CHEMICAL ENGINEERING 4G03

Module 01 – Problem Formulation

Dr. Jake Nease

Chemical Engineering

McMaster University

Updated January 13, 2023

Preface

This course module is meant to introduce optimization formulation. It uses several numerical and real-world examples to convey the subject, but bear in mind when reading this chapter that all formulations, depending on the features of the problem, have their own features (such as types of variables, types of constraints, *etc.*). Each type of problem encountered in 4G will therefore have its own examples on formulation in addition to the tools presented in the following document.

Some Potentially Useful Nomenclature

In all of these modules, if you encounter a **definition**, it will usually be in a red shaded box labelled as such. For example, the definition for an inequality constraint $g(\boldsymbol{x}) \leq \boldsymbol{0}$ will take the form:

Inequality Constraints

Inequality constraints are defined as any constraints conforming to:

$$g(\boldsymbol{x}) \leq \boldsymbol{0}$$

- **One-Way** limits on the system and are *essential* for optimization.
- Inequality constraints of the form ≥ 0 may be re-written to appear as ≤ 0.
- There can be many of these constraints $\Rightarrow g(\boldsymbol{x})$ is a *vector* that can be **indexed**.
- Inequalities are critical to defining the bounds of a feasible region and preventing **unbounded solutions**.

Moreover, **class workshops and problems** will usually be presented in a gray shaded box labeled as below. For example, a workshop asking for the definition of a linear objective function will take the form:

Class Workshop

Consider a **linear** optimization problem with variables x_i each with associated costs c_i for the set $i = \{1, 2, \dots, n\}$. Determine the *objective function* for this problem if it is to represent total cost:

1. As a summation of indexed variables
2. As a product of vectors

Workshops will typically be followed by a space, plot, or area for the solution to be taken up in class:

Workshop Solution

Outline of Module

This module consists of the following topics:

- Defining decision variables
  - Independent versus dependent
  - Discrete versus continuous
- Defining variable bounds
- Defining types of constraints
  - Inequality versus equality
  - Independence of inequality direction
- Defining an objective function
- Feasible regions
  - Defining boundaries
  - Plotting and visualizing feasible regions
  - Feasible search directions
  - Improving search directions
  - Examples of graphical solutions
- Scenarios of optimal outcomes

Suggested Readings

Rardin (1st edition): Chapters 1, 2

Rardin (2nd edition): Chapters 1, 2

The Importance of a Good Formulation

In model-based optimization, conclusions are drawn from the *model* of the system, NOT from the problem itself. As a result, it is critical to **develop an appropriate model for your system**. An inadequate model will lead to false conclusions, even if the solution of the model itself is "correct" according to the model equations.

All models are wrong. Some models are useful. – George Box

Moreover, it is important that the model is **computationally tractable** for it to be practical. Enormous models with very high levels of detail and thousands of complicated nonlinear equations may seem nice from a modeling accuracy perspective, but it will prove very difficult to solve computationally. A **trade-off** exists between model accuracy and computational efficiency (*aka* tractability).

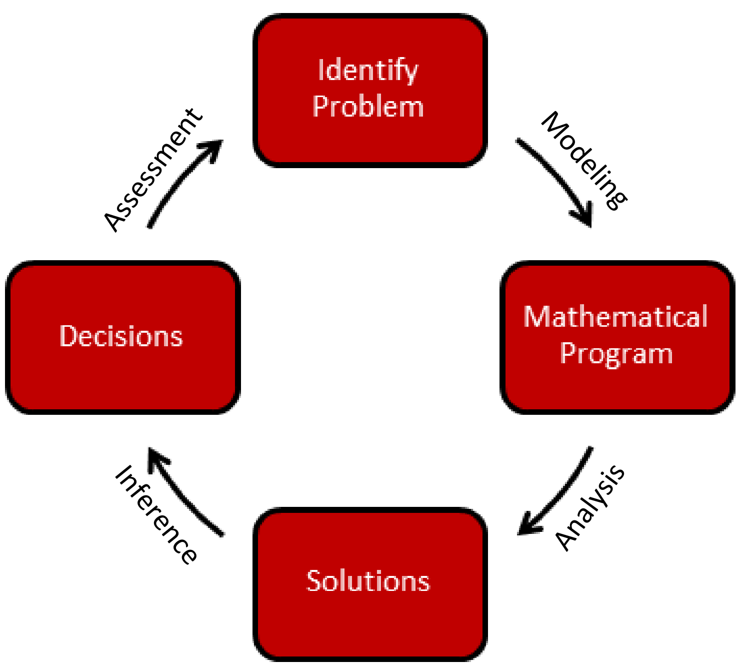
Model Validity versus Tractability

- The **VALIDITY** of a model is the degree to which inferences drawn from it hold real meaning for the system we are modeling.
- The **TRACTABILITY** of a model is the degree to which the model permits convenient analysis. This includes the ability for us (the users) to analyze results as well as the ability of the computer to generate those results.

There is always a **trade-off** between validity and tractability. We may make a "better" model, but this usually comes at the cost of additional complexity.

The model-based optimization procedure is typically *iterative* as shown in the figure below. Often, **defining the problem** is just as important as actually solving it; by defining an optimization problem, you are identifying the areas in which your system can improve.

We find it useful to formulate an optimization problem even if we cannot solve it. – Dofasco



Forming Optimization Models

Optimization models (mathematical programs) represent problem choices as *decision variables* that attempt to seek the maximum (or minimum) of an *objective function*. An objective function must be dependent on one or more decision variables. The ranges over which decision variables are permitted are typically subjected to *constraints*.

Generalized Optimization Program

$\min_{\boldsymbol{x}} \phi = f(\boldsymbol{x})$	← Objective Function
s.t.	← "Subject to"
$h(\boldsymbol{x}) = 0$	← Equality Constraints
$g(\boldsymbol{x}) \leq 0$	← Inequality Constraints
$\boldsymbol{x}_{lb} \leq \boldsymbol{x} \leq \boldsymbol{x}_{ub}$	← Variable Bounds

Decision Variables

Decision variables $\boldsymbol{x}$ may be grouped into one of two classes: **independent** and **dependent**. This delineation does not mean much in the long run but can be useful at the formulation stage to minimize the complexity of the problem. During the solution stage, dependent variables are usually constrained according to the model constraints and independent variables.

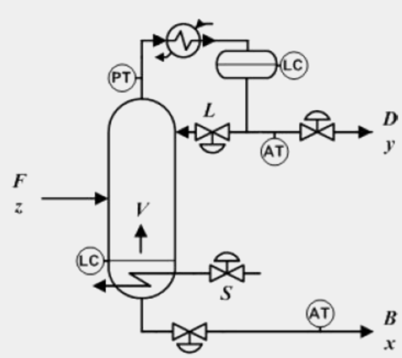
- **Independent Variables** can be manipulated to change the behaviour of the system.
- **Dependent Variables** are often expressed as more "convenient" combinations of independent variables for the sake of visualization, but in practice do not have an impact on the solution.

Examples of Decision Variables

- **Engineering Design:** reactor volume; distillation trays; HX area; reaction pathway
- **Operations:** temperature; flow rate; pressure; valve position; heat duty
- **Management:** feed composition; purchase prices; sales prices; supplier locations
- **Others:** maximum investment permitted; risk tolerance; environmental impact

Class Workshop – Decision Variables

List some common variables associated with a distillation column. Identify those variables as independent or dependent with a brief justification.



Workshop Solution – Decision Variables

Discrete Versus Continuous Variables

A decision variable is **discrete** if it is limited to a fixed or countable set of values (all-or-nothing, either-or, must be produced/purchased in integer quantities, *etc*.). Mathematically speaking, it means a set of discrete decision variables $\boldsymbol{x}_D$ (sometimes called $\boldsymbol{y}$) takes the form:

$$\boldsymbol{x}_D \ \epsilon \ \boldsymbol{I} = \{0, 1, 2, \dots\}$$

A variable is **continuous** if it may take on *any value* in a specified interval. A set of continuous variables $\boldsymbol{x}_C$ is therefore defined to be:

$$\boldsymbol{x}_C \ \epsilon \ \boldsymbol{S} \subset \mathbb{R}$$

Where that weird "⊂" means "is a subset of." Problems that contain a mixture of integer/discrete and continuous variables are known as **Mixed Integer Programs (MIPs)**, which we will explore a little later.

A **heuristic** about selecting appropriate variables is that *modeling with continuous variables is preferred over modeling with discrete variables*. This is for tractability reasons (think of numerical methods – It is far easier to solve a continuous function than a discontinuous function).

Another **heuristic** regarding discrete versus continuous variables is that if a discrete variable scale is large enough so that rounding to the nearest integer is possible with *minimal loss of model accuracy*, that variable can be modeled *continuously*.

Class Workshop – Continuous versus Discrete Variables

Decide whether a continuous of discrete variable should be assigned to best model the following quantities:

1. The optimal temperature of a chemical reactor
2. The warehouse slot assigned to a particular product
3. Whether or not we select a product for capital investment
4. The number of golf balls produced in a plant that manufactures 10,000 + per day
5. The number of aircraft produced on a defense contract
6. That start-up cost of a process unit
7. The optimal number of hours spent studying for the 4G03 final exam

Workshop Solution – Continuous versus Discrete Variables

Indexing Variables

In real applications, optimization problems can become quite large. They can quickly grow to thousands of equations, millions of variables, and more! We clearly do not want to write them all out (no, really), so we typically exploit **indexed notational schemes** to keep large models manageable during formulation.

Indexing is typically used to condense multiple dimensions of decisions into a single variable. For example, we can decide how much of a reactant (of limited supply) is used to create a certain amount of product (of required demand) with a single variable!

Indexing

Indexes (or **subscripts**) permit the representation of collections of similar pieces of information with a single symbol. For example, we might define a variable x_i which represents a decision about each value of $\boldsymbol{x}$ for, say, 100 values of index i:

$$\{x_i : i = 1 \dots 100\}$$

The first step of defining an optimization model is to choose appropriate indexes for each dimension of the problem. **Multiple indexes** in the same problem (for the same variable) are very common.

Class Workshop – Indexing

A large food manufacturer operates 20 different manufacturing plants. Each plant has the capability of producing 30 different types of pre-packaged foods and can supply 25 different customers across a given market. We may denote the dimensions of the problem as:

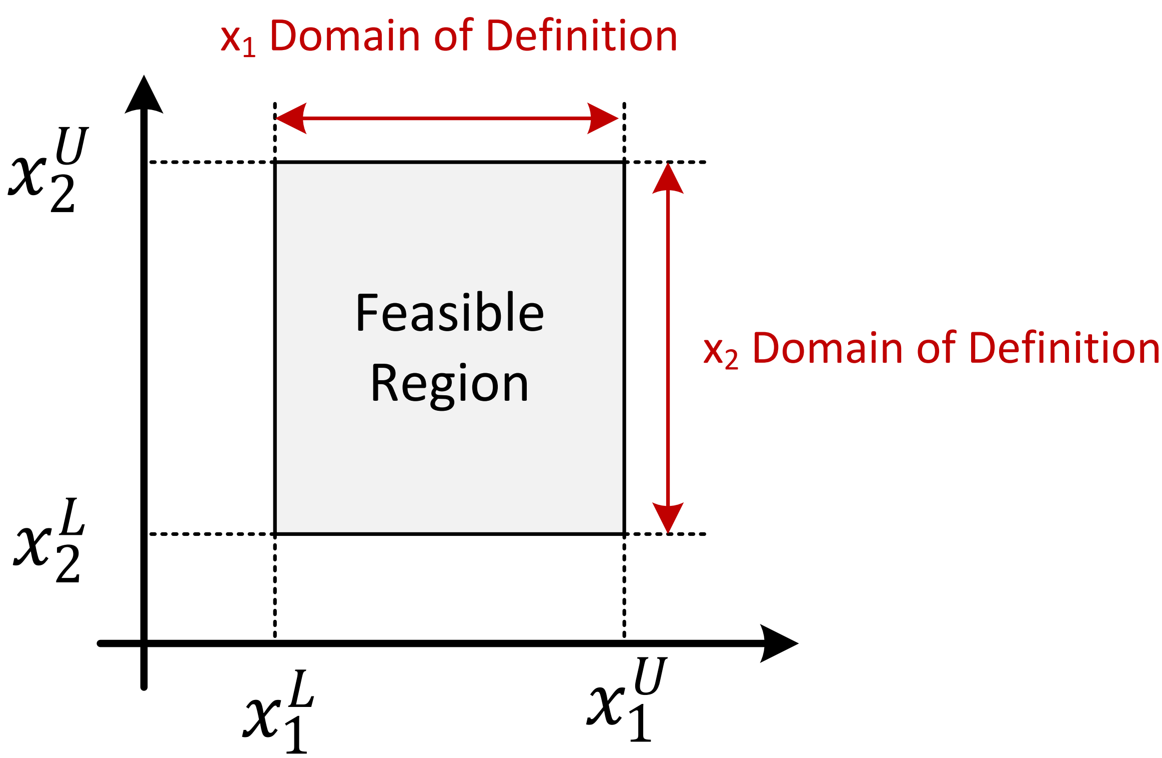
- p is the plant facility number: $p = \{1, \cdots, 20\}$
- f is the food product produced: $f = \{1, \cdots, 30\}$
- r is the sales region that can be supplied by a given plant: $r = \{1, \cdots, 25\}$

1. Based on this scenario, what might be the variables for this problem (in words)?
2. Using indexing, define the appropriate decision variables for this problem (mathematically)
3. What is the total number of decision variables for this model?

Workshop Solution – Indexing

Variable Bounding

Decision variables are usually *bounded* in optimization problems to reflect physical reality (cannot purchase negative raw materials, cannot have a temperature less than 0 K) and model assumptions (cannot sleep less than 6 hours per day). The upper ($\boldsymbol{x}^U$) and lower bounds ($\boldsymbol{x}^L$) for given variables define their so-called **domain of definition**. The combination of domains of definition provides what is known as the **feasible region** or **feasible set** (more on this later).



Remarks

- Assigning the minimum and maximum values to the same number sets that variable to a specific value → $x_i^L = x_i^U = x_i$ for some i.
- The most common variable bound is **non-negativity** → $x_i \geq 0 \ \forall \ i$

Notation comment – the upside-down "A" (∀) is the "for all" symbol.

Class Workshop – Variable Bounds

Propose appropriate bounds for the variables in the distillation column workshop example.

Workshop Solution – Variable Bounds

Constraints

Constraints are self-imposed or physically realizable restrictions and interactions between decision variables that further limit the allowable values of those variables. There are two types of constraints generally used in optimization: **inequality constraints** and **equality constraints**.

Inequality Constraints

Inequality constraints are defined as any constraints conforming to:

$$g(\boldsymbol{x}) \leq \mathbf{0}$$

- **One-Way** limits on the system and are *essential* for optimization.
- Inequality constraints of the form ≥ 0 may be re-written to appear as ≤ 0.
- There can be many of these constraints $\Rightarrow g(\boldsymbol{x})$ is a *vector* that can be **indexed**.
- Inequalities are critical to defining the bounds of a feasible region and preventing **unbounded solutions**. (ALL THE SPORTS in our final year experience example)

There are several reasons that we might want to limit the values a given variable can take, such as:

- Safety
- Product qualities (contracts, performance)
- Equipment damage (long term, short term [failure])
- Operating windows
- Legal/ethical concerns
- Prevents regions of mathematical/physical inconsistencies (*e.g.*, steady state that is not achievable)

Examples of inequality constraints

- Maximum allowable investment
- Maximum/minimum fluid flow through a pump/compressor
- Minimum vapour flow rates on distillation trays (prevents weeping)
- Maximum pressure in a sealed vessel
- Maximum region in which a simplification/approximation is valid (*e.g.*, linear controllers)

Equality Constraints

Equality constraints are defined as any constraints conforming to:

$$h(\boldsymbol{x}) = \mathbf{0}$$

- These describe interactions or physical relationships between variables in the model.
- Written with a **zero RHS** according to convention.
- There can be multiple equality constraints so that $h(\boldsymbol{x})$is a vector that can be **indexed**.
- There **CANNOT** be more independent equality constraints than decision variables in the model. Think back to your numerical methods class and convince yourself as to why this is true.

Examples of equality constraints

- Material, energy, and momentum balances.

$$Accumulation = In - Out + Generation$$

- Constitutive relationships.

$$Q = UA\Delta T_{\ell m}$$

$$k = k_0 e^{-\frac{E}{RT}}$$

$$A + B \rightarrow 2C + D \Rightarrow X_A = -2X_C$$

- Equilibrium conditions such as VLE or $\frac{d}{dt} = 0$.
- Decisions made by the engineer during the design phase.

$$F_A = 2F_B$$

- Behaviour enforced by controllers (set-points). Note that this can lead to a dependent variable.

$$T_S = 273.15\ K$$

The Objective Function

The objective function is our method to evaluate the impacts of decision variables.

Objective Function

The objective function ϕ is a function of the decision variables $\boldsymbol{x}$ for which we want to find the *minimum* or *maximum*. For example:

$$\min_{\boldsymbol{x}} \phi = f(\boldsymbol{x})$$

- We require the objective function to be **quantitative** (numerical), rather than qualitative.
- A scalar objective function is preferred, although many multi-objective problems exist. In these cases, it is sometimes possible to combine objectives into a single scalar quantity.
- There is no fundamental (mathematical) or practical difference between *maximization* and *minimization* problems (why?). That is: $\min_{\boldsymbol{x}} \phi_1 = f(\boldsymbol{x}) \Leftrightarrow \max_{\boldsymbol{x}} \phi_2 = -f(\boldsymbol{x})$

Examples of scalar quantities representing performance

- Maximize profit (or minimize cost)
- Maximize product quality
- Minimize energy use
- Minimize environmental footprint
- Minimize time required to complete a task
- Maximize anticipated safety (how do you get a single value for this?)

Class Workshop – Objective Function

Consider a **linear** optimization problem with variables x_i each with associated costs c_i for the set $i = \{1, 2, \ldots, n\}$. Determine the *objective function* for this problem:

1. As a summation of indexed variables
2. As a product of vectors

Workshop Solution – Objective Function

Model Formulation Lecture Case Study 1 – Crude Distillation

A refinery under your supervision distills crude petroleum into **three** products: gasoline, jet fuel, and lubricants. Your plant receives crude oil shipments from **two** locations: Saudi Arabia and Venezuela. The chemical compositions of the crude from each location are slightly different, and thus yield different products per unit of crude oil refined. You may assume that the qualities from each location are constant.

- Each barrel from Saudi Arabia yields 0.3 barrels of gasoline, 0.4 barrels of jet fuel, and 0.2 barrels of lubricants.
- Each barrel from Venezuela yields 0.4 barrels of gasoline, 0.15 barrels of jet fuel, and 0.35 barrels of lubricants.
- The remaining 0.1 (10%) from both sources is lost to the refining process.

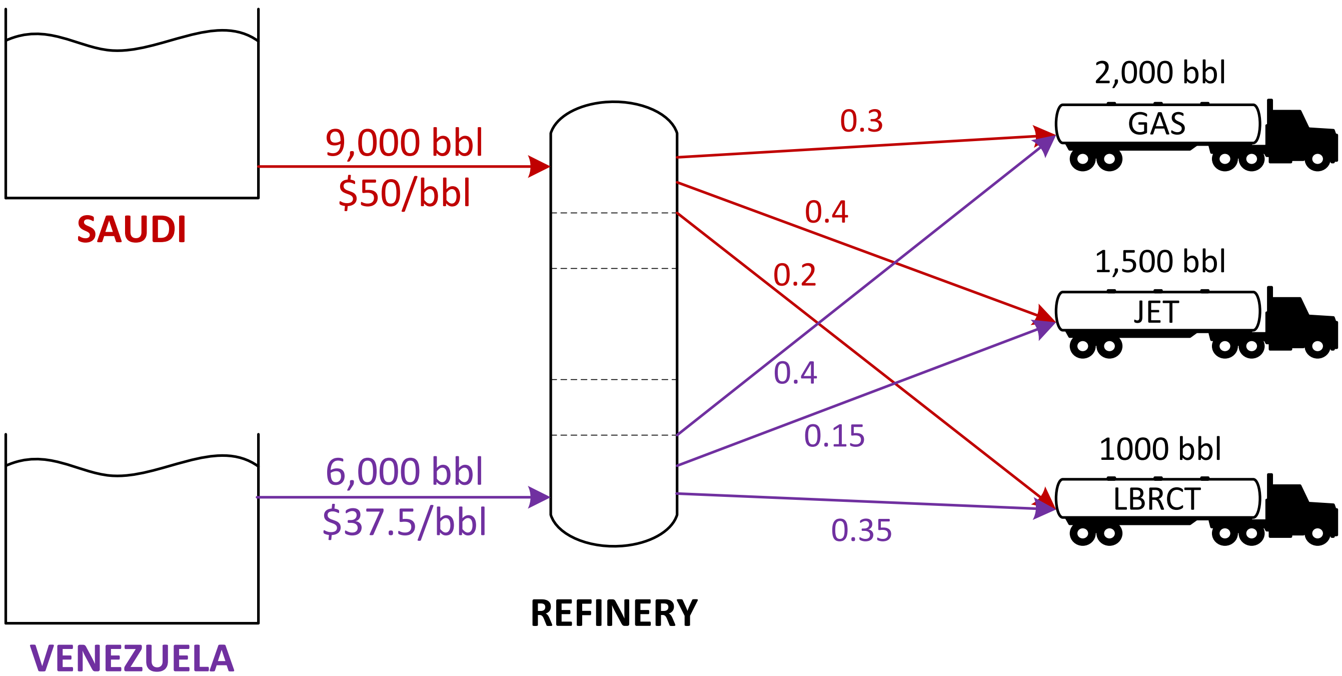
The crudes also differ in availability and cost:

- The Saudi oil costs your refinery \$50 per barrel and is available up to 9,000 barrels per day.
- The Venezuelan crude costs your refinery \$37.5 but only available up to 6,000 barrels per day.

You have a contact with local distributors to provide **2,000 barrels of gasoline**, **1,500 barrels of jet fuel**, and **1000 barrels of lubricants** per day.

Your task is to **draw a diagram of the supply network** and **formulate** the optimization problem for this scenario.

Lecture Case Study 1 Solution



The formulation of this problem can be written as follows (note that other formulations are possible). In the following formulation, x_1 represents the number of barrels purchased from Saudi Arabia and x_2 is the number of barrels purchased from Venezuela.

$$\min_{x_1,x_2} \phi = 50x_1 + 37.5x_2$$

Subject to

$$\begin{aligned}
0.3x_1 + 0.4x_2 &\geq 2,000 \\
0.4x_1 + 0.15x_2 &\geq 1,500 \\
0.2x_1 + 0.35x_2 &\geq 1,000 \\
x_1 &\leq 9,000 \\
x_2 &\leq 6,000 \\
x_i &\geq 0 \qquad (\forall\, i)
\end{aligned}$$

This formulation can then be translated into `GAMS`, `MATLAB`, or Microsoft Excel for solving. This solution can also be done *graphically*, but we will get to that later. Do you see what the answer is before even optimizing? It might not be that obvious.

The solution to this problem has been coded up in GAMS with the file name: `Module_01_CaseStudy1`.

Graphing Model Constraints

In many small models, such as 2- or 3-dimensional models, it is useful to plot constraints and objective contours to arrive at a graphical solution. When graphing a potential optimization problem, we are faced with the requirement of forming the **feasible set**:

Feasible Set

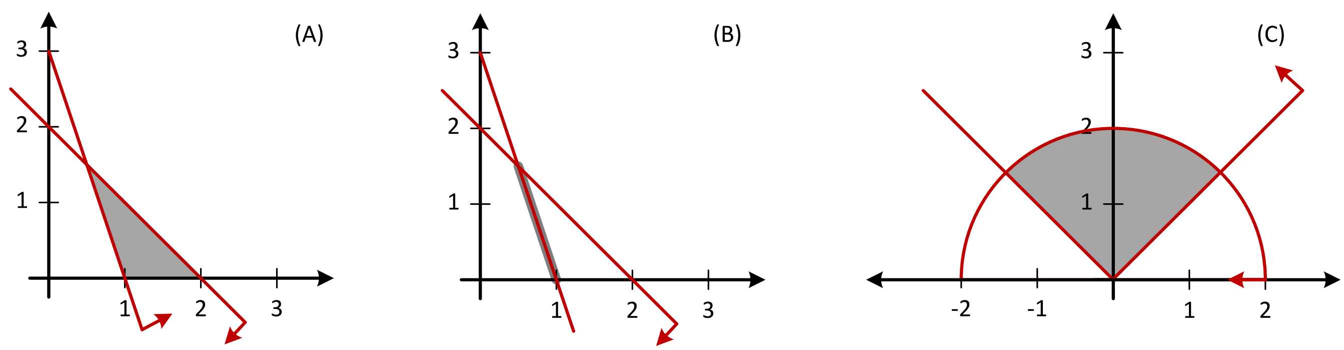
The feasible set (or region) $\mathcal{S}$ of an optimization model is the collection of decision variables that satisfy *all* the model constraints:

$$\mathcal{S} \triangleq \{\boldsymbol{x} : g(\boldsymbol{x}) \leq 0, h(\boldsymbol{x}) = 0, x^L \leq x \leq x^U\}$$

- The set of all points satisfying $h(x) = 0$ results in a **line** or **vector** when plotting the feasible set.
- The set of all points satisfying $g(x) \leq 0$ results in a **region** bounded above ($\leq$) or below ($\geq$) by a line defining the inequality.

For example, consider the following three sets of constraints, which result in the feasible regions shown in the figure below:

Constraint set (A)	*Constraint set (B)*	*Constraint set (C)*
$x_1 + x_2 \leq 2$	$x_1 + x_2 \leq 2$	$x_1^2 + x_2^2 \leq 4$
$3x_1 + x_2 \geq 3$	$3x_1 + x_2 = 3$	$\lvert x_1 \rvert - x_2 \leq 0$
$x_1 , x_2 \geq 0$	$x_1 , x_2 \geq 0$	



Class Workshop – Feasible Region

On the grid provided on page 13, plot the feasible set for the two-crude case study from Lecture Case Study 1. The work area for this proceeds the objective function and optimum discussions.

Graphing Objective Functions

Graphing objective functions in two dimensions requires the plotting of the objective function as **contours**. Knowing the objective function and how it behaves in the feasible region is critical to solve an optimization problem graphically.

Objective Contour

The contour C_ϕ of an objective function ϕ (in the decision variable space) is the line or curve passing through values of the decision variables $\boldsymbol{x}$ having a **constant value for the objective function** ϕ:

$$C_\phi \triangleq \{\boldsymbol{x} : f(\boldsymbol{x}) = \phi\}$$

The **easiest way to plot objective contours** is to assign a value for ϕ, treat it as a constant, and plot the resulting profile on the (x_1, x_2) axes. For example, consider the following two small, constrained optimization problems, resulting in the feasible regions and objective contours shown below:

Problem (A)

$$\min_{x_1, x_2} \phi = 3x_1 + x_2$$

Subject to

$$x_1 + x_2 \leq 2$$

$$x_1, x_2 \geq 0$$

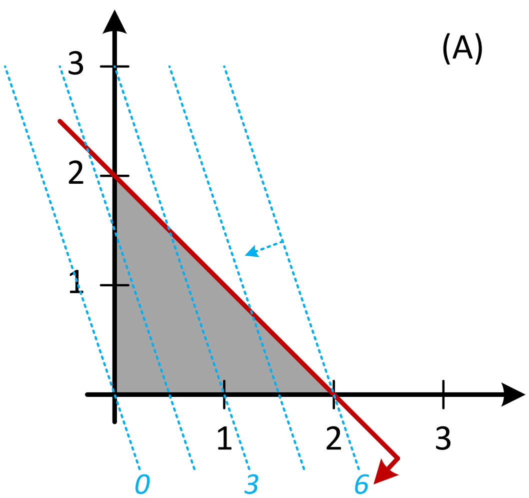
Problem (B)

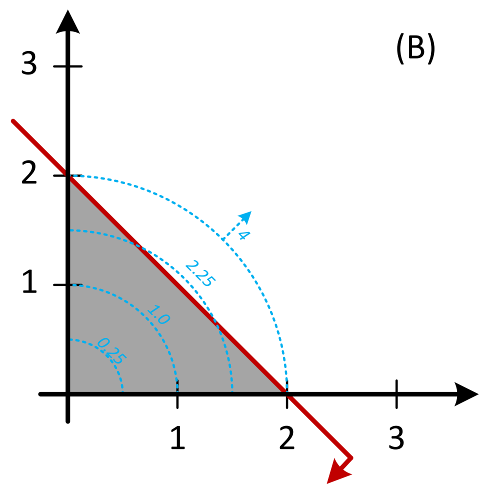
$$\max_{x_1, x_2} \phi = x_1^2 + x_2^2$$

Subject to

$$x_1 + x_2 \leq 2$$

$$x_1, x_2 \geq 0$$





Class Workshop – Objective Contours

On the graph paper provided on page 15, plot the objective contours for the two-crude case study from Lecture Case Study 1. The work area for this proceeds the optimal outcome discussions.

Optimization Outcomes

The Optimum

An optimum solution $\boldsymbol{x}^\star$ to an optimization problem is a *feasible* choice of the decision variables with an objective function value ϕ *at least* as low (high) as any other solution satisfying the constraints:

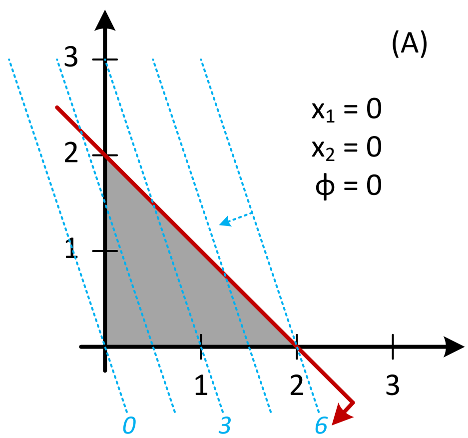
$$\phi(\boldsymbol{x}^\star) \leq (\geq)\, \phi(\boldsymbol{x}) \quad \forall\, \boldsymbol{x} \in \mathcal{S}$$

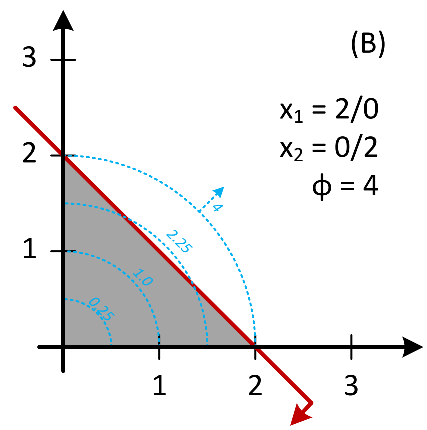
We will get into more detailed definitions of optimality later.

As far as nomenclature and graphical solutions are concerned, please consider the following **remarks**:

- Optimal solutions are shown graphically to be the point(s) lying on the best objective function contour that intersects with at least one boundary of the feasible region.
- The *optimal value* $\phi^\star$ is defined to be the value of the objective at the optimum(s): $\phi^\star \triangleq \phi(\boldsymbol{x}^\star)$.
- An optimization model can have only **one** true optimal value. It may have:
  - A **unique** optimal solution.
  - Several **alternative** solutions $\boldsymbol{x}^\star$ yielding the *same* optimal $\phi^\star$.
  - **No** optimal solutions (either the problem is unbounded or infeasible).

For example, considering the same optimization problems as above, we can identify the *optimal solutions* for each of them according to the plots below:





However, we run into situations of **degeneracy** and **unboundedness** (more on these in a later section) when we slightly modify the problem as follows:

Problem (A)

$$\max_{x_1,x_2} \phi = 3x_1 + 3x_2$$

Subject to

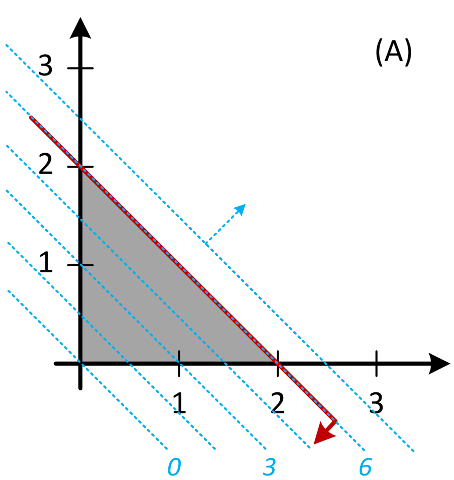
$$x_1 + x_2 \leq 2;\; x_1, x_2 \geq 0$$

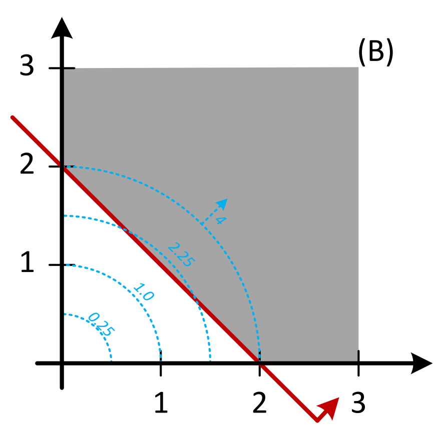
Problem (B)

$$\max_{x_1,x_2} \phi = x_1^2 + x_2^2$$

Subject to

$$x_1 + x_2 \geq 2;\; x_1, x_2 \geq 0$$

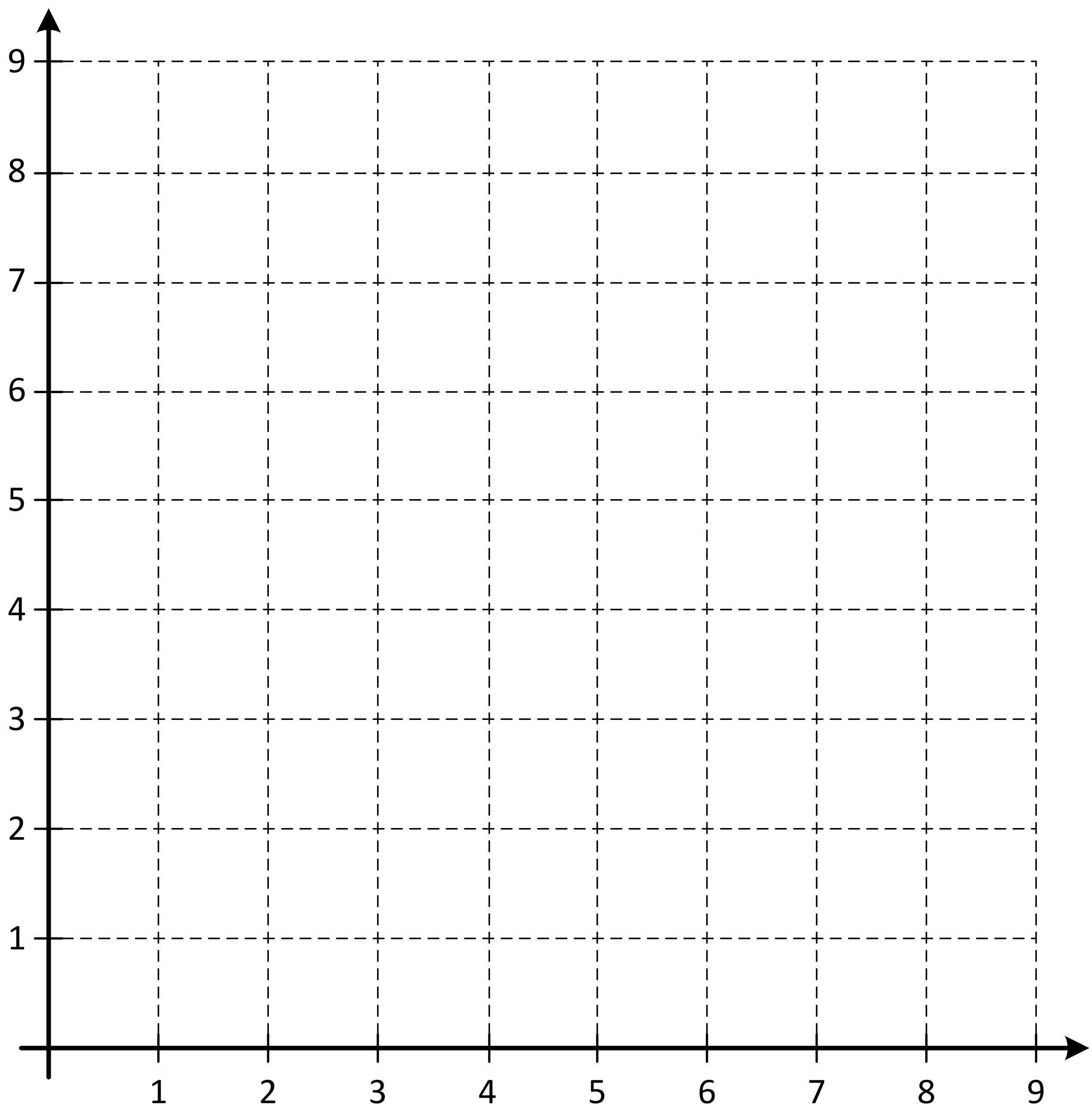




Class Workshop – Locating Optimum

On the graph paper provided below, locate the optimum for the two-crude case study from Lecture Case Study 1.

Workshop Solution – Graphical Optimization of 2-Crude Distillation Study



The solution to this problem has been coded up in GAMS with the file name:

`Module_01_CaseStudy1`

Active Constraints
A constraint $g(\boldsymbol{x}) \leq 0$ is said to be:

- **Active** (or **binding**) at some point $\boldsymbol{x}^\star$ if $g(\boldsymbol{x}^\star) = 0$.
- **Inactive** at some point $\boldsymbol{x}^\star$ if $g(\boldsymbol{x}^\star) < 0$.

And finally, consider the following **remarks** regarding active constraints:

- The set of constraints that are active at the optimal solution are known as the **active set**.
- Equality constraints are **ALWAYS** active at any feasible optimal point.
- No constraints (inequality or equality) may be violated at any optimal point.

Conclusions

We now have a good handle on the formulation of optimization problems. We will be revisiting this topic *continuously* throughout the course as we define different types of problems and apply different methods to solving each. Formulation and graphical solutions are **imperative** to enhance one's understanding of the types of problems we may encounter, and how to turn real-world opportunities into formal mathematical programs that we can handle.

An EXTRA model formulation problem is presented below for your practice/interest. The solution is available and discussed in detail.

\|O‿O|/

Model Formulation Case Study 2 – Distribution Network

Via Rail Canada must reallocate its rail cars to ship a large harvest of corn and grain across Canada. There are five (5) regions that require a minimum number of rail cars. Moreover, each region currently possesses several rail cars that can be added to (or taken from) as necessary. The following table breaks down the availability of cars and number of cars required for each region, as well as the *cost* of moving a car from one region to another.

	Region				
From	1	2	3	4	5
1	—	10	12	17	35
2	10	—	18	8	46
3	12	18	—	9	27
4	17	8	9	—	20
5	35	46	27	20	—
Present	115	385	410	480	610
Need	200	500	800	200	300

Your task is to **formulate a problem** that will move cars between the regions to meet their requirements in the most cost-efficient way possible.

Lecture Case Study 2 Solution

In this case, it is obvious that the variables we are dealing with are *discrete*. However, since their scale is large (in the hundreds), chances are that we can treat them as *continuous* variables and round them to the nearest integer once we find a solution. As a matter of fact, we can formulate it the same way and solve it either as continuous or discrete in GAMS to see how much of a difference it will make. Let's define some variables:

- Let $x_{i,j}$ be the number of rail cars shipped *from* region i *to* region j.
- Let $C_{i,j}$ be a matrix of cost coefficients for shipping one rail car from i to j. Note that this matrix will take on the same shape and values of the 5×5 matrix in the table above.
- Let P_j be the number of cars presently at each region before transit.
- Let N_j be the number of rail cars required in each region.

We can now define the optimization problem as follows (note each of the constraints). We can do a couple of different things here, such as change the inequality for the first constraint to be an equality. Notice we do not constrain one location to avoid shipping *from* if it is also being shipped *to*. Why is this? If it *were* not allowed, would the optimization problem choose it to be so?

Note also the $i \neq j$ condition in the summation for the first constraint. That can be a bit of a logic nightmare. Can you think of any ways to enforce such a constraint more easily? *Hint*: We do not have any cost assigned to shipping from one location to itself (for good reason). If we *DID* allow this, would our first constraint make any sense?

Would you classify this problem as linear (LP), nonlinear (NLP), integer (IP), or mixed-integer (MILP/MINLP)? Provide a brief justification.

$$\min_{x} \phi = \sum_{i} \sum_{j} x_{i,j} C_{i,j}$$

Subject to

$$\sum_{\substack{i \\ i \neq 1}} x_{i,1} - \sum_{\substack{j \\ j \neq 1}} x_{1,j} + 115 \geq 200$$

$$\sum_{\substack{i \\ i \neq 2}} x_{i,2} - \sum_{\substack{j \\ j \neq 2}} x_{2,j} + 385 \geq 500$$

$$\sum_{\substack{i \\ i \neq 3}} x_{i,3} - \sum_{\substack{j \\ j \neq 3}} x_{3,j} + 410 \geq 800$$

$$\sum_{\substack{i \\ i \neq 4}} x_{i,4} - \sum_{\substack{j \\ j \neq 4}} x_{4,j} + 480 \geq 200$$

$$\sum_{\substack{i \\ i \neq 5}} x_{i,5} - \sum_{\substack{j \\ j \neq 5}} x_{5,j} + 610 \geq 300$$

$$x_{i,j} \geq 0 \qquad (\forall\, i, j)$$

The solution to this problem has been coded up in GAMS with the file name: `Module_01_CaseStudy2`.

Variable Aliasing

It is sometimes the case, as in the problem above, that the only difference between two indices of **the same dimensionality** (i and j, for example) is that they may have different index positions (for example, "coming from" versus "going to"). If this is the case, we can **alias** the indices i and j to each other, meaning that although they are in fact unique indices, we can treat them as equivalent variables in our mathematical representation of the problem. If this is the case, it may be possible to write our first constraint above as:

$$\sum_{\substack{i \\ i \neq 1}} \left(x_{i,1} - x_{1,i} \right) + 115 \geq 200$$

We note here that we only have *one* summation (indexed to i) with the actual location of the subscript i reversed between the two variables. This is allowed if we alias i and j. The **benefit** of this idea extends beyond simply having one summation: Now that j has been freed up, we can safely include it in the constraint instead of a hard-coded 1, 2, *etc*. In fact, we can write the *entire set of constraints in the problem above as*:

$$\sum_{\substack{i \\ i \neq j}} \left(x_{i,j} - x_{j,i} \right) + P_j \geq N_j \qquad (\forall\, j)$$