SAS® GLOBAL FORUM 2018

USERS PROGRAM

April 8 – 11 | Denver, CO
Colorado Convention Center

#SASGF

Prescriptive Analytics: Using Optimization with Predictive Models to find the Best Action

Dr. Mamdouh Refaat, Angoss Software (Datawatch)

Mamdouh Refaat is the SVP and Chief Data Scientist at Angoss Software where has been since 1999. He Leads the data science group in Angoss, a provider of data science software modeling tools.

He has used for more than 20 years and he authored two books on using SAS in data preparation for data mining, and for the development of credit risk scorecards.

Agenda

- About Optimization
- About Collections
- Collections Optimization
- A detailed example & a real case study
- Q&A

Why now ?

What, why, how …

Evolution of “Analytics”

Reports / Charts

Reports / Charts

What happened?

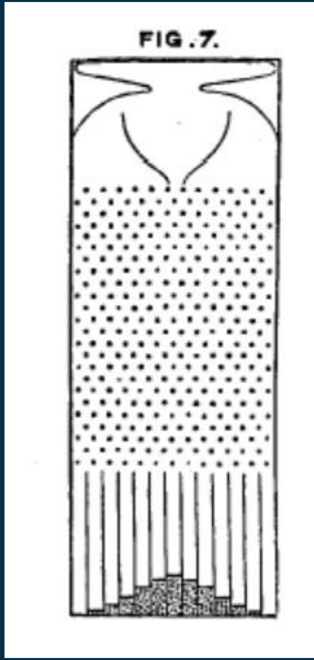
A Pie chart from William Playfair's "Statistical Breviary", **1801**

Statistical Analysis – Predictive Models

Regression models, correlation analysis, ANOVA

↓

How did/will it happen?



The original Galton's diagram explaining the frequency distribution (~1880)

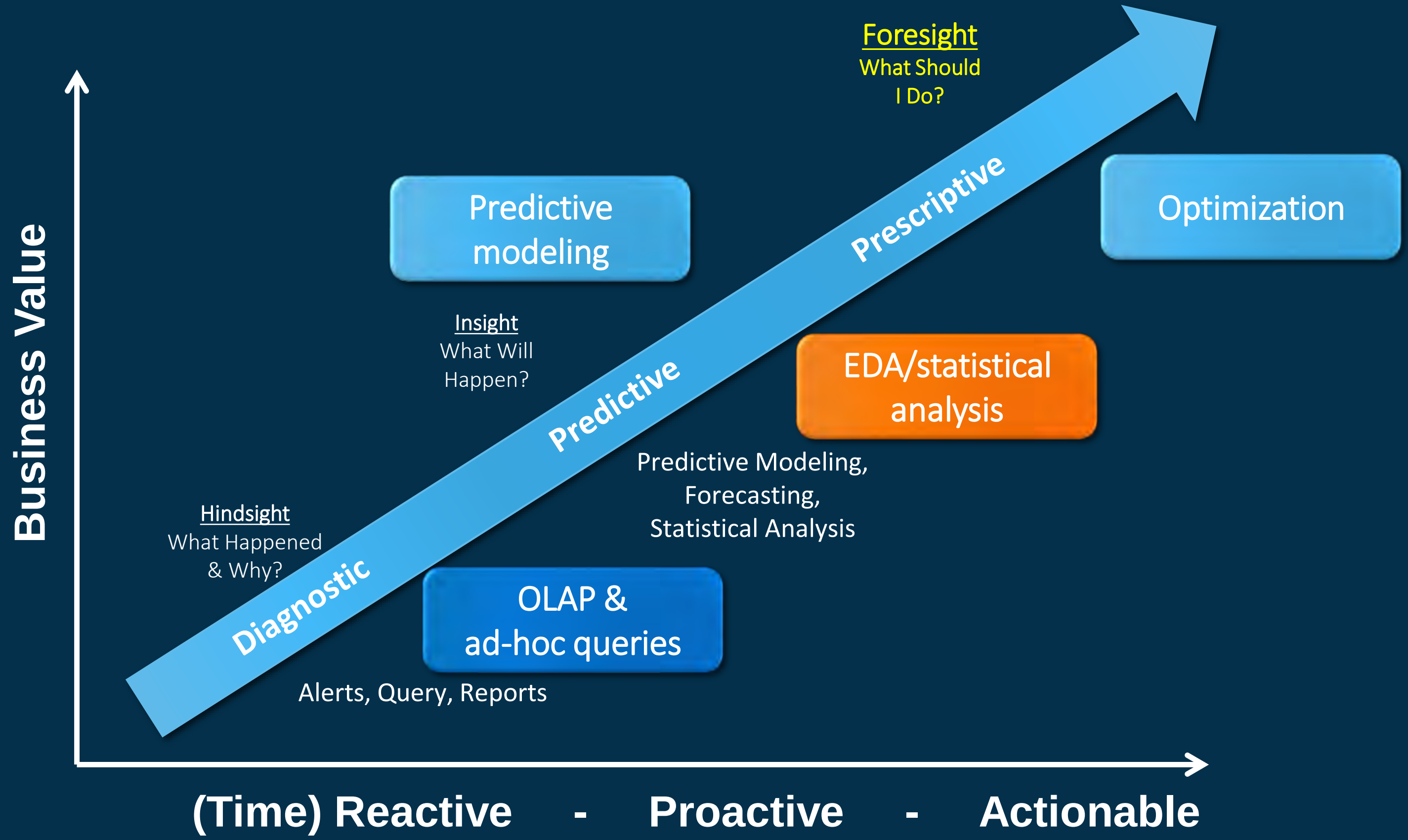
OLAP – Dynamic Reports

OLAP applications, Pivot Tables, Dynamic reporting

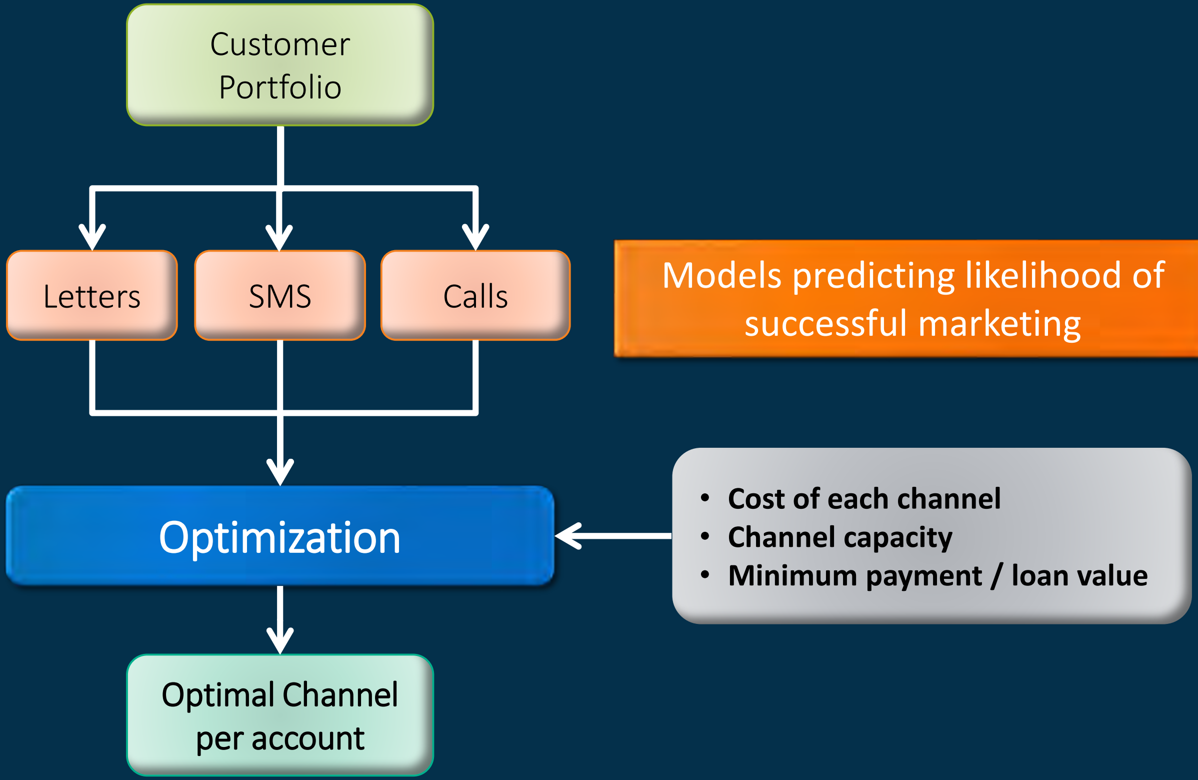
What happened (again!)?

Computers / Graphics → dynamic reporting / charting possible

Then and Now…The Analytics Journey



Optimization – what is the **best** action?



Optimization 101

Overview

Value

Make the <u>best decisions</u> under business <u>constraints</u>

Examples

- Marketing channel optimization
- Best of Collections Channel
- Optimization of Credit Offer – optimal price
- Best Product Offer for Acquisition

Constrained Optimization

Example:

Find **min** $F(x_1, x_2, x_3)$

subject to

$7x_1 - 2x_2 + 3x_3 \le 7$

$8x_1 + 9x_2 - 2.5x_3 \le 25$

$x_1, x_2, x_3 \ge 0$

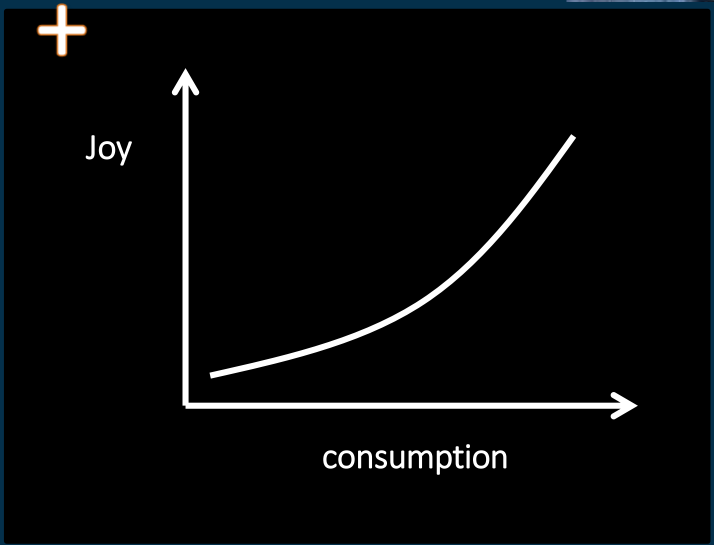
Objective Function

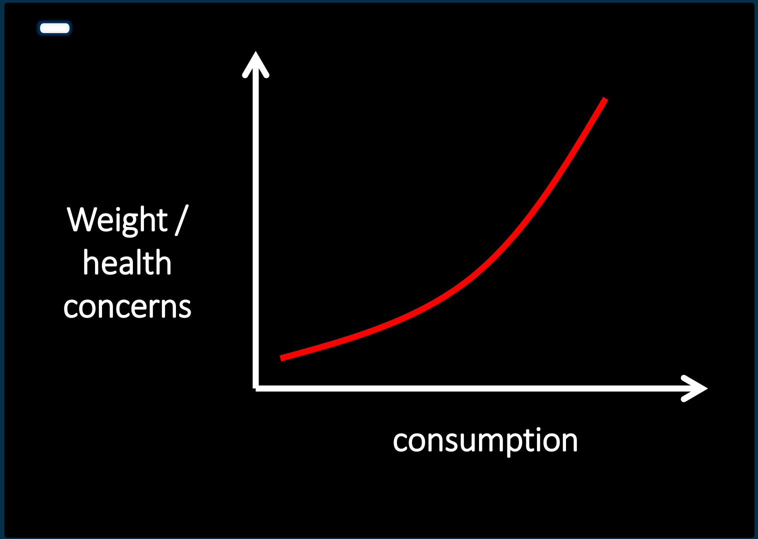
Decision variables

constraints

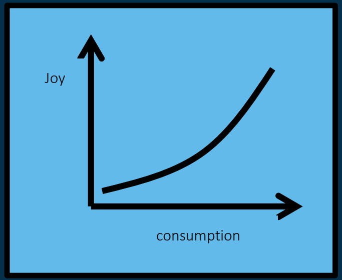
We try to find the decision variables x_1, x_2, x_3 that will minimize F while satisfying the given constraints

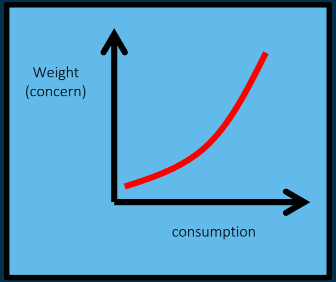
A more fun Example!





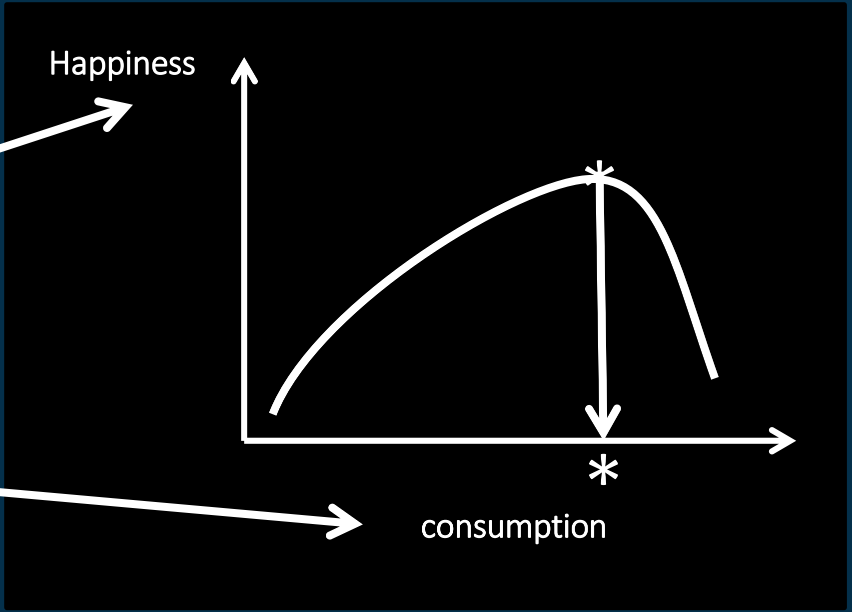
Happiness = Joy - Concerns



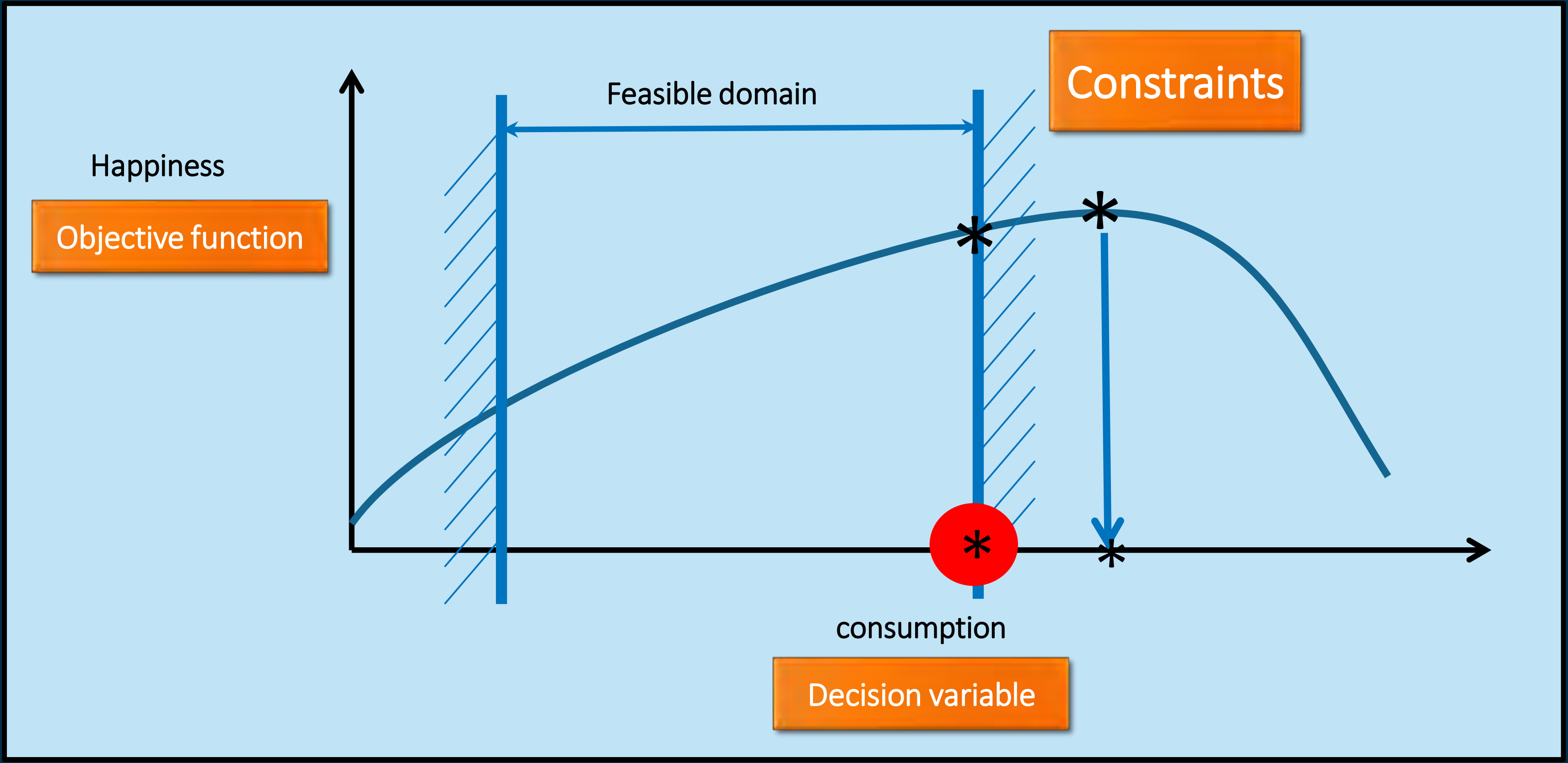


Objective function

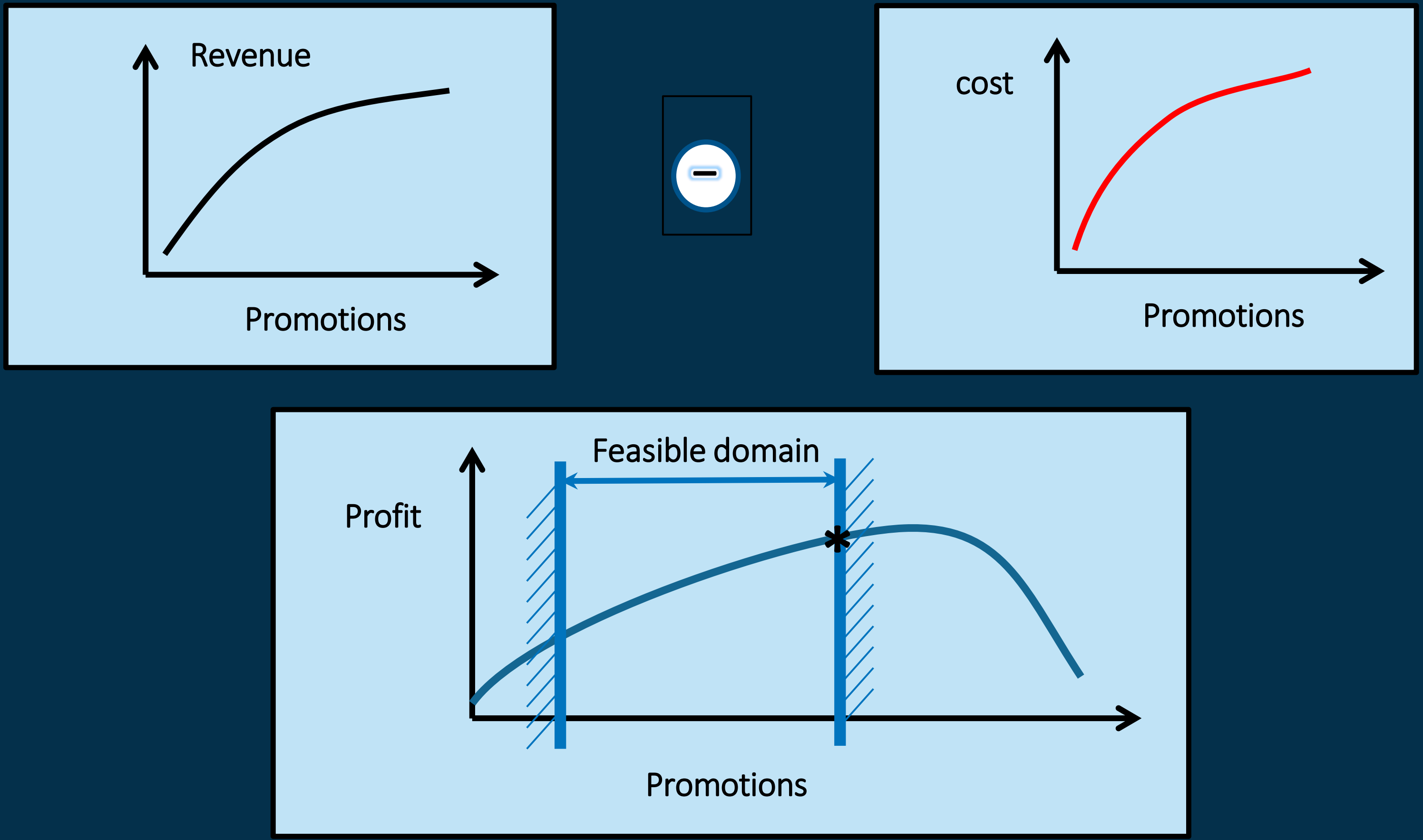
Decision variable



Constraints



Profit = Revenue - Cost



Marketing Channel →Maximum profit

Model 1: Response to a Telephone call P_T	$4.00	3,000/day
Model 2: Response to an email P_E	$0.02	200,000/day
Model 3: response to a letter P_L	$1.15	3,000/day
Daily budget		$8,000

Cust. ID	P_T	P_E	P_L
000-001	0.98	0.20	0.00
000-002	0.41	0.21	0.74
000-003	0.31	0.39	0.89
000-004	0.67	0.27	0.14
000-005	0.08	0.12	0.00
000-006	0.00	0.32	0.45
000-007	0.48	0.00	0.12

Additional constraints

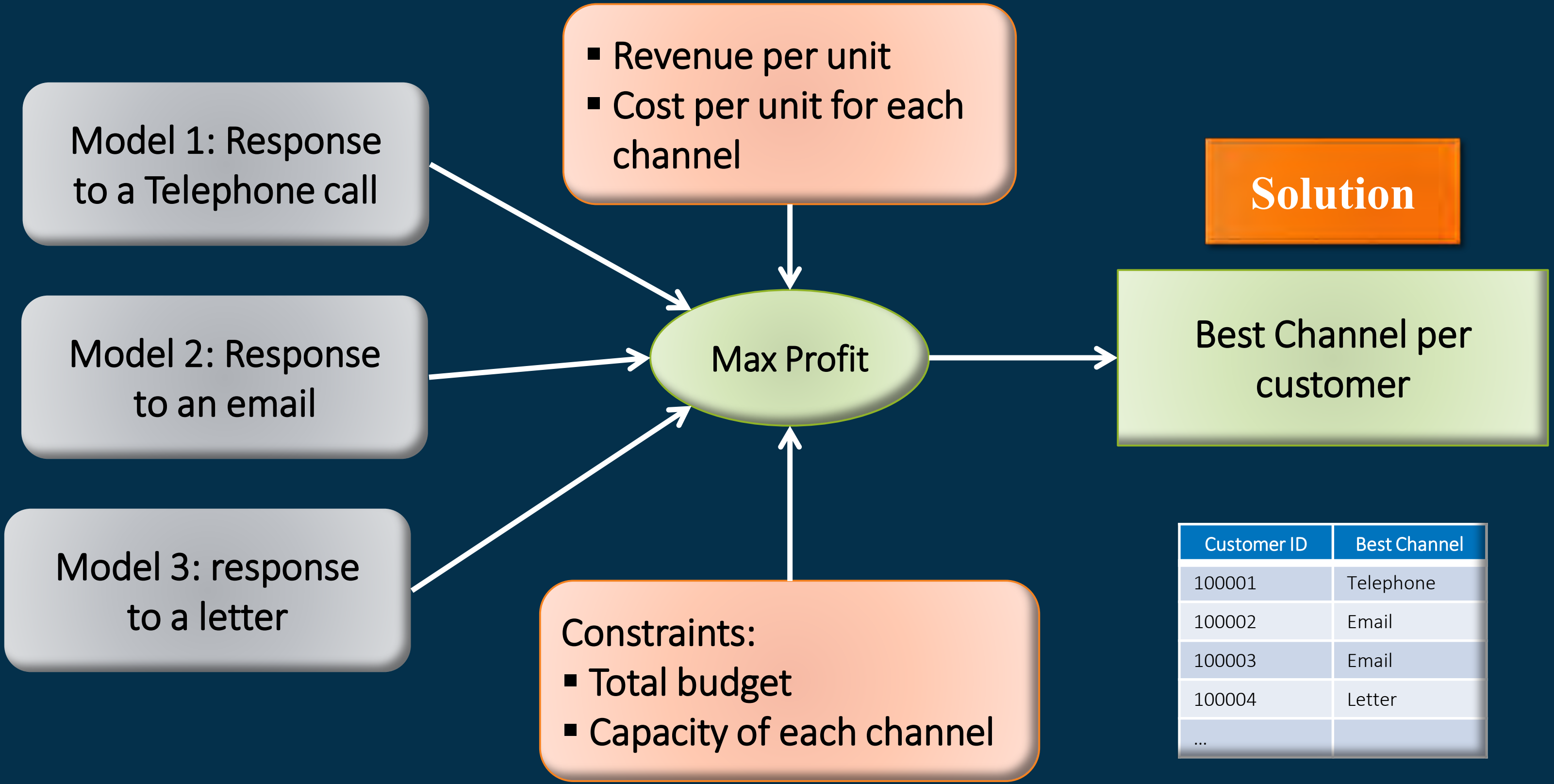
- Each customer is contacted by only **ONE CHANNEL**
- Some customers should not be contacted at all
- Any customer cannot be contacted more than twice a quarter

Cust. ID	P_T	P_E	P_L
000-001	0.98	0.20	0.00
000-002	0.41	0.21	0.74
000-003	0.31	0.39	0.89
000-004	0.67	0.27	0.14
000-005	0.08	0.12	0.00
000-006	0.00	0.32	0.45
000-007	0.48	0.00	0.12

Cust. ID	Marketing Channel - today
000-001	Telephone call
000-002	Letter
000-003	---- (no contact) ---
000-004	…
000-005	…
000-006	…
000-007	…

Required Solution

The Marketing Channel Optimization Problem



Customer ID	Best Channel
100001	Telephone
100002	Email
100003	Email
100004	Letter
…	

Classification of Optimization Problems

and Solution Techniques !

Components of Optimization

Decision variables

- Continuous
- Integer
- Binary

Objective Function

- Linear
- Quadratic
- Nonlinear

Constraints

- Linear
- Nonlinear
- Conditional

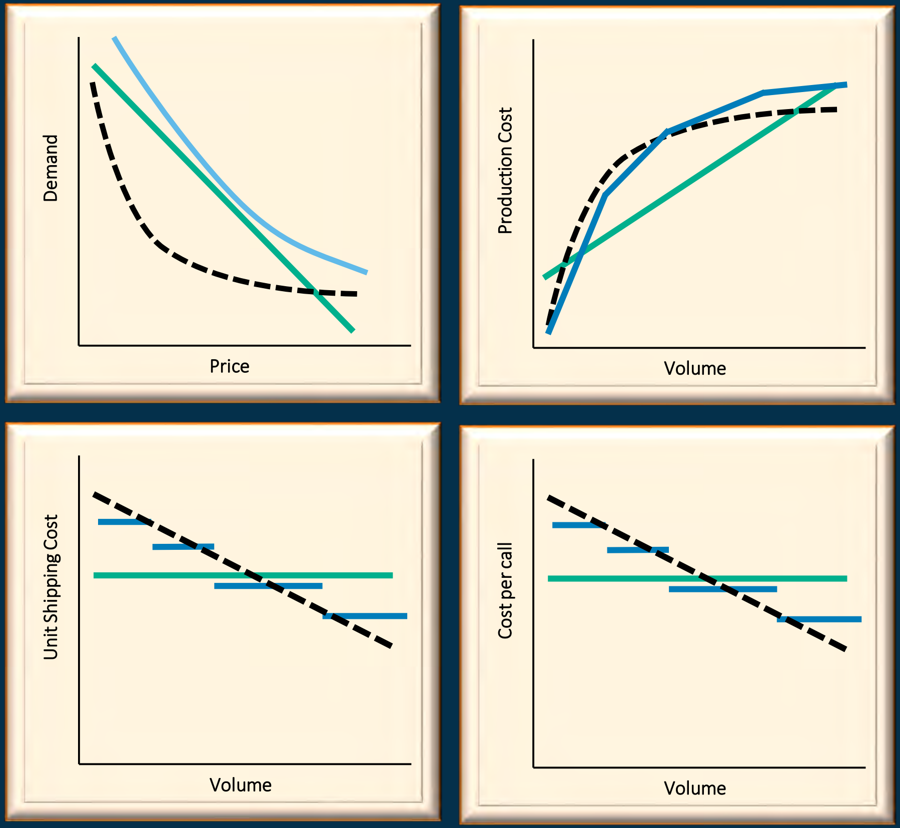
Programming ←→ Optimization !

Decision variables – which type?

- Continuous values (real numbers)
  - Balance, Profit, loss
  - Interest rate
  - Temperature
  - Volume / weight / times …
- Discrete (Integer) values (any non-divisible entity)
  - Number of units
  - Number of products
  - Number of customers
- Binary variables
  - Actions (Call/Don’t call)
  - Mutually exclusive states (Good/Bad status)

Objective Function

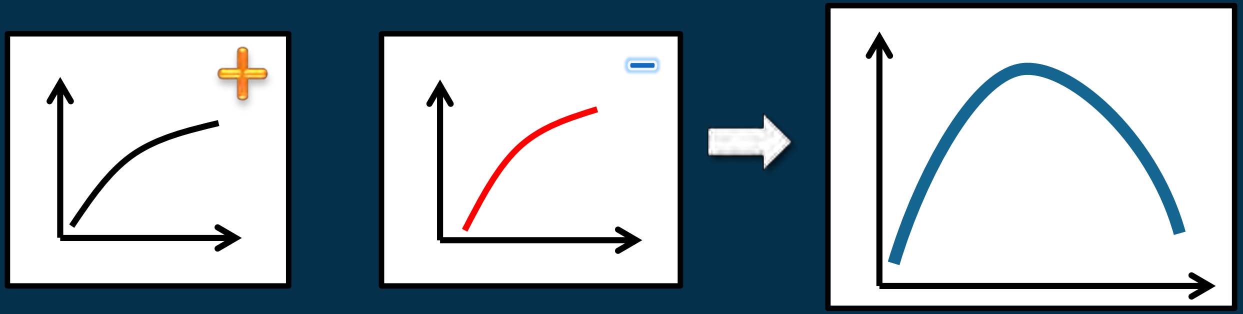
- Price-demand is (linear, quadratic, other?)
- Production Cost - volume is (linear, piecewise linear, quadratic)
- Shipping cost –volume is (constant, piecewise constant, linear)
- Cost per calls (in a call center)-volume is (constant, piecewise constant, linear, other)



Creating Objective Functions

Conflicting objectives

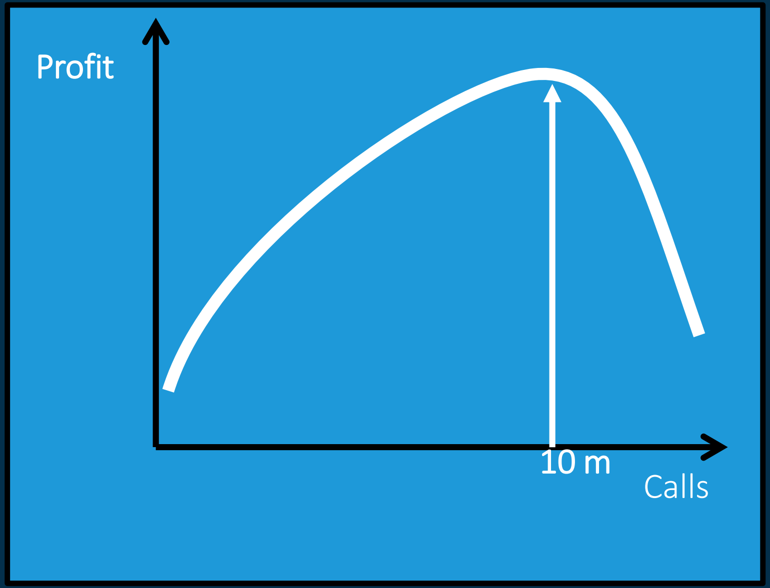
Represents the business objective as a "**Function**" of the decision variables



- The units of the objective function are the **units** of the business objective
- We attempt to translate conflicting relations using a **single unit**
- **Money/time/counts** are the most common units of objective functions

Constraints

Reality of the business problem



Limitations of real life

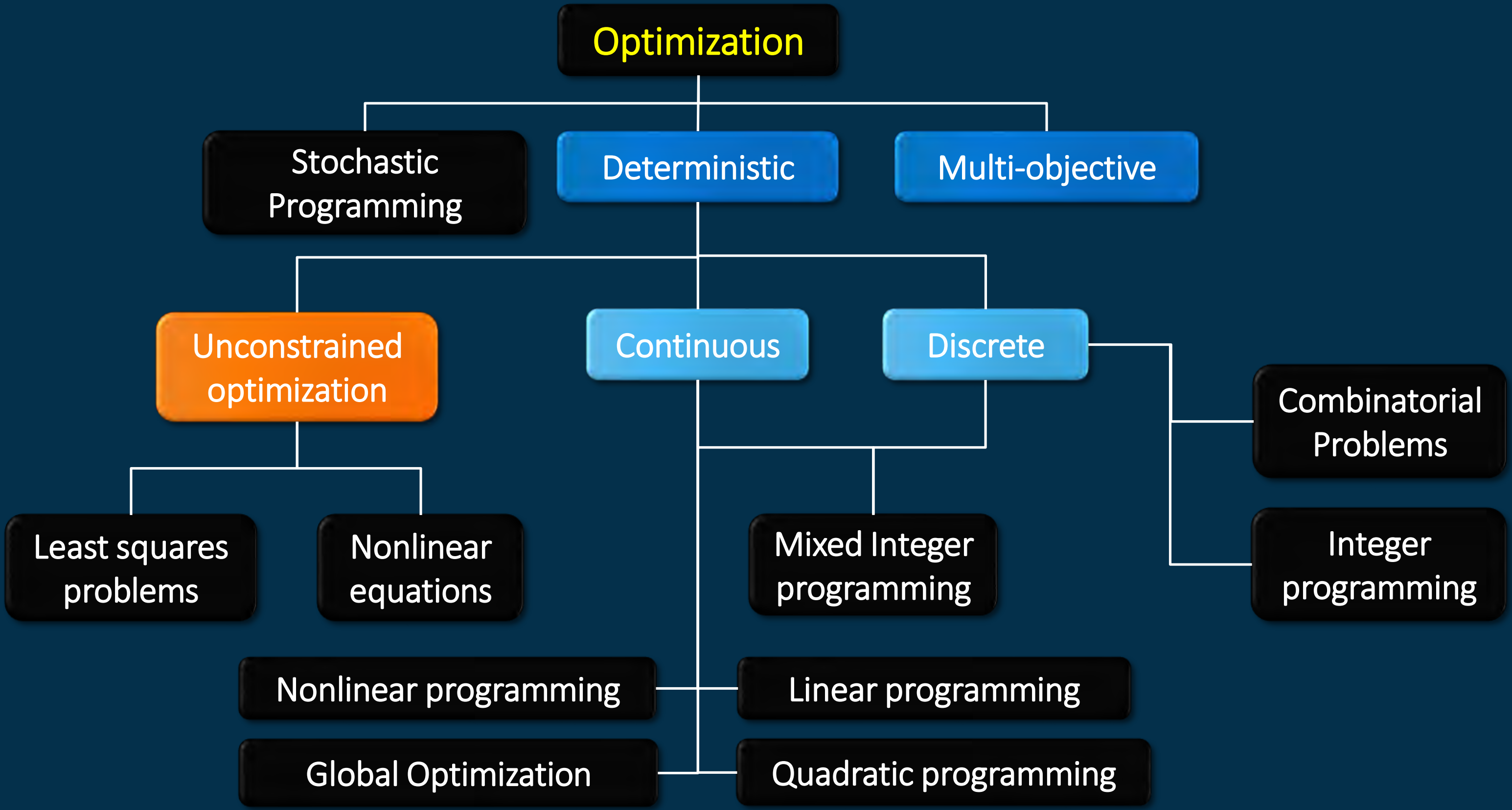
- Resources are not unlimited
- Actions are not independent
  - If we call the customer, no need for SMS
  - We cannot call a customer more than a certain number of times every quarter

Classification of Optimization Problems (1)

Decision Variables	Objective Function	Constraints	Problem Name
Continuous	Linear	Linear	Linear Programming (LP)
Continuous	Quadratic	Linear	Quadratic programming (QP)
Integer / binary	Linear	Linear	Integer programming (IP)
Mixed integer and continuous	Linear	Linear	Mixed Integer-linear programming
Continuous	Nonlinear – convex	Nonlinear – convex	Convex programming (CP)
Continuous	Nonlinear	Nonlinear	Nonlinear programming (NLP)
Mixed integer and continuous	Nonlinear	Nonlinear	General mixed nonlinear programming

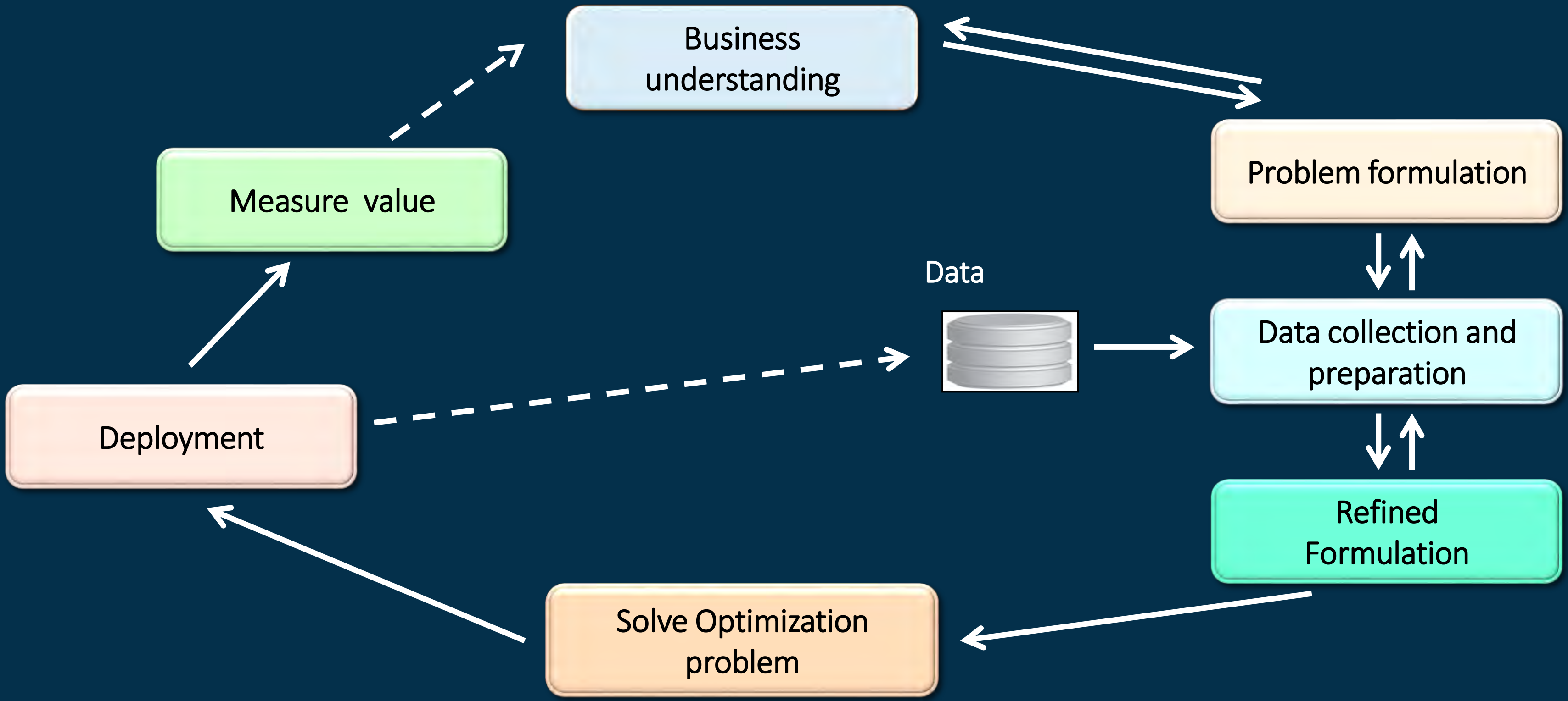
Programming ≡ Optimization

Classification of Optimization Problems (2)



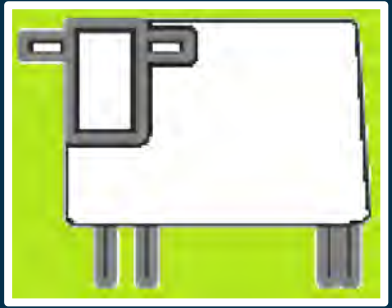
Optimization Methodology

Optimization Methodology



Realism vs. simplicity

Approximation of a <u>cow</u>'s geometry

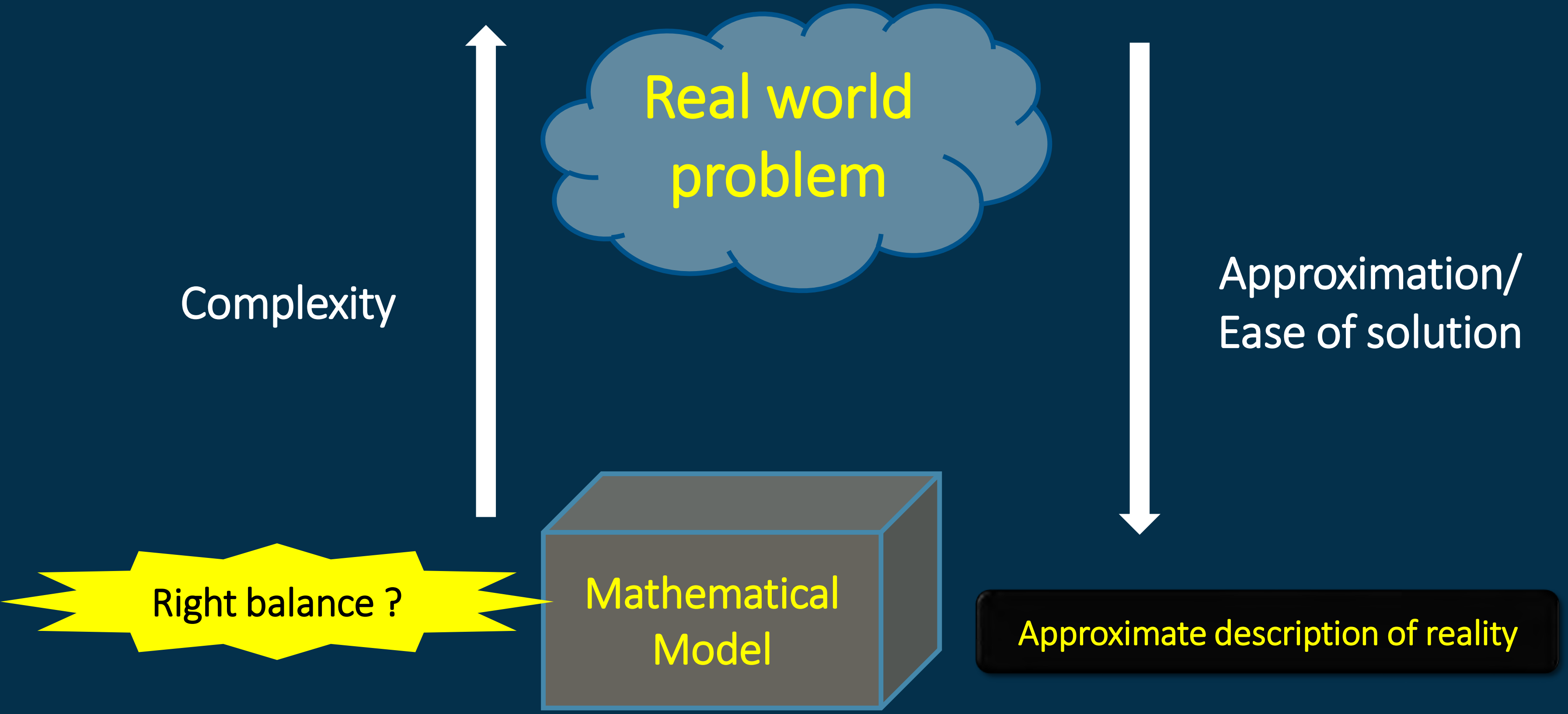


The rectangular cow model

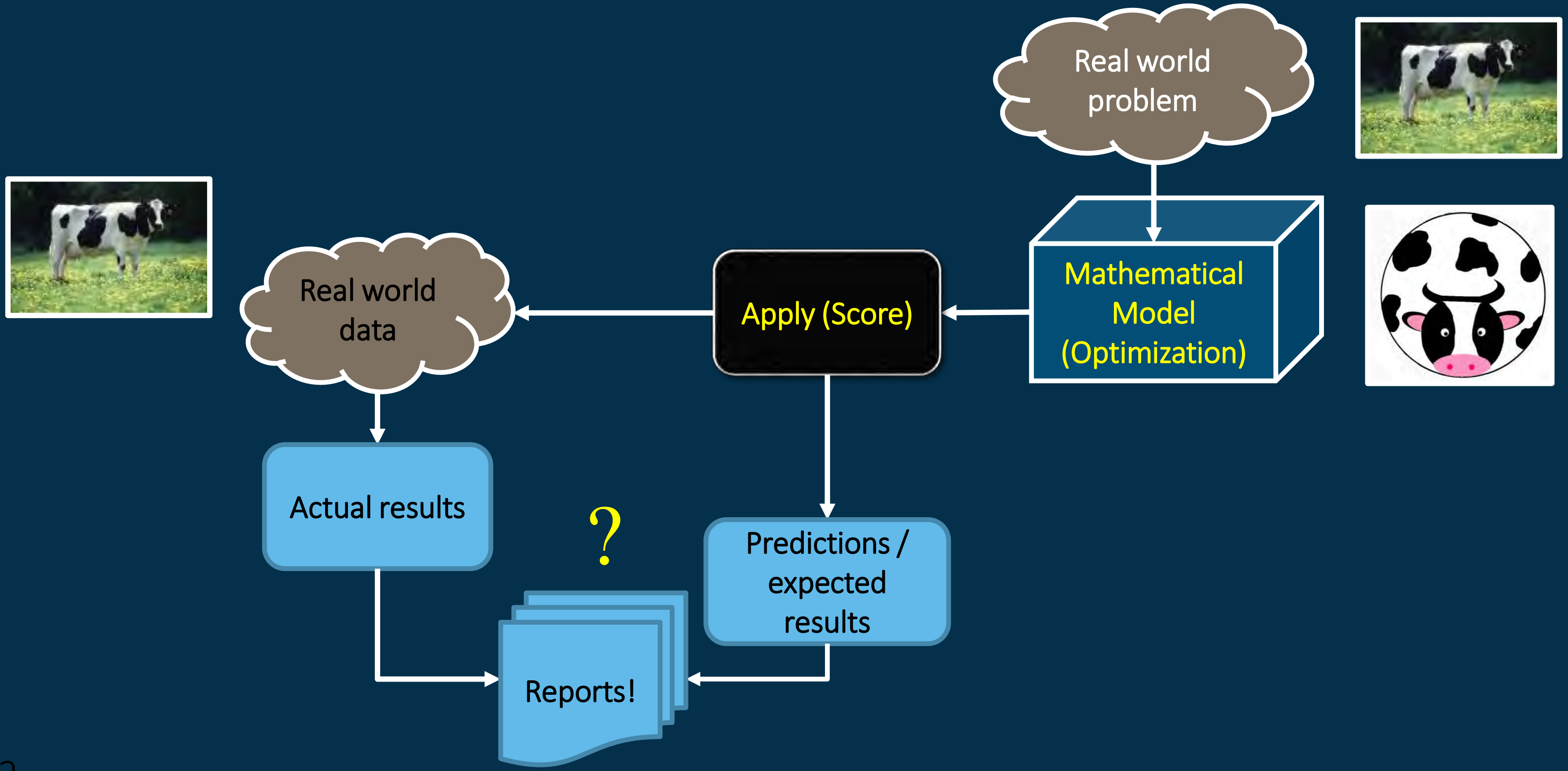
?

The idealized spherical cow

Real World ←→ Mathematical Model



Implications – Accuracy, validation, reporting, …



The Solution of Optimization Problems

Find: decision variables

$x_1, x_2, \ldots$

→ Min(max) $F(x_1, x_2, \ldots)$

Subject to

$g_1(x_1, x_2, \ldots) \leq 0$

$g_2(x_1, x_2, \ldots) \leq 0$

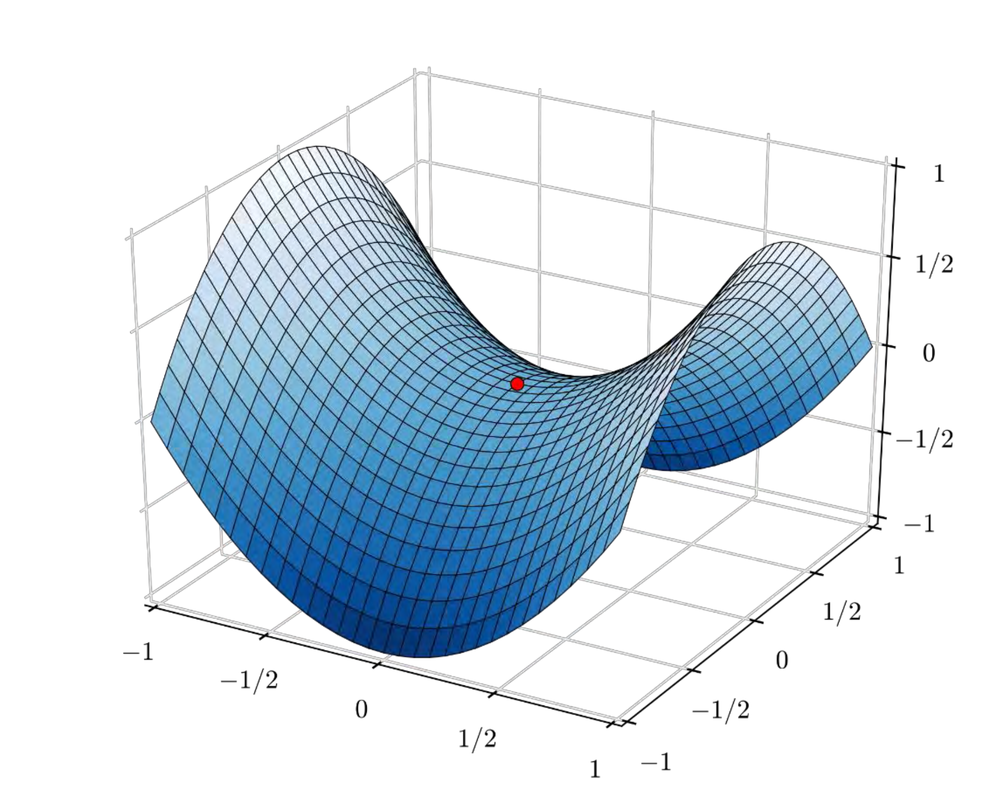
$\ldots$

- The solution
  1. Existence of a solution (exists, bounded, feasible)
  2. Value of decision variables $x_1, x_2, \ldots$
  3. Value of objective function at optimal solution
  4. State of constraints (free or active)

Can't we always find a solution ? !!

No Optimal Value: Minimum = Maximum !

Saddle Point



How can a real problem have a saddle point?

x=Balance

y = Interest rate

Risk = $(x-1000)^2 - (y-1.2)^2$

Saddle Point at

x = 1000

y = 1.2

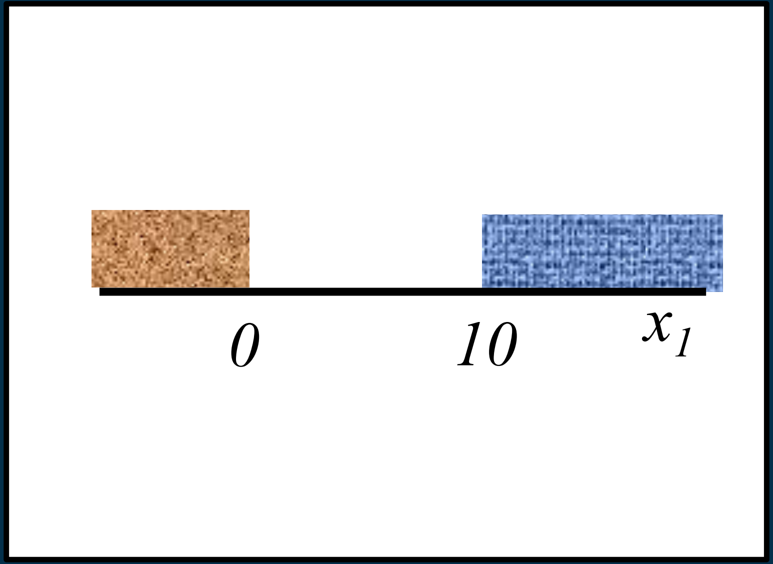
Infeasible Problems

Infeasible problem

Max $F = 9\,x_1$

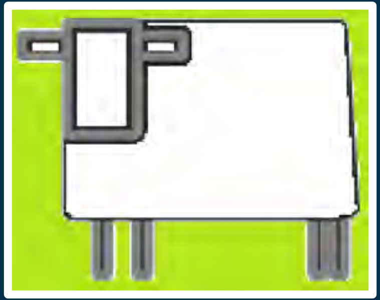
Subj $x_1 > 10,$

$x_1 < 0,$



How can a real problem be infeasible?

Maximize



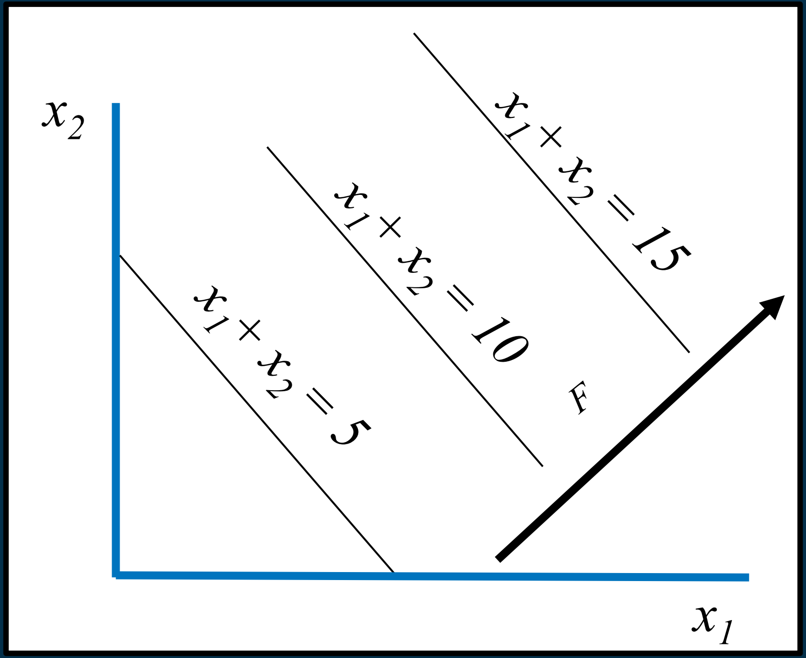
Cows ≥1

Not Feasible

Unbounded Objective Function

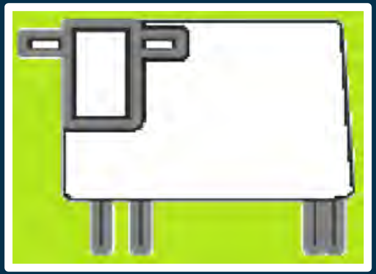
Unbounded problem

Max $F = x_1 + x_2$



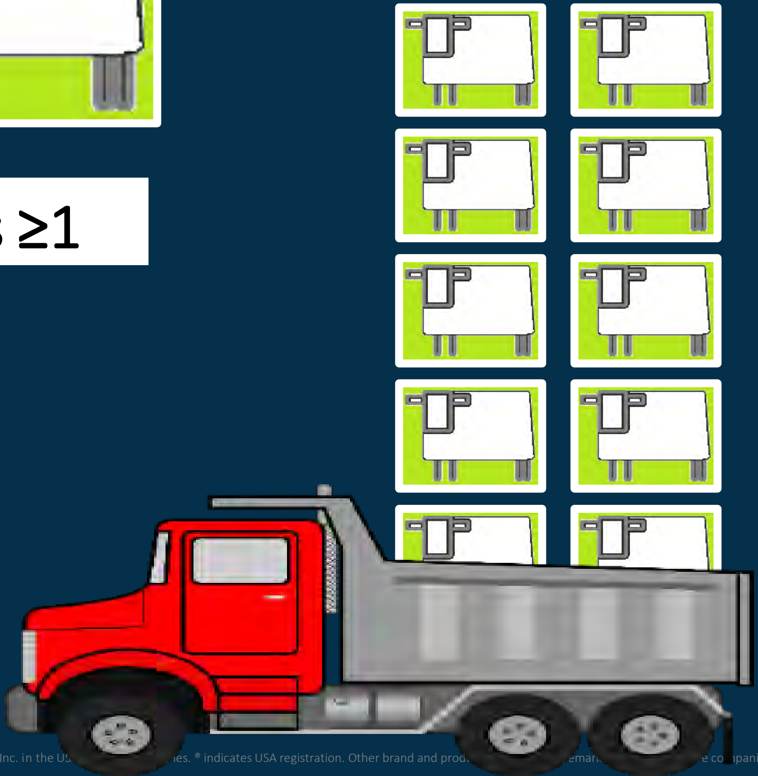
How can a real problem be unbounded ?

Maximize

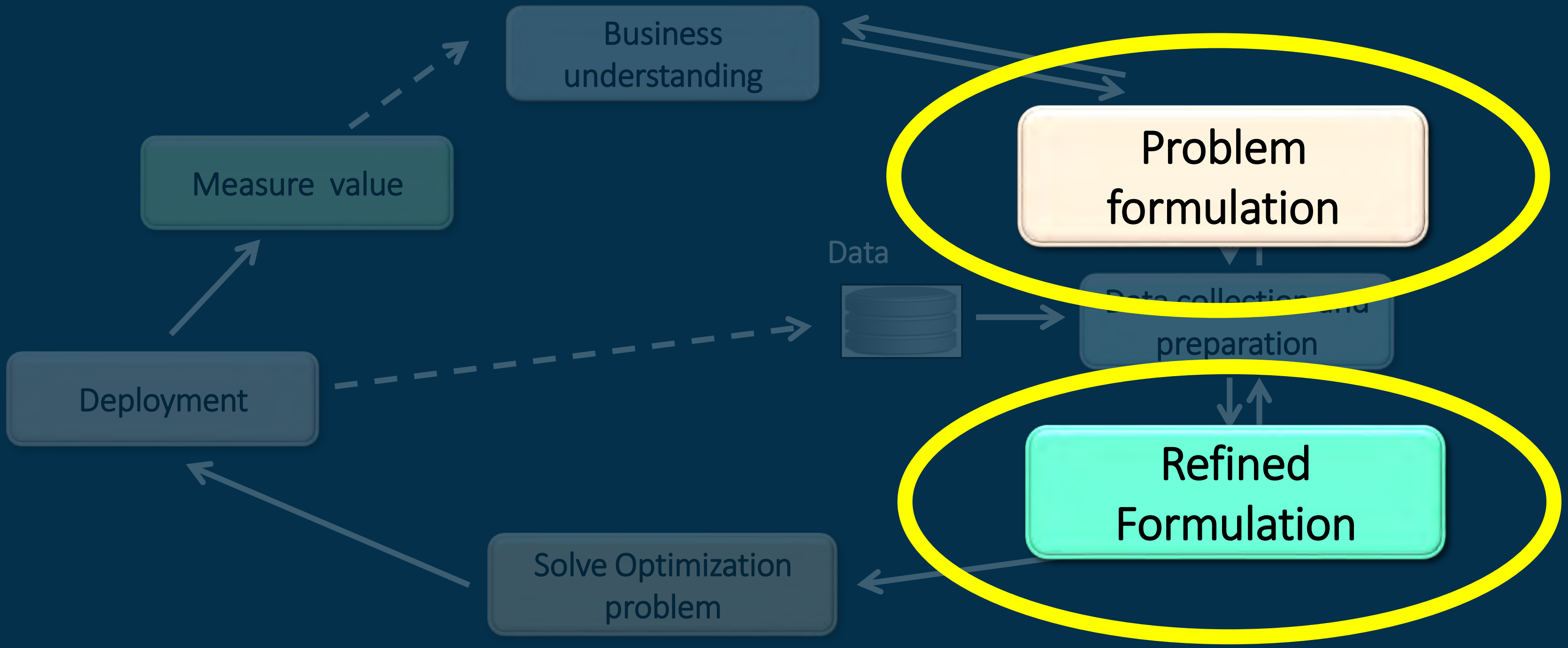


Cows ≥1

Unbounded !!



Source of these problems is …



The Solution of Optimization Problems

Find: decision variables
$x_1, x_2, \ldots$

→

Min(max) $F(x_1, x_2, \ldots)$

Subject to

$g_1(x_1, x_2, \ldots) \le 0$
$g_2(x_1, x_2, \ldots) \le 0$
$\ldots$

- The solution
  1. Existence of a solution (exists, bounded, feasible)
  2. Value of decision variables $x_1, x_2, \ldots$
  3. Value of objective function at optimal solution
  4. State of constraints (free or active)

Lessons after the solution – **Sensitivity Analysis**

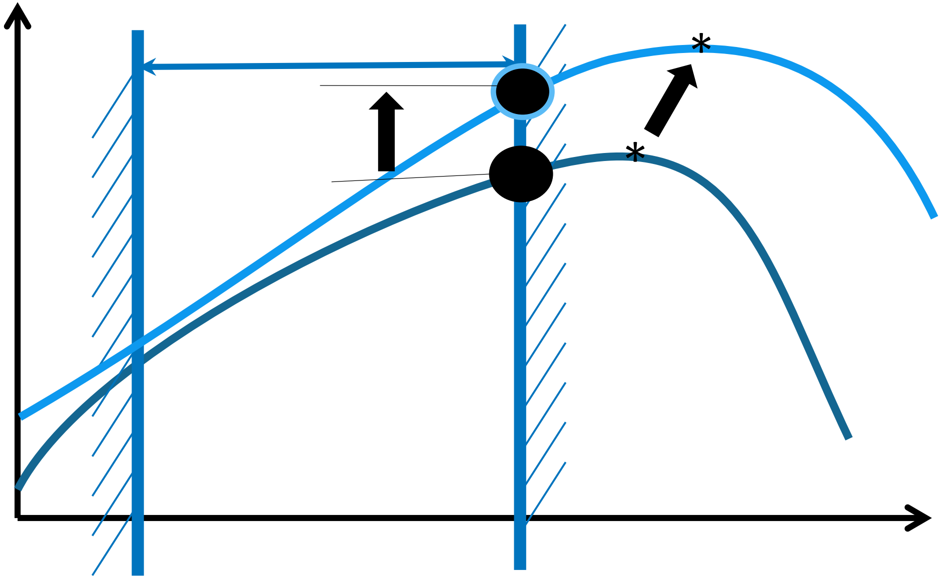
$F(x_1, x_2, \ldots)$

$g_1(x_1, x_2, \ldots) \leq 0$
$g_2(x_1, x_2, \ldots) \leq 0$
$\ldots$

- Value of decision variables $x_1, x_2, \ldots$?
- Value of objective function at optimal solution ?
- State of constraints (free or active) ?

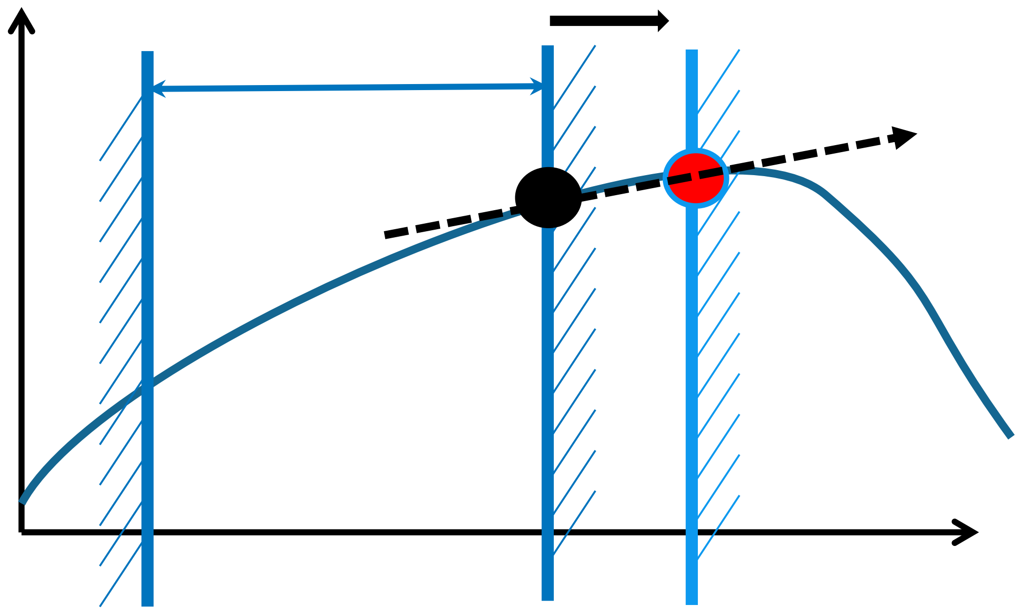
Values and trends

Sensitivity analysis – Objective Function



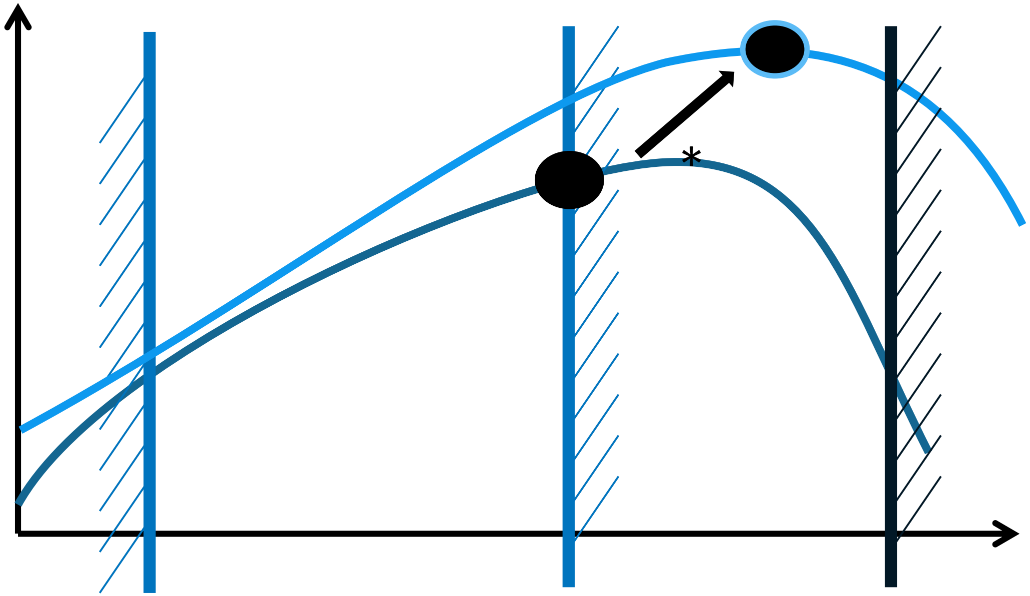
Effect of change in the objective function

Sensitivity analysis - constraints



Effect of change in the Constraints

More on Sensitivity Analysis



The optimization problem may turn into an unconstrained problem!

How to solve optimization problems

How to NOT solve optimization problems!

Optimal Channel Assignment

Email Call

Solution 1

ID	Ch
1	E
2	E
3	E
4	E
5	E

2

ID	Ch
1	C
2	E
3	E
4	E
5	E

3

ID	Ch
1	E
2	C
3	E
4	E
5	E

4

ID	Ch
1	E
2	E
3	C
4	E
5	E

5

ID	Ch
1	E
2	E
3	E
4	C
5	E

ID	Ch
1	E
2	E
3	E
4	E
5	C

...

ID	Ch
1	C
2	C
3	E
4	E
5	E

32

ID	Ch
1	C
2	C
3	C
4	C
5	C

Number of possible solutions = 2x2x2x2x2= $(2)^5$ = (channels)^(records)

A Realistic Small Collections Portfolio

~ 200,000 accounts

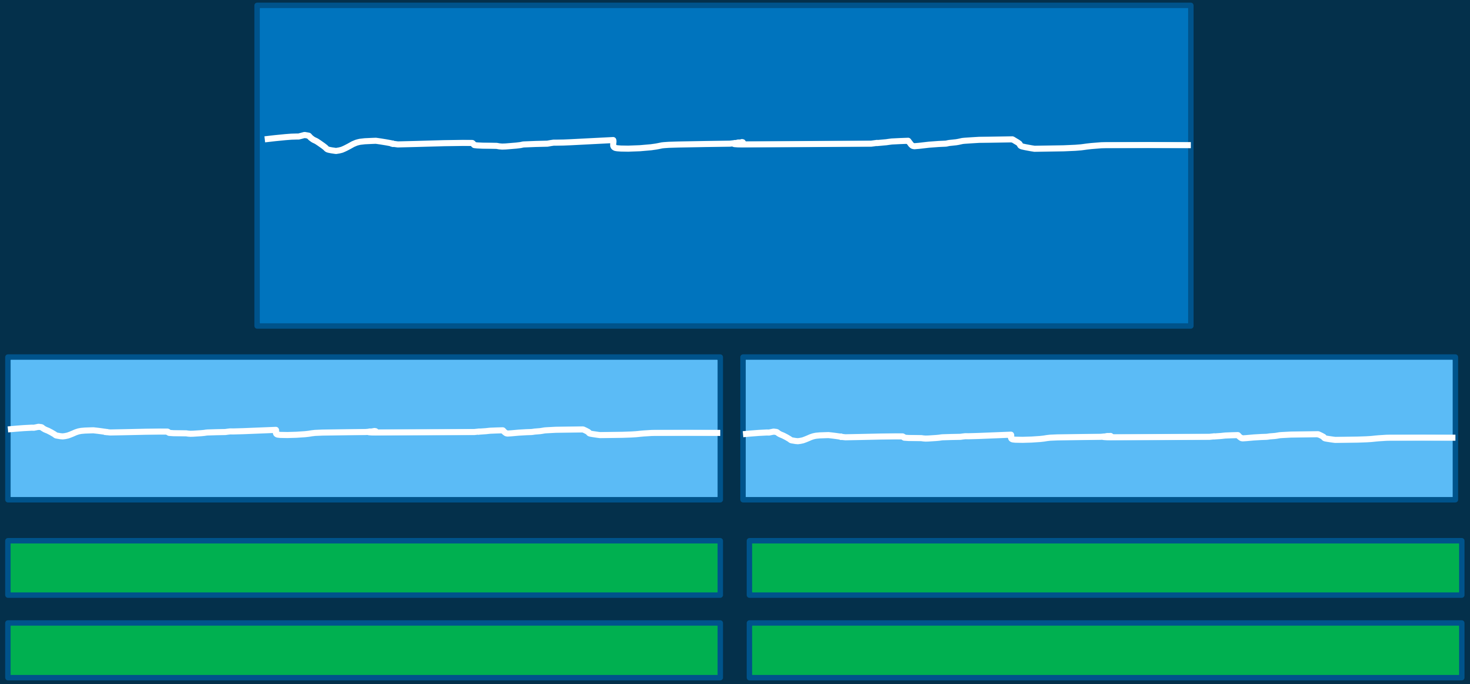
Number of solutions $(2)^{200,000} = (2x2x2x2)^{50,000} = (16)^{50,000} > (10)^{50,000}$

This is a **VERY LARGE NUMBER** of solutions to test

How large is $10^{50,000}$?

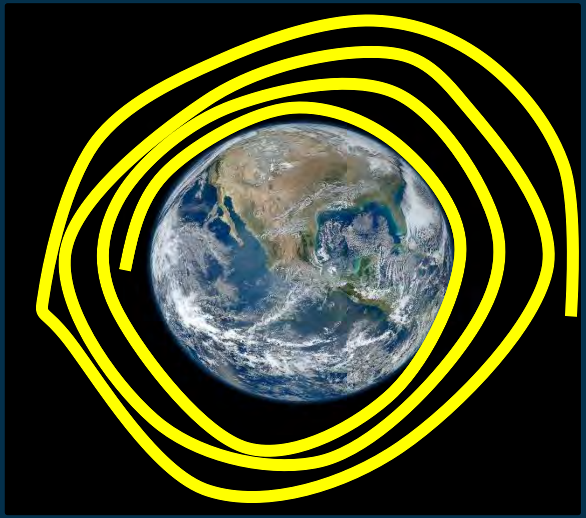
Atoms are everywhere!

How big is an atom?



Atoms are really, really ***SMALL***

The atoms in a tablespoon of salt would create a chain that is **3.5 times** the circumference of the **planet Earth**



Putting things in perspective!

$5 \times (10)^{23}$ Atoms in a tablespoon of water

$(10)^{82}$ Atoms in the entire Universe

Our **Small** Collections problem has more than **$(10)^{50,000}$** Solutions

That is $(10)^{49,918}$ times more than the number of atoms in the universe

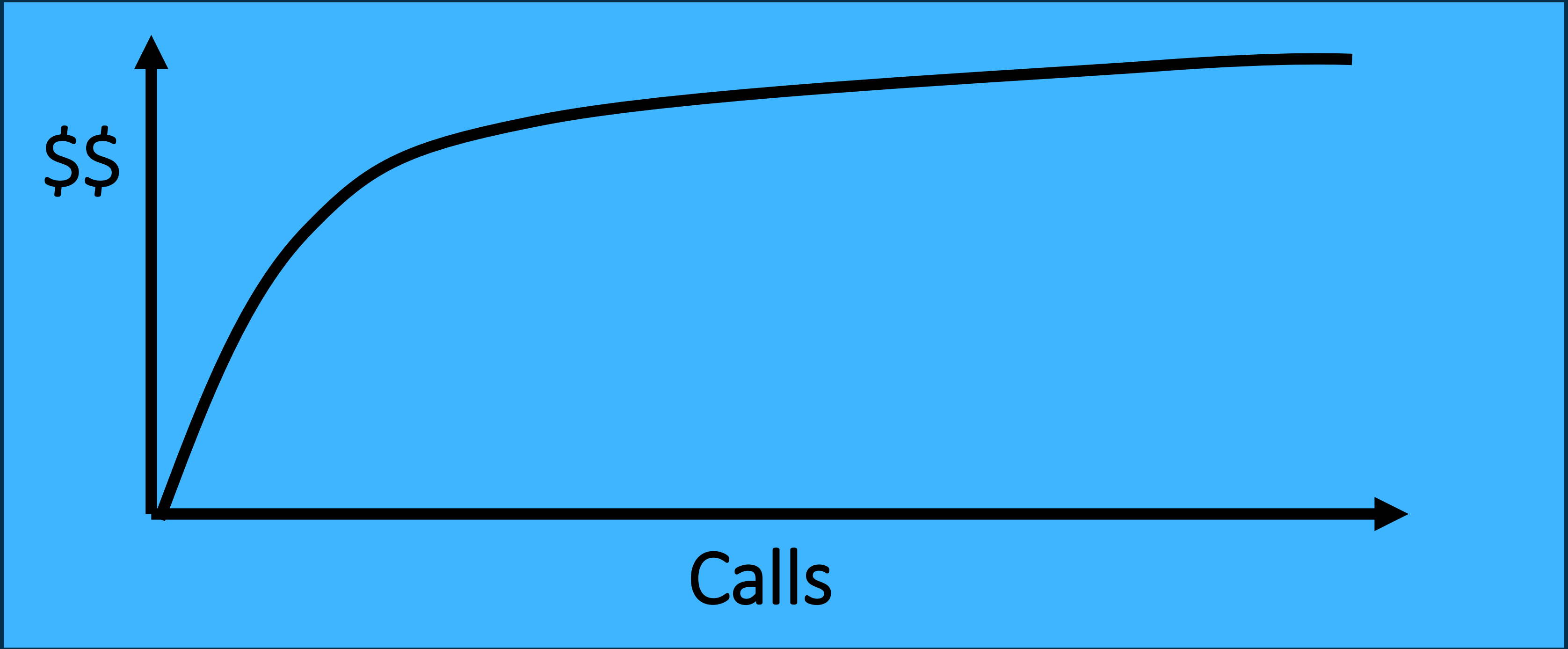
So, what is the lesson?

- Combinatorial problems are HARD
- We don’t solve them by trying EVERY possible solution
- We need special algorithms ➔ Software tools

Few things about Collections

(difficulties!)

Effort vs. Results



How? (Channels)

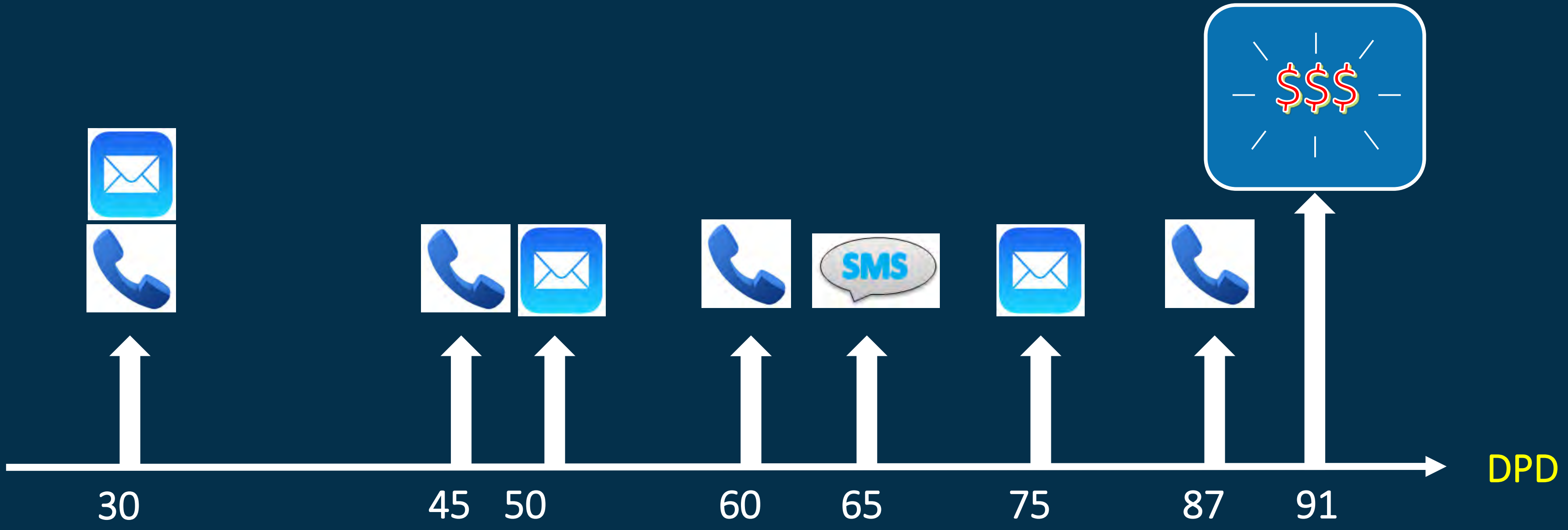
- Not all channels have the same effectiveness
- Not all respond same way to contacts using the different channels
- There are regulatory restrictions on writing off debts
- Collections budgets are limited
- There exist a large number of *accepted wisdom* in the form of "<u>Collections Strategies</u>"
- Collections strategies focus on sequence and timing of actions (30 DPD: Letter → 45 DPD: Call → …)

Definition of Successful Collection

- Spoke / Connected / Engaged with account holder
- Promised to pay
- Agreed on a payment plan
- Made a partial payment
- Made a full payment

When to remove the account from the collections queue ?

Attribution



Which action caused (triggered) payment ?

Vantage Analysis

Current Period (month) (columns) / **Previous Period (month)** (rows)

	30-60 DPD	60-90 DPD	90-120 DPD	120-180 DPD	> 180 DPD
30-60 DPD					
60-90 DPD					
90-120 DPD					
120-180 DPD					
> 180 DPD					

Bad

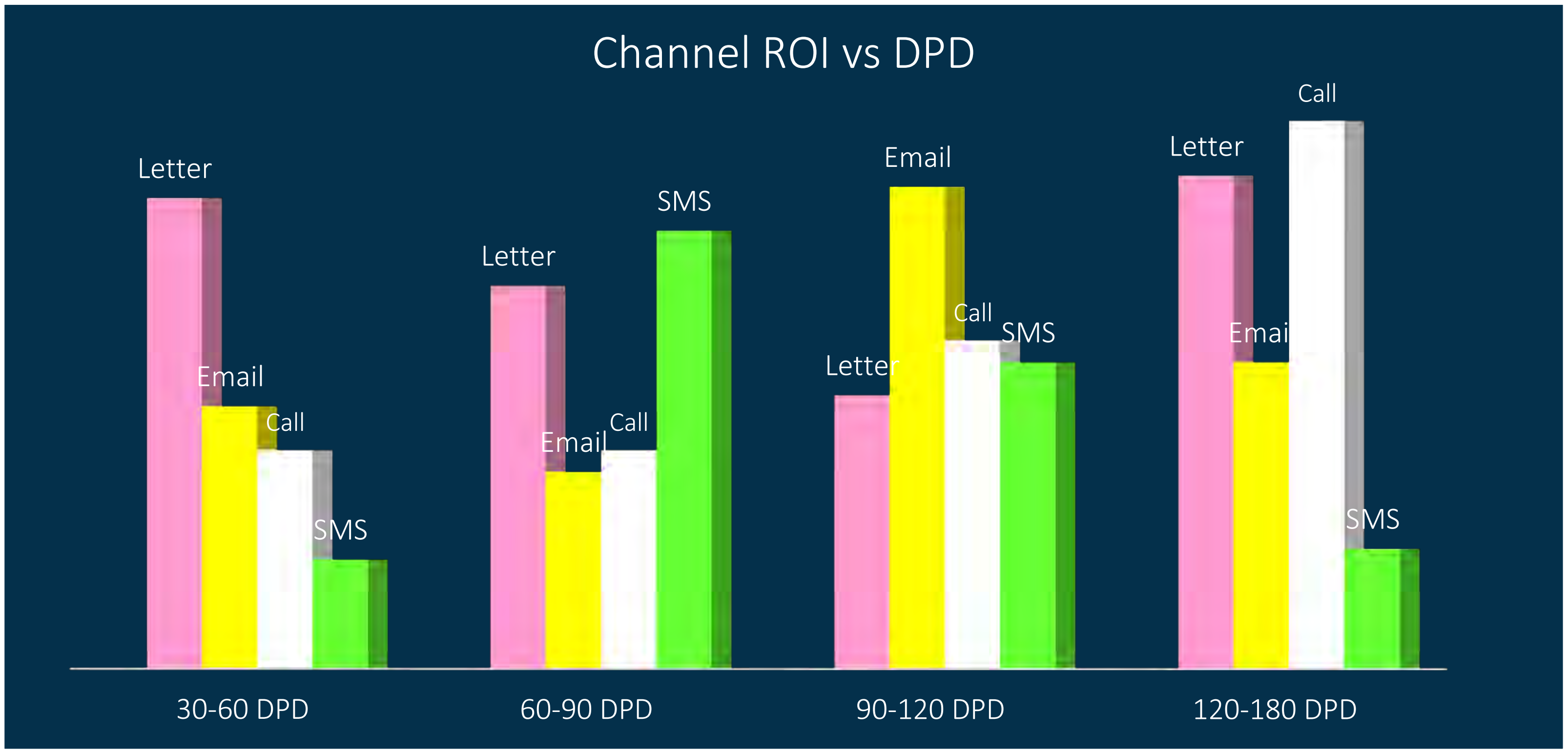
Good

Models … Models …

- Likelihood to Self-cure
- Likelihood to pay back (any amount, %, all)
- Estimate of account to recover
- Likelihood to Write-off
- Likelihood to accept restructuring
- Likelihood to promise to pay
- Likelihood to connect/engage with the right person
- Likelihood to "respond" using a specific channel

Which model ?

Channel ROI

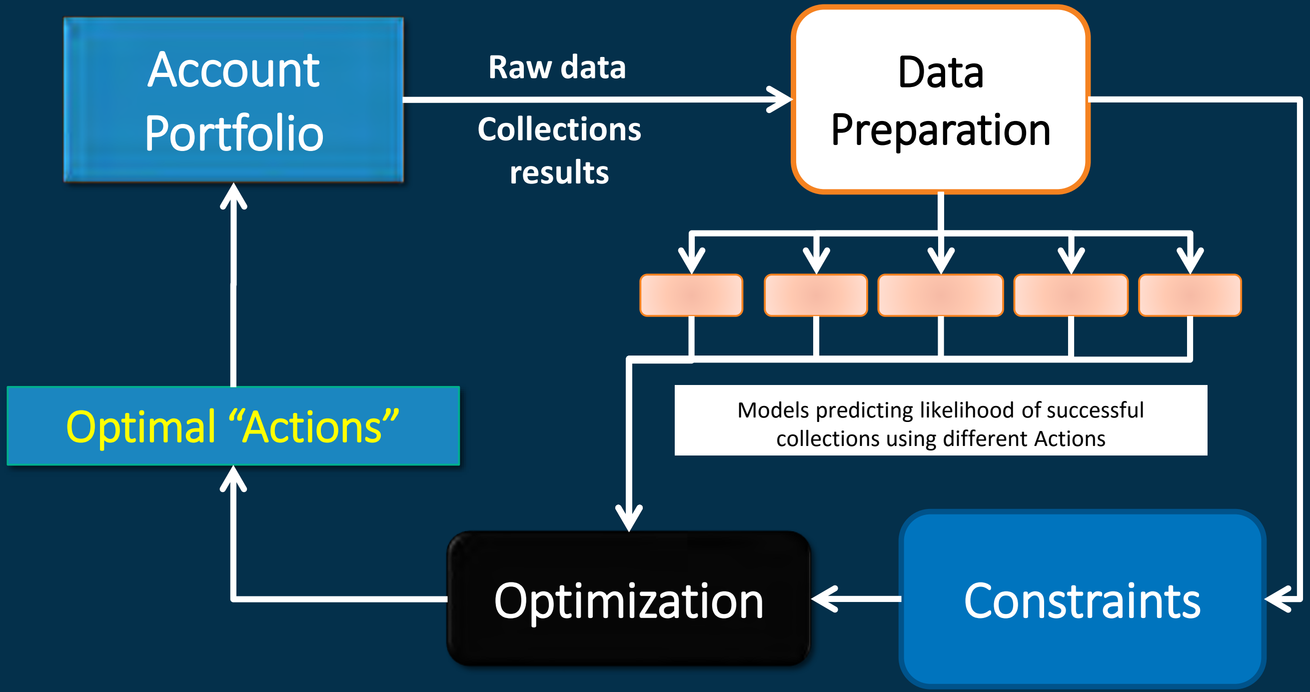

Channel ROI vs DPD
Letter
Email
Call
SMS
Letter
Email
Call
SMS
Letter
Email
Call
SMS
Letter
Email
Call
SMS
30-60 DPD
60-90 DPD
90-120 DPD
120-180 DPD

Collections Optimization

Finally !!!

Collections Optimization

The framework



Back to the **basics**!

Optimization problems consist of

1. Decision variables
2. Objective function
3. Constraints

Problem Setup

Decision Variables

- Optimal channel → Channel(s) (1/0)
- When to contact → DPD / time of day (1/2/3)
- Message intensity → Level (1/2/3/...)
- Best action → collect, write-off, restructure
- Who Collects → internal, outsource (a/b)
- Agent assignment → which agent (1, ..., 86)

Contact strategy → Channel + when + intensity + who

Problem Setup

Complexity

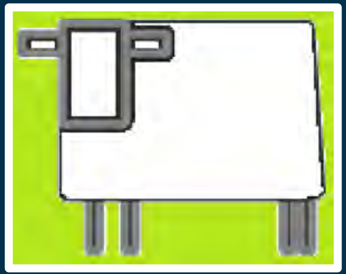
Real world problem

Approximation/
Ease of solution

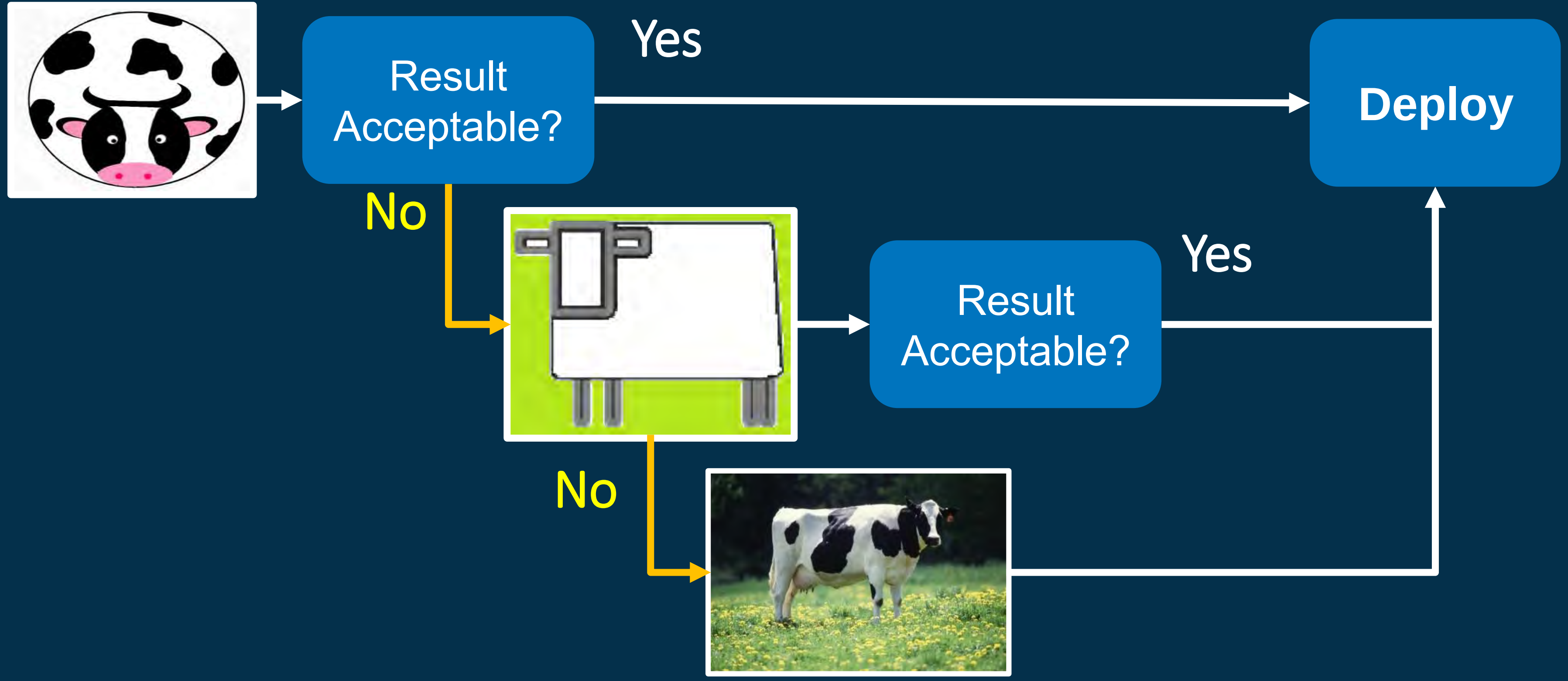
Right balance ?

Mathematical Model

Approximate description of reality

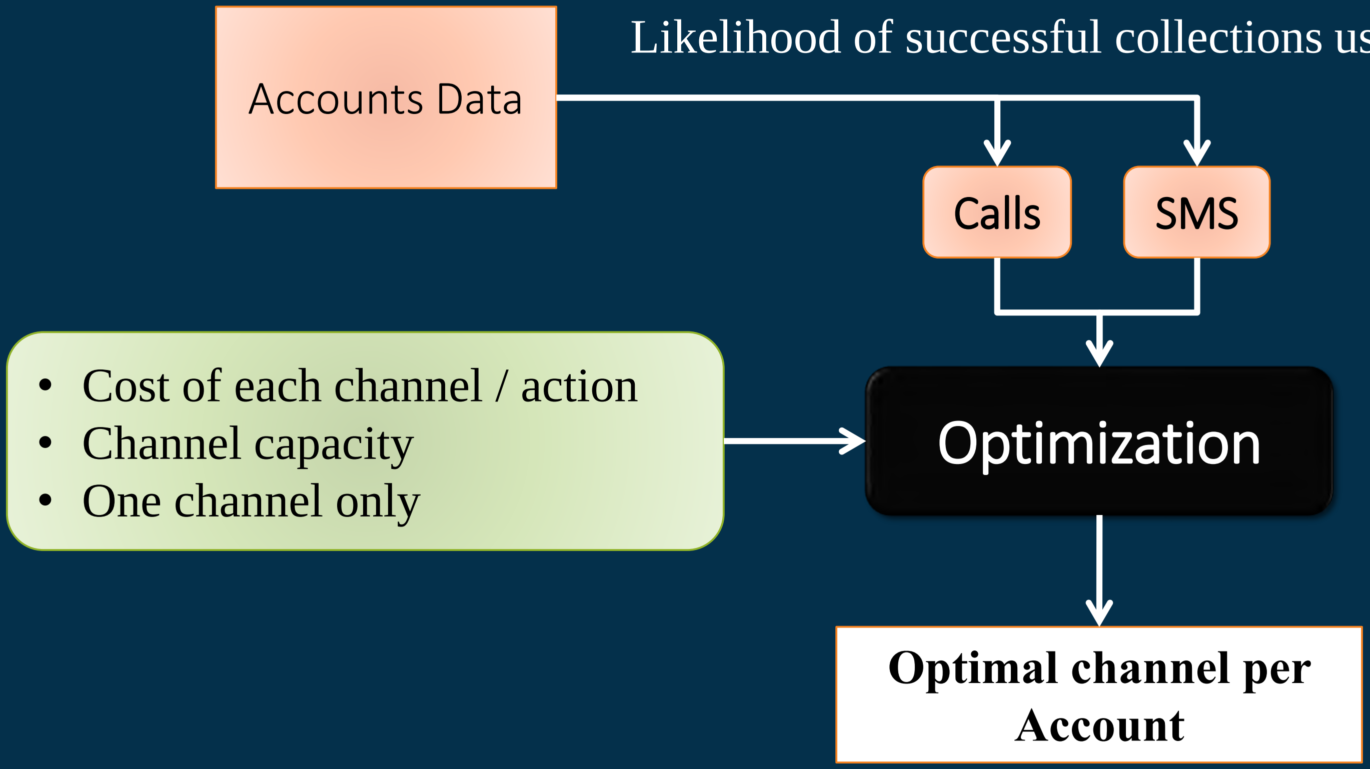


From Simple → Complex !



A Detailed Example

Collections Channel Optimization



A Two-channel Collections Problem

ID	Response Probability		Actions		Revenue (R)	Cost per action	
	Call (α)	SMS (β)	Call (x)	SMS (y)		Call (c)	SMS(s)
1	α1	β1	*x1*	*y1*	R1	c	s
2	α2	β2	*x2*	*y2*	R2	c	s
3	α3	β3	*x3*	*y3*	R3	c	s
4	α4	β4	*x4*	*y4*	R4	c	s
5	α5	β5	*x5*	*y5*	R5	c	s

Decision variables

Objective function

Constraints

Step 1 – Decision Variables

ID	Response Probability		Actions		Revenue (R)	Cost per action	
	Call (α)	SMS (β)	Call (x)	SMS (y)		Call (c)	SMS(s)
1	α1	β1	*x1*	*y1*	R1	c	s
2	α2	β2	*x2*	*y2*	R2	c	s
3	α3	β3	*x3*	*y3*	R3	c	s
4	α4	β4	*x4*	*y4*	R4	c	s
5	α5	β5	*x5*	*y5*	R5	c	s

Decision variables

⬇

Channels to optimize

- x1, x2, …, y1, y2, … are binary (0/1)
- x1=0 ➔ No Call
- x1=1 ➔ Call
- …
- Y1=0 ➔ No SMS
- Y2=1 ➔ SMS

Step 2 – Objective Function: Profit

ID	Response Probability		Actions		Revenue (R)	Cost per action	
	Call (α)	SMS (β)	Call (x)	SMS (y)		Call (c)	SMS(s)
1	$\alpha 1$	$\beta 1$	*x1*	*y1*	R1	c	s
2	$\alpha 2$	$\beta 2$	*x2*	*y2*	R2	c	s
3	$\alpha 3$	$\beta 3$	*x3*	*y3*	R3	c	s
4	$\alpha 4$	$\beta 4$	*x4*	*y4*	R4	c	s
5	$\alpha 5$	$\beta 5$	*x5*	*y5*	R5	c	s

Account 1

Expected Revenue $= max(\alpha_1 * x_1 + \beta_1 * y_1) * R_1$ **Constraint:** $x_1 + y_1 = 1$

$= (\alpha_1 * x_1 + \beta_1 * y_1) * R_1$

Cost $= x_1 * c + y_1 * s$

All Accounts

Profit $= \Sigma\{(a_i * x_i + b_i * y_i) * R_i - (x_i * c + y_i * s)\}$

Step 3 – Constraints

ID	Response Probability		Actions		Revenue (R)	Cost per action	
	Call (a)	SMS (β)	Call (x)	SMS (y)		Call (c)	SMS(s)
1	α1	β1	*x1*	*y1*	R1	c	s
2	α2	β2	*x2*	*y2*	R2	c	s
3	α3	β3	*x3*	*y3*	R3	c	s
4	α4	β4	*x4*	*y4*	R4	c	s
5	α5	β5	*x5*	*y5*	R5	c	s

<u>Account level constraint:</u>
Use only one channel

$x_1 + y_1 = 1$

$x_2 + y_2 = 1$

$x_3 + y_3 = 1$

$x_4 + y_4 = 1$

$x_5 + y_5 = 1$

<u>Channel level constraint:</u>
(Maximum Budget)

$x_1 + x_2 + x_3 + x_4 + x_5 \leq 3$

$y_1 + y_2 + y_3 + y_4 + y_5 \leq 3$

Formal Problem Statement (Integer Programming)

Maximize

$$
\begin{aligned}
\{ \; & (\alpha_1 * x_1 + \beta_1 * y_1) * R_1 - (x_1 * c + y_1 * s) \\
& + (\alpha_2 * x_2 + \beta_2 * y_2) * R_2 - (x_2 * c + y_2 * s) \\
& + (\alpha_3 * x_3 + \beta_3 * y_3) * R_3 - (x_3 * c + y_3 * s) \\
& + (\alpha_4 * x_4 + \beta_4 * y_4) * R_4 - (x_4 * c + y_4 * s) \\
& + (\alpha_5 * x_5 + \beta_5 * y_5) * R_5 - (x_5 * c + y_5 * s) \\
\} &
\end{aligned}
$$

Subject to

$x_1 + y_1 = 1;$

$x_2 + y_2 = 1;$

$x_3 + y_3 = 1;$

$x_4 + y_4 = 1;$

$x_5 + y_5 = 1;$

$x_1 + x_2 + x_3 + x_4 + x_5 \leq 3;$

$y_1 + y_2 + y_3 + y_4 + y_5 \leq 3;$

$x_1, x_2, x_3, x_4, x_5,$
y_1, y_2, y_3, y_4, y_5 *Integer*

Numerical Example

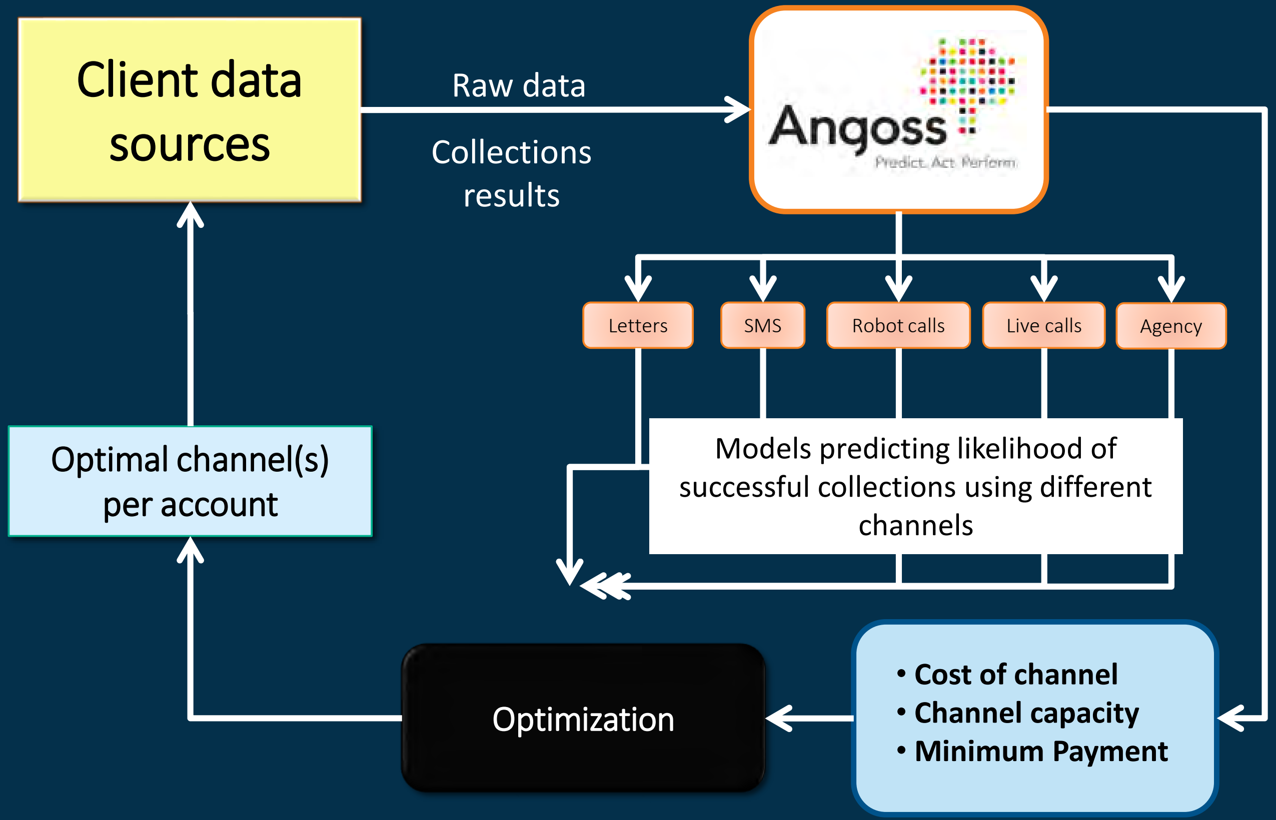
ID	Response Probability		Revenue (R)	Cost per action		Actions	
	Call (α)	SMS (β)		Call (c)	SMS(s)	**Calls (x)**	**SMS (y)**
1	0.42	0.42	80	$ 1.00	$ 0.50	**1**	**0**
2	0.40	0.72	200	$ 1.00	$ 0.50	**0**	**1**
3	0.66	0.82	80	$ 1.00	$ 0.50	**0**	**1**
4	0.35	0.41	150	$ 1.00	$ 0.50	**1**	**0**
5	0.10	0.20	120	$ 1.00	$ 0.50	**0**	**1**
						2	3

Calls ≤ 3; SMS ≤ 3;

Observations

- We may not use all the assigned budget
- Some accounts will not be contacted
- Accounts may be assigned the channel with the least likelihood to respond !
- To take account of previous contacts, use them as variables in the models
- We optimize "day-by-day" ←→ not the optimum value over a period of time
- Did not take account of write-offs

Actual Project – A North American Canadian Bank



Results – Lessons Learned

- Higher dollars collected
- Higher channel ROI

- More attention to higher balances
- Issues with mixed portfolios (more focus on some at the expense of others)

- Revising the models to consider write-offs and effect on capital reserve

Final Words of Wisdom !

- Optimization is fun
- Use it if it makes sense
- Pay more attention to the problem set up than anything else
- Use a good software
- Don’t attempt to write your own algorithm before trying to use available tools (don’t re-invent the wheel)

Thank You all …

mrefaat@Angoss.com

Your Feedback Counts!

Don't forget to complete the session survey
in your conference mobile app.

1. Go to the Agenda icon in the conference app.
2. Find this session title and select it.
3. On the Sessions page, scroll down to Surveys and select the name of the survey.
4. Complete the survey and click Finish.


#SASGF
SAS®
GLOBAL
FORUM
2018
April 8 – 11 | Denver, CO
Colorado Convention Center