Linear Programming as a tool for Refinery planning

Geoffrey Gill
Commercial Division
NZ Refining Company
Private Bag 9024
Whangarei, New Zealand

Abstract

Determining the best way to operate a refinery is a difficult task. In many Refineries world-wide a linear programming (LP) model is used to assist Refinery planners and schedulers. This paper will discuss some of the advantages and shortfalls a Refinery LP model has in terms of theoretical accuracy and practical use. Modelling issues that will be addressed include, optimal selection of crudes for processing, and the pooling problem. The unusual contractual application of the LP model between New Zealand Refining Company and the New Zealand oil companies will be described.

1 Introduction

A generalised Refinery can be thought of as a complex network of process units which process raw crude and produce ‘on-grade’ products. A process unit will be defined as something that has at least one input stream which is then processed in some manner to produce one or more output streams of desired quality. At the ‘front-end’ of this network is the Crude Distillation Unit (CDU). Crude(s) will go into a CDU and be cut into various fractions according to boiling point. The cut-points (or temperature at which the splits are made,) are chosen in order to ensure the most efficient use of the crude being processed. The boiling point of any fraction of a crude will usually be proportional to its weight. Having been processed through the CDU, the lightest fraction will then be an input for another particular unit where it is further processed, while simultaneously the other components will either be input for different specified units or be directly blended into a final product. The CDU will be the unit focussed on here.

Crude is not a homogeneous substance like pure water, but is made up of many different components all mixed together. Mogas (petrol, the fuel that many cars run on) is produced from some of the lighter fractions of crude, while bitumen which we can see on most roads, has some of the heaviest fractions of crude. Other products that originate from fractions between mogas and bitumen are jet fuel, diesel and fuel oil. No two crudes are identical in composition. The ratio of light components to heavy components will vary between two different crudes. Therefore, if a lighter crude was processed in a refinery (consisting of only one CDU), it would tend to produce more ‘mogas-like’ products, while a heavier crude would tend to yield proportionally less light products.
The CDU separates crude into (typically five) streams whose components differ in weight. Details concerning how the separation is done will be omitted here but can be found in Leffler. The cut-point at which a separation is made is not some arbitrary fixed constant, but can be varied. This means that one particular crude can be made to yield slightly more mogas and correspondingly less heavier product if desired. Alternatively, the same crude could be processed to make less mogas and more heavier products by lowering the temperature of the mogas cut-point. The restrictions are that the total demand must be met for each product and each product must be 'on-grade,' or satisfy certain specifications (for example mogas volatility and fuel oil viscosity.)

Ideas used in the construction of an LP model for the New Zealand Refining Company (NZRC) will be developed here for the two CDUs at NZRC. Other units can be modelled in a similar manner. The accuracy of the LP model will be assessed and the general LP contractual arrangements with the four major New Zealand oil companies (BP, Caltex, Mobil, and Shell) will be outlined.

2 LP modelling of the CDU

The approximate compositions for one of the lightest feedstocks, Maui, which is found in New Zealand, is shown below in figure 1. Contrasting that is the approximate distillation curve for a heavy Middle Eastern crude, Arabian Heavy.

![Figure 1: Crude compositions for 100 KT of Maui and Arabian Heavy at atmospheric pressure](image)

When Maui is processed in a CDU, most of the feed will be converted into mogas, jet-fuel and gas-oil with virtually no residue which would otherwise eventually become fuel-oil. The heavy fractions (residues) of crudes are sometimes 'cracked' into lighter components, but cracking will not be considered here. On the other hand, when Arabian Heavy is processed there is a relatively high percentage yield of residue available for bitumen.

Suppose that at atmospheric pressure, the percentage of Maui that evaporates is given in the following table.

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>0-140</th>
<th>140-220</th>
<th>220-400</th>
<th>400-550</th>
<th>&gt;550</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stream</td>
<td>Mogas</td>
<td>Jet fuel</td>
<td>Gas-Oil</td>
<td>Fuel-Oil</td>
<td>Bitumen</td>
</tr>
<tr>
<td>Percent</td>
<td>35</td>
<td>30</td>
<td>34</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
If 100 Kilo Tonnes (KT) of Maui was processed and the first cut-point was set at 140 °C, then 35 KT would be converted into mogas. Also, if the second cut-point was set at 220 °C, then 30 KT of jet fuel would result, and so on for the other products in this example. This will be called the ‘140-Maui’ mode of operation of the CDU. Now let the first cut-point be increased to 180°C. The difference in this mode of operation is that the fraction of Maui that evaporates between 140°C and 180°C (say 10%) is now part of the mogas stream and not the jet fuel stream any more. Therefore a typical yield of 100 KT of Maui in this ‘180-Maui’ mode would be 45 KT of mogas, only 20 KT of jet fuel, and the ratio of the other products would remain unchanged. Density is proportional to boiling point, which means that the average density of both the mogas and jet fuel streams in the ‘180-Maui’ mode would be higher than the respective densities in the ‘140-Maui’ mode. This concept also applies to many other properties in two adjacent streams, which provides refinery planners with additional opportunities.

A mode of operation can be used as a column in an LP constraint matrix, and extra elements can be added to the column representing various properties of the product streams such as the average density of the mogas cut. Its cost in the objective function is the sum of the cost of the crude and the processing cost for using that mode of operation. The cost would be measured in dollars per tonne. The value of an LP variable representing a mode of operation in an optimal LP solution is interpreted physically as the amount of crude that should be processed in that mode. When an LP has two (or more) columns representing the same crude in two different modes of operation in the optimal solution, then the optimal cut-point is a weighted ratio. For example, if the variable representing the ‘140-Maui’ mode was 75 and the ‘180-Maui’ variable was 25, then 100 KT of Maui should be processed with the first cut-point set at 150 °C. Similar columns can be generated for the Arabian Heavy crude, and any other crudes that may be available. Rows can be constructed to compute various properties of final products while ensuring that the properties satisfy given specifications. The assumption in this model is that properties blend linearly. This means if one fuel oil stream with a sulphur content of 4% by weight was mixed in equal ratio with another fuel oil stream with 2% sulphur, then the final sulphur content of the fuel oil blend would be 3%.

3 Accuracy of the LP model

3.1 Processing more than one crude

Crude is heated up and injected into a CDU where the components are separated according to planned boiling point. The components with the highest boiling point rise to the top of the CDU. If a light crude such as Maui was processed on its own then the bottom of the CDU, where the high boiling point components arrive at, would be barely utilised. Likewise, if Arabian Heavy was processed on its own, the top of the CDU would be virtually empty. These inefficiencies would suggest that it may be a better idea to mix the crudes in such a way that allows increased utilisation of the whole CDU. In practice it is desirable to maximise total input into a CDU, because in this way the rest of the processing units in the refinery are equally utilised. Also, since in real life the number of different crudes usually exceeds the number of crude tanks, a tank may contain a mixture of two or more crudes. A CDU with two crudes being pooled before being processed, can be seen in figure 2.
However, the model we have seen so far assumes that both Maui and Arabian Heavy could be co-processed and then the favourable qualities from the Maui only, could be drawn into a particular stream without the qualities from Arabian Heavy 'contaminating' that stream. For example the low sulphur content from the Maui could be drawn into the gas-oil stream, without the high sulphur content of Arabian Heavy also entering the gas-oil stream. The result would be a gas-oil stream with a supposedly low sulphur content, according to this segregated LP model. This would be acceptable if the crudes were processed in the CDU separately, but there are various reasons mentioned earlier why pooling of crudes is desirable or necessary.

3.2 LP Recursion to overcome pooling

Suppose two input streams I₁ and I₂, with qualities Q₁ and Q₂ respectively, are blended together. The actual amounts of the input streams (I₁ and I₂) are unknown. If the quality of the blend has to satisfy a product quality specification, then the problem is a blending problem which is linear and can be solved using standard LP techniques such as the Revised Simplex Method. However if the blend is not a final product, but has two or more output streams which are sent to other destinations, then the problem is non-linear¹. The case with two inputs, two outputs, one type of quality (indicated in brackets) and one pool has been illustrated in figure 3.

Figure 3: Illustration of the pooling problem

\[
I₁ (Q₁) \rightarrow \text{Pool} \rightarrow O₁ ((I₁Q₁ + I₂Q₂) / (I₁ + I₂)) \quad I₂ (Q₂) \rightarrow \text{Pool} \rightarrow O₂ ((I₁Q₁ + I₂Q₂) / (I₁ + I₂))
\]

In oil refining, there is not only the pooling of crudes, but streams from the CDU from one day to the next that can be put to tankage (and consequently pooled) before being further processed in another unit. Product tanks are often mixed together to obtain final products that meet all the required specifications. Most attempts at modelling pooling involve estimating either the qualities of the pool, or the quantities or qualities of the outputs and optimising the resulting LP, then adjusting the estimates based on the optimal solution and reoptimising the LP. This process of adjusting estimates and then reoptimising a new LP continues until (hopefully) the estimates coincide with the LP solution. Such techniques are called LP recursion, and some variations are discussed in further detail in Ciriani and Leachman (1993). There has been some success with the implementation of LP recursion to find optimal solutions to pooling problems consistently².
3.3 Other non-linearity features

Lead is added to the premium grade of petrol in order to boost the octane number which in turn provides more power when combustion takes place. When 0.1 g/L of lead is combined with an unleaded grade of petrol, the octane increases by about seven octane numbers. (The increase in octane can vary depending on the quality of components of the crude). However when a second 0.1 g/L of lead is added, the increase in octane number is approximately three. This 'diminishing returns' type behaviour of octane improvement with extra lead continues for any amount of lead added. The Healy method is a piecewise linear approximation to the lead response curve, and it uses reliable experimental information about the values of the derivatives of the lead response curve at two or three different lead levels. The method has been used satisfactorily in many refinery LP codes including that at NZRC. There are other octane boosting additives that may be used after September 1996, when lead will no longer be allowed, in order to increase the octane number. Some of these have an increasing effect on the octane number for each extra quantity added. This type of situation may lead to local optima when standard LP techniques are used.

Flash-point is defined as the temperature at which a product will give off enough vapours to self-ignite. It is important for obvious reasons that operating temperatures of an engine should always be below the flash-point of the fuel that is being used. A property like sulphur content blends linearly. However, when a component with a low flash-point is combined with an equal quantity of high flash-point material, the resulting mixture has a flash-point that lies closer to the low flash-point than the high flash-point. The pour-point is the lowest temperature at which a product will 'flow' under certain laboratory conditions. This property is very difficult to predict accurately, and tests do not appear to suggest any kind of linear blending relationship.

As well as non-linear blending behaviour and pooling, the constraints of shipping can add 'integer' complications. For example an (operationally poor but) optimal LP solution may be to process only one barrel of Arab Heavy crude. However it is far from economically viable to send a ship all the way from Saudi Arabia to New Zealand with only one barrel of crude on board! Likewise, processing units in a refinery have fixed costs associated with operating them that are independent of the input into the processing units themselves such as electricity cost. For this reason it is more economic to shut a unit down than operate it at a low throughput. An oil refinery would generally be classified as a continuous operation, but when final products are either prepared in a tank by blending, or taken away by ship or by pipeline, the activity would be more reminiscent of a batch type operation which is implicitly integer in nature. The LP model ignores this fact. For medium to long term planning, (one month to one year) this continuity of products assumption does not conflict with daily operational planning.

With the overall improvement in both computing power and optimisation algorithms, mixed integer programming for refineries is becoming a popular enhancement to LP models. NZRC and the four major New Zealand oil companies use an LP model as one of their tools to assist in the planning of the purchase of crudes, refinery operation and product production. However a standard LP can not prepare for 'the unexpected,' so quick decisions still must be made manually by experienced refinery planners.
4. The LP environment at NZRC

Access the facilities at NZRC are shared between the four major New Zealand oil companies, BP, Caltex, Mobil and Shell according to an agreed market share formula. The four user companies must share not only each process unit, but the purchase of crude and the product offtakes. NZRC develop an LP model which gives a good representation of the actual refinery as a whole. This information is made up of technical data pertaining to all the process units such as maximum (and minimum) capacities in tonnes of throughput per day, how the units may be operated, what crudes and components are allowed to feed each units, expected yields of crudes when they are processed through different units for various modes of operation, properties of component streams and final product blending restrictions. Each of the four user companies run the same LP model but must not optimise the units in such a way that exceeds their allocated share of capacity for all the process units. This means that although they are all sharing the same refinery, they effectively each have a different refinery configuration. The users use the LP model to assist them in crude purchases.

The contractual LP model runs on a three-monthly (or quarterly) basis. Plans must be made two months beforehand, because this is how long it can take for Middle-Eastern crudes to be sourced and shipped to New Zealand. Each quarter is not just determined by the collection of four LP outputs collected at one time, but is built up in a step-wise fashion. Approximately every two weeks an updated revision from each user company is sent to NZRC who have to assess the feasibility of the individual programmes both from a ‘stand-alone feasibility perspective’, and whether or not the programmes have adverse effects on the operation of the refinery as a whole. No user company is allowed to develop an LP solution that would infringe on the rights of another user. Once all revisions have been accepted, a ‘bulked-up’ LP is developed and published. In this way, each user knows exactly the others are doing.

In terms of optimisation, the users have the ability to purchase the best crudes, and produce desired quantities of products. In the (segregated crude) LP model that is used, the good qualities of a particular crude can be realised. NZRC have control over the best way to operate the refinery and in what order the crudes should be processed. At the end of each quarter, the LP is used retro-actively as an accounting tool to assist in determining the ownership of feedstocks processed and products produced. Any differences between actual refinery performance and that determined by the LP for each user are then settled by negotiation between companies. For example one user may settle for less gas oil in lieu of jet fuel.
5. Conclusion

The LP model has been used to assist in medium to long term planning successfully for fifteen years in its present format. It forms the basis for contractual arrangements regarding crude supply, crude purchasing, product manufacturing, product off-takes, and percentage utilisation share of each process unit. There are deficiencies in the LP model such as no pooling which allow 'over-optimisation' of the qualities of some crudes and component streams. A spreadsheet model is mainly used to perform the day-to-day scheduling and blending of products.

References


