formal optimization is a practical proposition;
- have sufficient understanding of the theory of optimization to select an appropriate optimization strategy, and to evaluate the results which it returns.

1. Definitions

The goal of an optimization problem can be stated as follows: find the combination of parameters (independent variables) which optimize a given quantity, possibly subject to some restrictions on the allowed parameter ranges. The quantity to be optimized (maximized or minimized) is termed the *objective function*; the parameters which may be changed in the quest for the optimum are called *control* or *decision variables*; the restrictions on allowed parameter values are known as *constraints*.

A maximum of a function f is a minimum of $-f$. Thus, the general optimization problem may be stated mathematically as:

$$\begin{array}{lll} \text{minimize} & f(\mathbf{x}), & \mathbf{x} = (x_1, x_2, \dots, x_n)^T \\ \text{subject to} & c_i(\mathbf{x}) = 0, & i = 1, 2, \dots, m' \\ & c_i(\mathbf{x}) \geq 0, & i = m'+1, \dots, m. \end{array} \tag{1.1}$$

where $f(\mathbf{x})$ is the objective function, $\mathbf{x}$ is the column vector of the n independent control variables, and $\{c_i(\mathbf{x})\}$ is the set of constraint functions. Constraint equations of the form $c_i(\mathbf{x}) = 0$ are termed *equality constraints*, and those of the form $c_i(\mathbf{x}) \geq 0$ are *inequality constraints*. Taken together, $f(\mathbf{x})$ and $\{c_i(\mathbf{x})\}$ are known as the *problem functions*. If inequality constraints are simply a restriction on the allowed values of a control variable, e.g. minimum and maximum possible dimensions:

$$x_{i\min} \leq x_i \leq x_{i\max}, \tag{1.2}$$

these are known as *bounds*.

2. Classifications

There are many optimization algorithms available to the engineer. Many methods are appropriate only for certain types of problems. Thus, it is important to be able to recognize the characteristics of a problem in order to identify an appropriate solution technique. Within each class of problems there are different minimization methods, varying in computational requirements, convergence properties, and so on. Optimization problems are classified according to the mathematical characteristics of the objective function, the constraints and the control variables.

Probably the most important characteristic is the *nature* of the objective function. If the relationship between $f(\mathbf{x})$ and the control variables is of a particular form, such as *linear*, e.g.

$$f(\mathbf{x}) = \mathbf{b}^T\mathbf{x} + c, \tag{2.1}$$

where $\mathbf{b}$ is a constant-valued vector and c is a constant, or *quadratic*, e.g.

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} + \mathbf{b}^T\mathbf{x} + c, \tag{2.2}$$

where $\boldsymbol{A}$ is a constant-valued matrix, special methods exist that are guaranteed to locate the optimal solution very efficiently. These, along with other, classifications are summarized in Table 2.1.

Table 2.1: Optimization Problem Classifications.

Characteristic	Property	Classification
Number of control variables	One	Univariate
	More than one	Multivariate
Type of control variables	Continuous real numbers	Continuous
	Integers	Integer or Discrete
	Both continuous real numbers and integers	Mixed Integer
Problem functions	Linear functions of the control variables	Linear
	Quadratic functions of the control variables	Quadratic
	Other nonlinear functions of the control variables	Nonlinear
Problem formulation	Subject to constraints	Constrained
	Not subject to constraints	Unconstrained

3. Problem Formulation

The optimal solution to a problem is the one with the *best* combination of the parameters which fulfils the functions set out in the problem specifications. In engineering optimization is often performed in the context of design, where *best* frequently means the *cheapest* solution that does the *job*. The overall aim is thus to minimize the cost (or some other performance attribute) while maximizing the functionality.

The overall aim is to formulate the optimization problem in precise mathematical terms and then to solve for the optimal solution. The five main steps are:

1. Create a basic configuration
2. Identify the decision variables
3. Establish the objective function
4. Identify any constraints
5. Select and apply an optimization method

3.1 Create A Basic Configuration

To create a Basic Configuration it is usually necessary to produce a simplifying model of the problem to highlight the *Decision Variables*.

3.2 Identify The Decision Variables

These should be independent parameters. They can be *continuous* (e.g. linear dimensions, volume), *discrete* (e.g. bolt sizes: M10, M12, etc.), or *integer* (e.g. number of cylinders in a piston engine).

Discrete and integer variables can often be replaced by continuous variables and the necessary adjustments made at the end of the optimization process (see Johnson for further details). We shall therefore concentrate mainly on continuous variables.

3.3 Establish The Objective Function

The most common objective function is some measure of cost and the aim is to minimize this. Other objective functions, e.g. weight, may be used, but these can usually be related to cost.

The objective function must be established in terms of the decision variables, i.e. given values for the decision variables it must be possible to calculate the objective function.

3.4 Identify Any Constraints

Not all decision (control) variables can assume values over an infinite range. Constraints are therefore applied. They can be *equality* constraints (e.g. fixed volume) or *inequality* constraints (e.g. minimum radius, maximum height).

Equality constraints can often be incorporated directly into the objective function (by eliminating a variable).

Inequality constraints are more difficult to handle, but they can sometimes be left out initially and then considered when interpreting the results from candidate optima. Other approaches are discussed later (e.g. penalty functions).

3.5 Select And Apply An Optimization Method

There are two basic classes of optimization methods:

3.5.1 Optimality Criteria

Analytical methods. Once the conditions for an optimal solution are established, then either:

- a candidate solution is tested to see if it meets the conditions, or
- the equations derived from the optimality criteria are solved analytically to determine the optimal solution.

3.5.2 Search Methods

Numerical methods. An initial trial solution is selected, either using common sense or at random, and the objective function is evaluated. A move is made to a new point (2nd trial solution) and the objective function is evaluated again. If it is smaller than the value for the first trial solution, it is retained and another move is made. The process is repeated until the minimum is found.

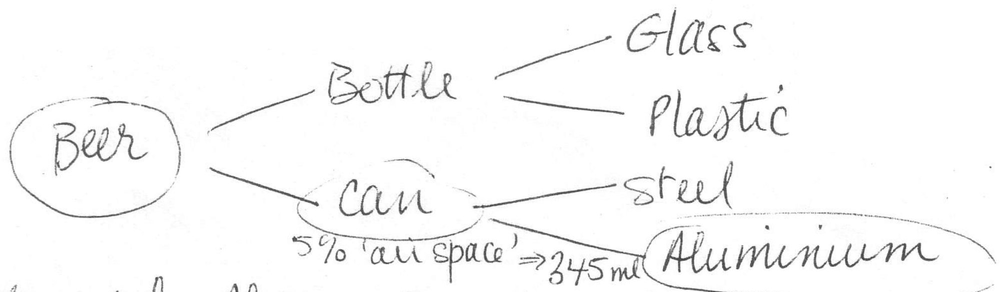
Search methods are used when:

- the number of variables and constraints is large;
- the problem functions (objective and constraint) are highly nonlinear; or
- the problem functions (objective and constraint) are implicit in terms of the decision/control variables — making the evaluation of derivative information difficult.

The most appropriate method will depend on the type (classification) of problem to be solved.

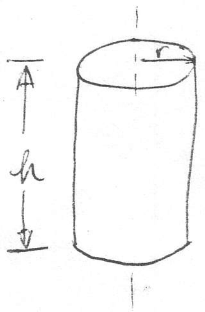
3.6 Example: Beer Container

Task: Design a beer container to hold 330 ml of beer.



1st model: Al can, simple cylinder;
cost of manufacture $\propto$ surface area of can

Step 2: Decision Variables



2 independent variables: r & h

so $\underline{x} = (x_1, x_2) = (r, h)$

Step 3: Objective Function

cost $\propto$ area (A)

$$A = 2\pi r^2 + 2\pi r h = 2\pi (r^2 + rh)$$

$$\therefore \text{minimize } f(\underline{x}) = f(r, h) = 2\pi (r^2 + rh)$$

Step 4: Constraints

Equality constraints: $V = \pi r^2 h = 345$ cc

Inequality constraints: $r_{min} \le r \le r_{max}$

say $25 \text{ mm} \le r \le 50 \text{ mm}$

$0 \le h$ (can be excluded)

Step 5: Optimization method

Can be treated as unconstrained problems by eliminating equality constraint & not worrying now about the inequality constraint.

substituting for h, from $V = \pi r^2 h = 345$ get

$$\boxed{A = f(r) = 2\pi \left(r^2 + \frac{V}{\pi r}\right)}$$ ← function of one varia

Use optimality criteria. Will return to this problem.

4. Optimality Conditions

4.1 Definitions And Background Theory

The *goal* of the optimization process is to:

$$\text{minimize } f(\mathbf{x}) \quad \text{subject to } \mathbf{x} \in S \,, \tag{4.1}$$

where S is the set of feasible values (the *feasible region*) for $\mathbf{x}$ defined by the constraint equations (if any), i.e. the values of $\mathbf{x}$ for which all constraints are satisfied. Obviously, for an unconstrained problem S is infinitely large.

The *gradient* vector of $f(\mathbf{x})$

$$\mathbf{g}(\mathbf{x}) = \nabla f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right]^T \tag{4.2}$$

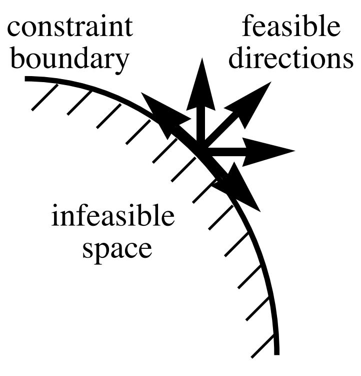
denotes the direction in which the function will increase most per unit distance travelled.

The *Hessian* of $f(\mathbf{x})$ is an $n \times n$ symmetric matrix giving the spatial variation of the gradient:

$$\boldsymbol{H}(\mathbf{x}) = \nabla(\nabla f(\mathbf{x})) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{4.3}$$

$\mathbf{d}$ is a *feasible direction* at $\mathbf{x}$ if an arbitrarily small move from $\mathbf{x}$ in the direction $\mathbf{d}$ remains in the feasible region, i.e. if there exists an α_0 such that

$$\mathbf{x} + \alpha\mathbf{d} \in S \qquad \forall\ 0 < \alpha < \alpha_0 \,. \tag{4.4}$$



The definition of the *global minimum* $\mathbf{x}^*$ of $f(\mathbf{x})$ is that

$$f(\mathbf{x}^*) \le f(\mathbf{y}) \qquad \forall\ \mathbf{y} \in S, \mathbf{y} \ne \mathbf{x}^* \,. \tag{4.5}$$

If we can replace $\le$ with $<$ then this is a *strict* or *strong global minimum.*

A *relative* or *weak local minimum* exists at $\mathbf{x}^*$ if an arbitrarily small move from $\mathbf{x}^*$ in any feasible direction results in $f(\mathbf{x})$ either staying constant or increasing, i.e.

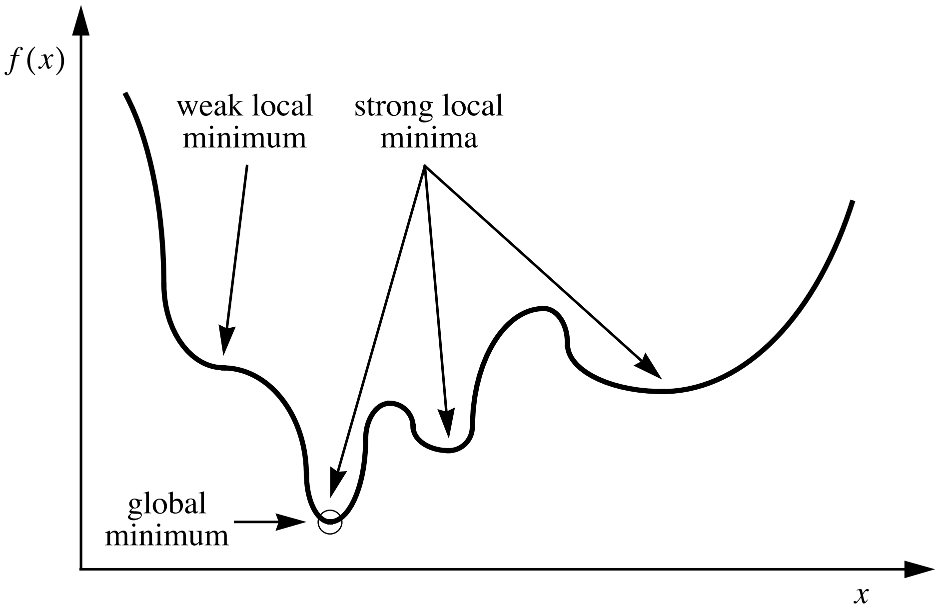
$$f(\mathbf{x}^*) \leq f(\mathbf{y}) \qquad \forall \mathbf{y} = \mathbf{x}^* + \varepsilon \mathbf{d} \in S, \mathbf{y} \neq \mathbf{x}^* . \tag{4.6}$$

If $f(\mathbf{x})$ is a smooth function with continuous first and second derivatives for all feasible $\mathbf{x}$, then a point $\mathbf{x}^*$ is a *stationary point* of $f(\mathbf{x})$ if

$$\mathbf{g}(\mathbf{x}^*) = \nabla f(\mathbf{x}^*) = 0 . \tag{4.7}$$

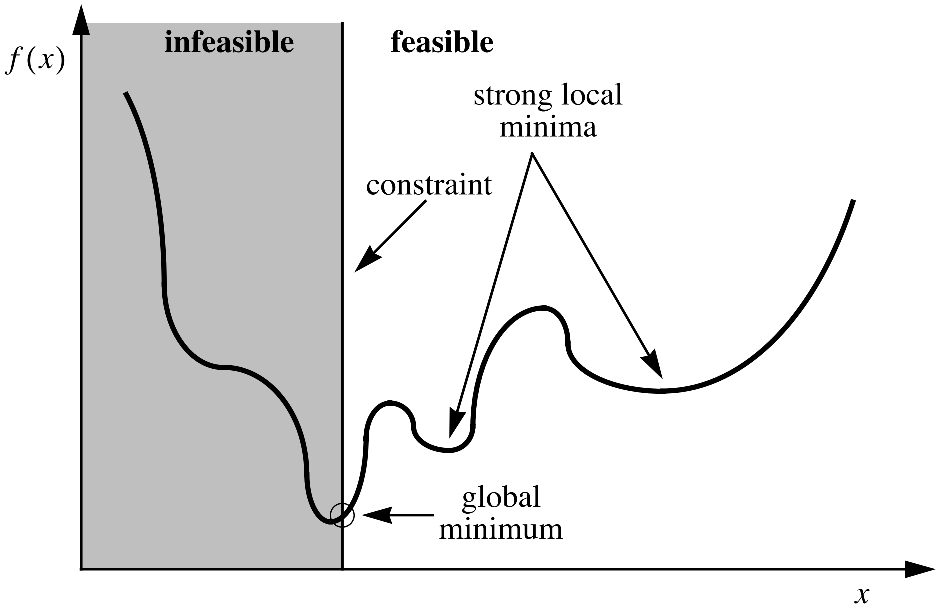
Figure 4.1 illustrates the different types of stationary points for unconstrained univariate functions.

Figure 4.1: Types of Minima for Unconstrained Optimization Problems.



As shown in Figure 4.2, the situation is slightly more complex for constrained optimization problems. The presence of a constraint boundary, in Figure 4.2 in the form of a simple bound on the permitted values of the control variable, can cause the global minimum to be an extreme value, an *extremum* (i.e. an endpoint), rather than a true stationary point.

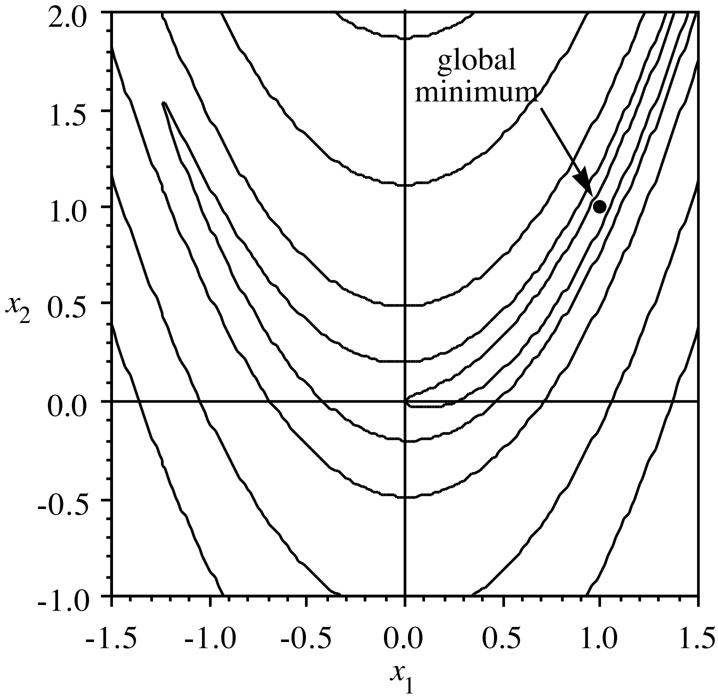
Figure 4.2: Types of Minima for Constrained Optimization Problems.



A function is *unimodal* if it has a single minimum value, with a path from every other feasible point to the minimum for which the gradient is < 0. An example is shown in Figure 4.3.

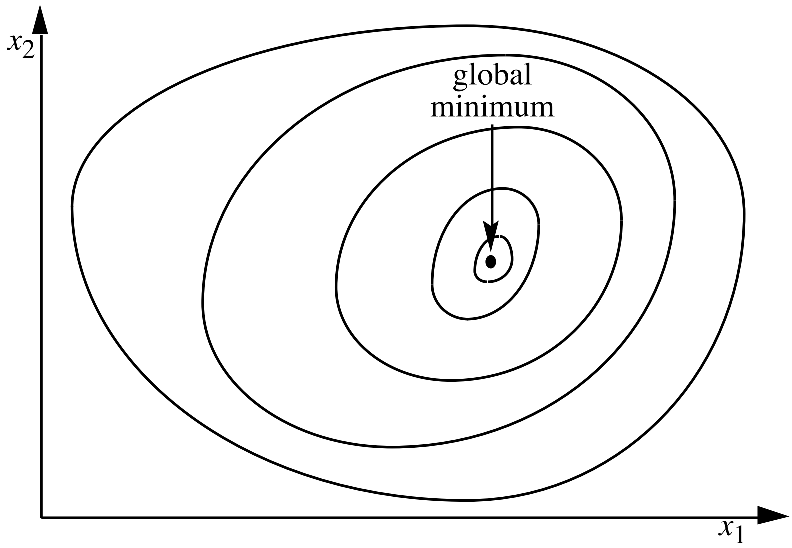
Figure 4.3: An Example of a Unimodal Function (Rosenbrock's Function

$$f(\mathbf{x}) \;=\; 100\,(x_2 - x_1^2)^2 + (1 - x_1)^2).$$



A function is *strongly unimodal* if a straight line in control variable space from every point to the minimum has a gradient < 0. An example is shown in Figure 4.4.

Figure 4.4: An Example of a Strongly Unimodal Function.



4.2 Conditions For A Minimum

A necessary condition for $\mathbf{x}^*$ to be a *local minimum* of $f(\mathbf{x})$ over S is that

$$\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0 \tag{4.8}$$

for all possible feasible directions $\mathbf{d}$ at $\mathbf{x}^*$, since any move from the minimum must cause the function to increase or stay the same.

$\nabla f(\mathbf{x}^*) \cdot \mathbf{d} = 0$ if $\mathbf{x}^*$ is an *interior point*, a point not on the boundary of the feasible region. Clearly, this condition also applies to a maximum and to a saddle point, since the gradient at both these points is zero. Thus this condition is *necessary* but not *sufficient* for a local minimum.

A second-order condition is needed.

4.2.1 Sufficient Condition for a Minimum

Any continuous function can be approximated in the neighbourhood of any point by a *Taylor series*. This series can thus be used to establish necessary and sufficient criteria for a minimum.

For the *single variable* case:

$$f(x) = f(x^*) + (x - x^*) f'(x^*) + \frac{1}{2}(x - x^*)^2 f''(x^*) + R , \tag{4.9}$$

where R is the remainder of the expansion, and is small compared with the first terms of the series, $f' = \frac{\mathrm{d}f}{\mathrm{d}x}$ and $f'' = \frac{\mathrm{d}^2 f}{\mathrm{d}x^2}$.

Let $x = x^* + d$ and $f(x) = f(x^*) + \Delta f$.

If x^* is a minimum and an interior point, then $f'(x^*) = 0$ and $\Delta f \geq 0$.

Thus, from the Taylor series (neglecting *R*):

$$\Delta f = \frac{1}{2} d^2 f''(x^*) .$$

Hence

$$f''(x^*) \geq 0 \tag{4.10}$$

is a *sufficient second-order* condition for a minimum. (If > then it is a *strict* or *strong minimum*.)

For a function of *several variables*, the Taylor series has the form

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \nabla f(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^T \boldsymbol{H}(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) + R . \tag{4.11}$$

Let $\mathbf{x}^*$ be an interior minimum, $\mathbf{x} = \mathbf{x}^* + \mathbf{d}$ and $f(\mathbf{x}) = f(\mathbf{x}^*) + \Delta f$. Then, $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ and $\nabla f(\mathbf{x}^*) = 0$. Thus, from the Taylor series (neglecting *R*):

$$f(\mathbf{x}^*) + \Delta f = f(\mathbf{x}^*) + \frac{1}{2}\mathbf{d}^T \boldsymbol{H}(\mathbf{x}^*)\mathbf{d}$$

$$\therefore \quad \mathbf{d}^T \boldsymbol{H}(\mathbf{x}^*)\mathbf{d} \geq 0$$

Thus, the conditions for a *local minimum* are:

$$\nabla f(\mathbf{x}^*) \cdot \mathbf{d} \geq 0 \tag{4.8}$$

$$\mathbf{d}^T \boldsymbol{H}(\mathbf{x}^*)\mathbf{d} \geq 0 \tag{4.12}$$

Equation (4.12) implies that the function is *convex* at $\mathbf{x}^*$, and will be satisfied if the Hessian $\boldsymbol{H}(\mathbf{x}^*)$ is a *positive definite matrix.*

4.2.2 Convex Functions

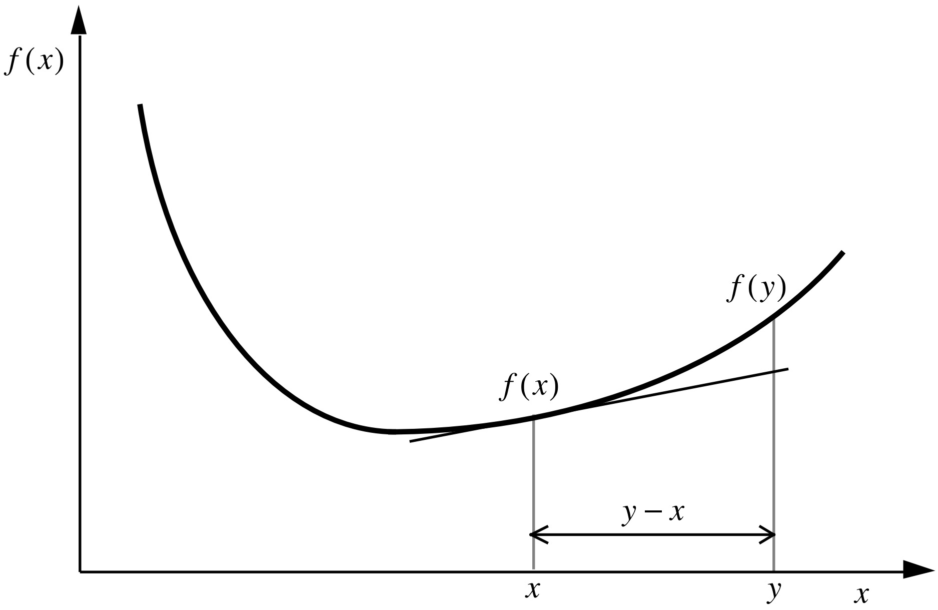
A function is said to be convex at $\mathbf{x}^*$ if for all arbitrarily small $\delta\mathbf{x} = \mathbf{x} - \mathbf{x}^*$ the value of the function is greater than or equal to the first two terms of the Taylor expansion. If the function is convex everywhere, i.e. if for all $\mathbf{x}$ and $\mathbf{y}$

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}) , \tag{4.13}$$

then *f* is called a *convex function* and $\mathbf{x}^*$ is a *global minimum.*

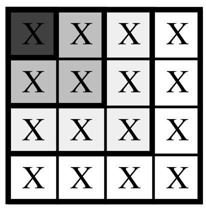
An example of a convex function is shown in Figure 4.5.

Figure 4.5: An Example of a Convex Function.



4.2.3 Positive Definite Hessian

The Hessian $\boldsymbol{H}(\mathbf{x}^*)$ is a *positive definite matrix*, if the determinant of each principal minor matrix of $\boldsymbol{H}$ is positive. For example, a 4×4 Hessian matrix will be positive definite if the determinant of each of these matrices is positive:



If $\boldsymbol{H}(\mathbf{x}^*)$ is positive definite then all its eigenvalues are positive.

If $\boldsymbol{H}(\mathbf{x}^*) = 0$ then higher-order terms need to be evaluated to determine whether $\mathbf{x}^*$ is a local minimum or not.

The Hessian of a quadratic function is the same everywhere. So, for a quadratic, a strict local minimum at an interior point is a global minimum.

4.3 Optimality Conditions Example: Beer Container

Objective $f(r) = r^2 + \frac{V}{\pi r}$

$$25 \text{ mm} \leq r \leq 50 \text{ mm}$$

Initial approach – ignore constraint

1st-order condition $\frac{df}{dr} = 0$

$$\frac{df}{dr} = 2r - \frac{V}{\pi r^2}$$

$$\Rightarrow r^* = \left(\frac{V}{2\pi}\right)^{1/3}$$

If $V = 345$ mL $\Rightarrow r^* = 38$ mm

Compare with constraints – constraints not violated $\therefore$ this is a genuine stationary point

Need to check 2nd-order condition

$$\frac{d^2f}{dr^2} = 2 + \frac{2V}{\pi r^3}$$

This is clearly > 0 for any +ve r

$\therefore r^* = 38$ mm is a minimum

$$h^* = \frac{V}{\pi r^2} = 76 \text{ mm}$$

Real cans have $r = 33$ mm