Utrecht, June 2019

Business Intelligence

Lecture 05 - Prescriptve Analystics Optimisation and Simulation

Georg Krempl

Algorithmic Data Analysis
Information and Computing Sciences
Utrecht University, The Netherlands

With particular thanks to

- Armel Lefebvre (tutor, A.E.J.Lefebvre@uu.nl)

Prescriptive Analytics: Optimisation and Simulation

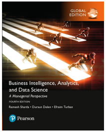


Figure: Textbook [Sharda et al., 2018, Chapter 3]
Sharda, Delen, Turban & King (2018). Business Intelligence, Analytics & Data Science: A Managerial Perspective 4th Global Edition, Pearson. ISBN-13: 9781292220567

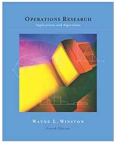


Figure: Textbook [Winston, 1997]
Winston (2003)). Operations Research: Applications and Algorithms 4th edition, Belmont ISBN: 0-534-42362-0

Outline and Summary[1]

- Prescriptive Analytics in BI: Optimisation
  See [Sharda et al., 2018, chapter 6]

- Note: The discussion of simulation will be skipped.

▸ Start ▸ Appendix

[1] Note: This lecture integrates knowledge from several further sources (cited where used, except if my own).

Outline and Summary[1]

- Prescriptive Analytics in BI: Optimisation
  See [Sharda et al., 2018, chapter 6]
- Note: The discussion of simulation will be skipped.

▸ Start ▸ Appendix

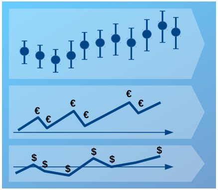
[1]Note: This lecture integrates knowledge from several further sources (cited where used, except if my own).

Prescriptive Analytics in BI

Advantages

- Makes it simpler to use for more people
  Play with multiple what-if-scenarios
- Combination of Decision Optimisation and Machine Learning

Optimisation - Motivating Questions

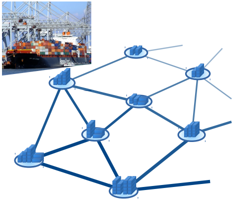


Portfolio Optimisation

You want to optimise your investment portfolio to maximise your profit, but at the same time to keep your risk low by ensuring that it is diversified between domestic and foreign index-based funds.

- How do you approach this problem?
- As potential data source, you can use the output of a predictive regression model that predicts the expected return of the index-based funds that you consider.

Optimisation - Motivating Questions

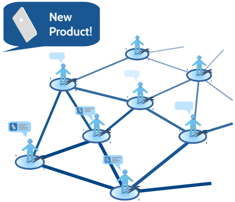


Optimising Logistics

As a junior consultant, a company asks you to develop a model that optimises distribution from its logistics centres to warehouses. Of course, the capacity of each logistics centre is limited, and delivery contracts have to be satisfied.

- How do you approach this problem?
- As potential data source, you can use data from a model that predicts the costs of delivery from each centre to each warehouse, as well as data about the required daily deliveries.

Optimisation - Motivating Questions



Optimising Marketing

Your company will soon release a new product. Which customers should the company invite as product testers, such that news and recommendations of the product reach the maximum number of other potential customers?

- How do you approach this problem?
- You have data how customers reacted on previous product offerings, and you can obtain from a third party social network data about your customers.

Optimisation

Optimisation

The aim is to find the *best* (optimal) solution to a *problem*.

In the Business Intelligence context, this means:

Not only describe the past/current or predict the future status, but suggest (re)actions to optimise the outcomes.

Answering the question: How and why make it happen?

We need to discuss what we mean with

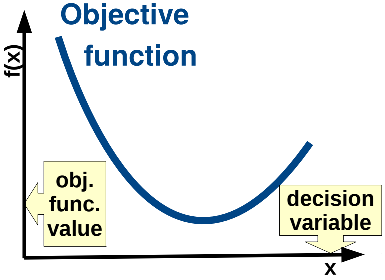
- **Problem**
- **Optimality**

Objective Function $f(x)$

The output that we minimise or maximise.

Decision Variables $x = (x_1, x_2, \cdots)$

The input to the objective function, which we control.



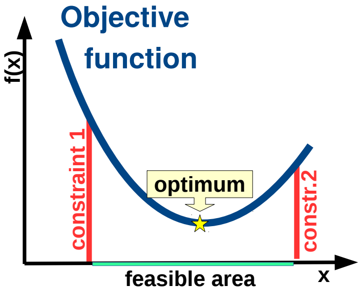
$$\min_{x_1, x_2, \cdots \in R^n} a_1 \cdot x_1 + a_2 \cdot x_2 \cdots$$

or

$$\max_{x_1, x_2, \cdots \in R^n} a_1 \cdot x_1 + a_2 \cdot x_2 \cdots$$

Constraints

- Restrict the decision variables, either
  - to be equal a certain value (equality constraint),
  - to be less (more, resp.) a certain value (inequality constraint)
- Constraints originate from
  - “Business rules”: e.g., constraints in working hours, diversity of a portfolio, ...
  - “Physical laws”: e.g., non-negativity of working hours, the new task’s starting time is later than the previous ones’ ending time...
- Define the **feasible area**



$$\min_{x_1, x_2, \dots \in R^n} a_1 \cdot x_1 + a_2 \cdot x_2 \cdots$$

subject to

$$\begin{aligned} -x_1 &\leq -0 \\ x_1 + x_2 &\leq +b_1 \\ \dots \end{aligned}$$

How to convert a maxisation into a minimisation problem?

$$\max_{x_1, x_2, \dots \in R^n} a_1 \cdot x_1 + a_2 \cdot x_2 + \cdots$$

Multiply with -1:

$$\min_{x_1, x_2, \dots \in R^n} -a_1 \cdot x_1 - a_2 \cdot x_2 - \cdots$$

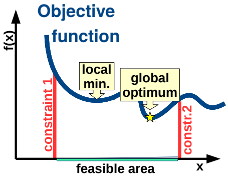
How to convert a $x \geq b$ constraint into a $\leq$-constraint?

$$x_1 + x_2 + \cdots \qquad \geq +b$$

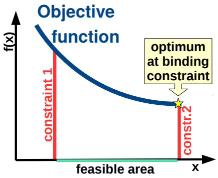
Again, multiply with -1:

$$-x_1 - x_2 \qquad - \cdots \leq -b$$

Optimisation - Mathematical Optimisation Problem Formulation



Local vs. Global Minimum (or Maximum)



Non-Binding vs. Binding (active) Constraint

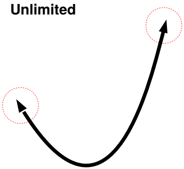
Categorisation of Optimisation Problems

What types of optimisation problems are there?

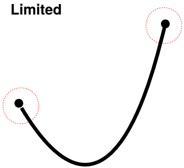
Categorisation of Optimisation Problems

Unlimited vs. Limited

- Limited: Optimisation problems bounded by constraints
- Unlimited: Optimisation problems without constraints



(a) Unlimited Problem

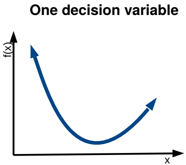


(b) Limited Problem

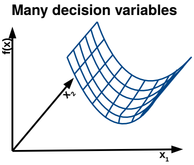
Categorisation of Optimisation Problems

One vs. Many Decision Variables

- Different number of decision (input) variables ...



(a) Single Decision Variable

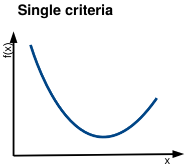


(b) Many Decision Variables

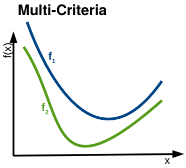
Categorisation of Optimisation Problems

Single vs. Multi-Criteria

- Single: A single objective function to consider
- Multi: Two or more objective functions to consider at once



(a) Single-Criterion



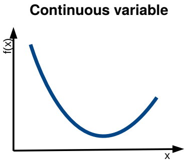
(b) Multi-Criteria

Categorisation of Optimisation Problems

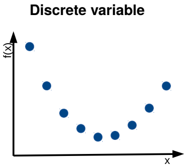
Continuous vs. Discrete

The involved variables can be

- continuous, e.g., $x \in \mathbb{R}$
- discrete, e.g., $x \in \mathbb{Z}$
- mixed, i.e., some continuous, some discrete



(a) Continuous



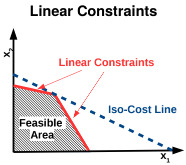
(b) Discrete

Categorisation of Optimisation Problems

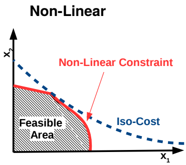
Linear vs. Non-Linear

Constraints (and objective function) can be

- linear
- non-linear



(a) Linear



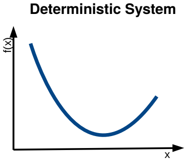
(b) Non-Linear

Categorisation of Optimisation Problems

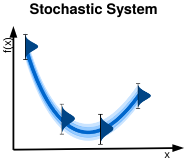
Deterministic vs. Stochastic System

The underlying system can be

- deterministic, i.e., the same cause produces always the same effect,
- stochastic, i.e., the effect depends on randomness



(a) Deterministic



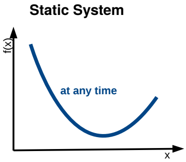
(b) Stochastic

Categorisation of Optimisation Problems

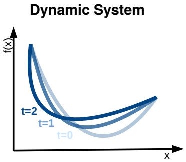
Static vs. Dynamic

The underlying system can be

- static
- dynamic (relationships change over time)

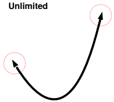


(a) Static

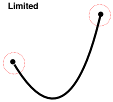


(b) Dynamic

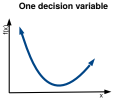
Categorisation of Optimisation Problems



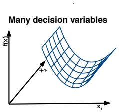
(a) Unlimited Problem



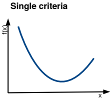
(b) Limited Problem



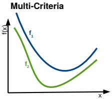
(c) Single Dec.Var.



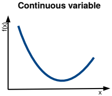
(d) Many Dec.Vars.



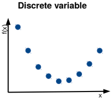
(e) Single-Criterion



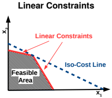
(f) Multi-Crititeria



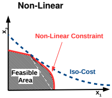
(g) Contin.



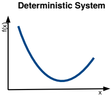
(h) Discrete



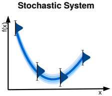
(i) Linear



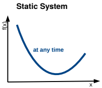
(j) Non-Linear



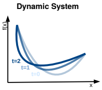
(k) Determ. System



(l) Stochastic System



(m) Static System



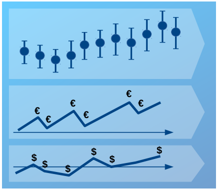
(n) Dynamic System

Optimisation - Revisiting Motivation's Examples

In the motivating examples,

- What are the objective function, the decision variables, and the constraints?
- What characteristics do these problems have?

Optimisation - Motivating Questions



Portfolio Optimisation

You want to optimise your investment portfolio to maximise your profit, but at the same time to keep your risk low by ensuring that it is diversified between domestic and foreign index-based funds.

- How do you approach this problem?
- As potential data source, you can use the output of a predictive regression model that predicts the expected return of the index-based funds that you consider.

Portfolio Optimisation[2]

- You want to optimise your portfolio of domestic and foreign index-based funds.
- Annual returns are 11% from domestic, and 17% from foreign stocks.
- Your available funds are 12.000
- At most 40, you want to invest 10.000 in domestic, and 7.000 in foreign stocks.
- At least half as much should be invested in domestic as in foreign,
  and at least half as much should be invested in foreign as in domestic.
- *What is the return-maximising investment plan?*

Formulation as Linear Program (LP)

- Decision variables are quantities invested in domestic x_1 and foreign x_2 funds
- Optimisation function is maximise $0.11 \cdot x_1 + 0.17 \cdot x_2$
- Non-negativity constraints: $x_1 \geq 0$, $x_2 \geq 0$
- Maximum fund capacity constraint: $x_1 + x_2 \leq 12$
- Diversity constraints: $x_1 \leq 10$, $x_2 \leq 7$
- Relative diversity constr.1: $x_1 \geq 0.5x_2 \Rightarrow -x_1 + 0.5x_2 \leq 0$
- Relative diversity constr.2: $x_2 \geq 0.5x_1 \Rightarrow +0.5x_1 - x_2 \leq 0$

[2] Based on [Winston, 1997, exercise 2-2].

Portfolio Optimisation

- You want to optimise your portfolio of domestic and foreign index-based funds.
- Annual returns are 11% from domestic, and 17% from foreign stocks.
- At most 40, you want to invest 10.000 in domestic, and 7.000 in foreign stocks.
- At least half as much should be invested in domestic as in foreign,
  and at least half as much should be invested in foreign as in domestic.
- *What is the return-maximising investment plan for your funds of* 12.000*?*

LP Formulation

$$\min_{x_1, x_2 \in R^2} -0.11 \cdot x_1 - 0.17 \cdot x_2$$

subject to

$$\begin{aligned}
-x_1 \quad\quad &\leq -0 \\
-x_2 &\leq -0 \\
x_1 + x_2 &\leq +12 \\
x_1 \quad\quad &\leq +10 \\
x_2 &\leq +7 \\
-x_1 + 0.5x_2 &\leq 0 \\
+0.5x_1 - x_2 &\leq 0
\end{aligned}$$

Portfolio Optimisation: R Code

Example

Solving the Portfolio Optimisation LP with lpSolve in R

```
library("lpSolve")
f.dir<-"min"
f.obj<-c(-0.11,-0.17)
f.con<-matrix(c(1,1,1,0,0,1,-1,+.5,+.5,-1),ncol=2,byrow=TRUE)
f.condir<-c("<=","<=","<=","<=","<=")
f.rhs<-c(12,10,7,0,0)
obj<-lp(f.dir,f.obj,f.con,f.condir,f.rhs)
print(obj) # not very informative
if (obj$status==0){
  print(paste("Success. Optimum at ",toString(obj$solution),
    " with objective function value ",obj$objval))
  print("Checking which constraints are active (binding):")
  print(cbind(f.con%*%obj$solution, f.condir, f.rhs))

}else{
  print("error: optimisation was not successful")
}
```

Note: lp assumes non-negativity of all variables

Portfolio Optimisation

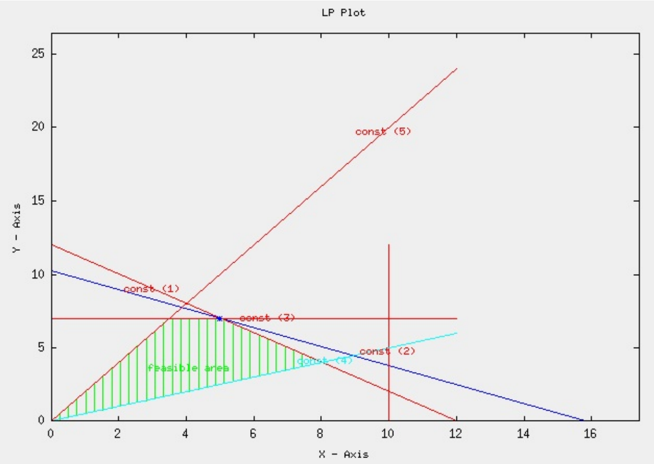
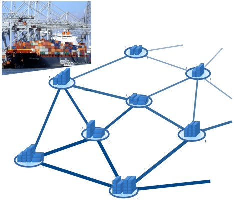


Figure: Plot of the Portfolio Optimisation LP with its optimum at $x_1 = 5$ and $x_2 = 7$ with $f(x_1, x_2) = 1.74$

Optimisation - Motivating Questions



Optimising Logistics

As a junior consultant, a company asks you to develop a model that optimises distribution from its logistics centres to warehouses. Of course, the capacity of each logistics centre is limited, and delivery contracts have to be satisfied.

- How do you approach this problem?
- As potential data source, you can use data from a model that predicts the costs of delivery from each centre to each warehouse, as well as data about the required daily deliveries.

Optimising Logistics

Logistics Optimisation

- A company owns two logistics centres A and B
- Contracts require at least 50 units to be delivered today
- At most 40 units can be delivered from A, and 30 units from B
- Delivery costs are 3 per unit from A, and 5 per unit from B
- *What is the cost-minimising delivery plan?*

LP Formulation

- Decision variables are quantities delivered from A and B: x_A and x_B
- Optimisation function is minimise $3 \cdot x_A + 5 \cdot x_B$
- Non-negativity constraints: $x_A \geq 0$, $x_B \geq 0$
- Minimum total delivery constraint: $x_A + x_B \geq 50$
- Maximum capacity constraints: $x_A \leq 40$, $x_B \leq 30$

Logistics Optimisation

- A company owns two logistics centres A and B
- Contracts require at least 50 units to be delivered today
- At most 40 units can be delivered from A, and 30 units from B
- Delivery costs are 3 per unit from A, and 5 per unit from B
- *What is the cost-minimising delivery plan?*

LP Formulation

$$\min_{x_A, x_B \in R^2} 3 \cdot x_A + 5 \cdot x_B$$

subject to

$$\begin{aligned}
-x_A \quad & \leq -0 \\
- x_B & \leq -0 \\
-x_A - x_B & \leq -50 \\
x_A \quad & \leq +40 \\
+ x_B & \leq +30
\end{aligned}$$

Optimising Logistics: R Code

Example

Solving the Logistics Optimisation LP with lpSolve in R

```
library("lpSolve")
f.dir<-"min"
f.obj<-c(3,5)
f.con<-matrix(c(-1,-1,1,0,0,1),ncol=2,byrow=TRUE)
f.condir<-c("<=","<=","<=")
f.rhs<-c(-50,40,30)
obj<-lp(f.dir,f.obj,f.con,f.condir,f.rhs)
print(obj) # not very informative
if (obj$status==0){
  print(paste("Success. Optimum at ",toString(obj$solution),
    " with objective function value ",obj$objval))
  print("Checking which constraints are active (binding):")
  print(cbind(f.con%*%obj$solution, f.condir, f.rhs))
}else{
  print("error: optimisation was not successful")
}
```

Note: lp assumes non-negativity of all variables

Optimising Logistics

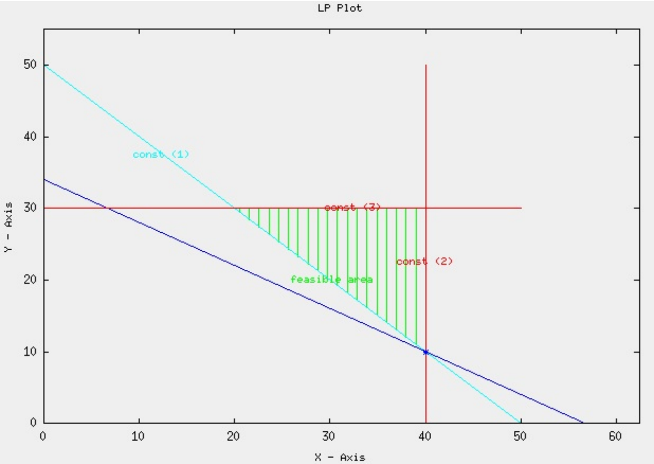
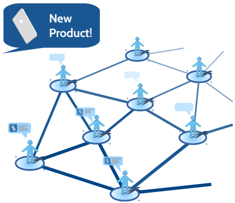


Figure: Plot of the Logistics Optimisation LP with optimum at $x_1 = 40$ and $x_2 = 10$ with $f(x_1, x_2) = 170$

Optimisation - Motivating Questions



Optimising Marketing

Your company will soon release a new product. Which customers should the company invite as product testers, such that news and recommendations of the product reach the maximum number of other potential customers?

- How do you approach this problem?
- You have data how customers reacted on previous product offerings, and you can obtain from a third party social network data about your customers.

Optimisation Problem Types: Overview

- Linear and Quadratic Programming Problems
  (mostly convex, "easy" to solve, e.g., SIMPLEX
- Quadratic Constraints and Conic Optimization Problems
- Integer and Constraint Programming Problems
  Challenge: inherently non-convex
- Smooth Nonlinear Optimization Problems
- Nonsmooth Optimization Problems

General Information on Optimisation with R

There are many packages for optimisation in R, see
`https://cran.r-project.org/web/views/Optimization.html`

Solving Linear Programs with R

- Package **lpSolve**
- Provides the `lp()`-function to solve LPs and MILPs, calls the freely available solver `lp_solve`
- Based on the revised simplex method and a branch-and-bound (B&B) approach
- **Assumes non-negativity** of all decision variables
- Provides also specialised solvers for assignment problems `lp.assign()` and transportation problems `lp.transport()`

Any More Questions?

Thank you!

Appendix

Bibliography I

Hand, D. J., Mannila, H., and Smyth, P. (2001).
Principles of Data Mining.
Adaptive Computation and Machine Learning. The MIT Press.

Sharda, R., Delen, D., and Turban, E. (2018).
Business Intelligence, Analytics, and Data Science: A Managerial Perspective.
Pearson, 4 edition.

Sherman, R. (2015).
Business Intelligence Guidebook: From Data Integration to Analytics.
Morgan Kaufmann.

Winston, W. L. (1997).
Operations Research: Applications and Algorithms.
Wadsworth Publishing Company, 3rd edition edition.